Presentation of the constant r-term Krichever–Novikov-type algebras

Jean Nuyts and Isabelle Platten
Departement de Physique Theorique et Mathematique, Faculté des Sciences, Université de Mons, 7000 Mons, Belgium

(Received 12 September 1989; accepted for publication 8 November 1989)

For each constant r-term Krichever–Novikov-type algebra, a minimal set of defining generators is given from which all the operators of the algebra can be constructed by sequential definitions. A finite system of polynomial conditions on the defining generators is given, guaranteeing the commutation relations of the full algebra.

I. INTRODUCTION

The r-term Krichever–Novikov (KN)-type algebras

$$[N_m,N_n] = (m-n)\sum_{k=0}^{r-1} C^k(m,n)N_{m+n-r+1+2k} \tag{1.1}$$

generalize the classical Virasoro algebra ($r=1$)

$$[L_m,L_n] = (m-n)L_{m+n}. \tag{1.2}$$

Let us remark that the step in k has been chosen to be 2, in analogy with the original article of Krichever and Novikov.[1]

We here extend the *presentation* recently obtained of the Virasoro algebra[2] to the simplest cases of (1.1) when the $C^k(m,n)$ are all chosen equal to 1:

$$[N_m,N_n] = (m-n)\sum_{k=0}^{r-1} N_{m+n-r+1+2k}. \tag{1.3}$$

These constant r-term KN-type algebras, which are truly infinite, are presented as *purely finite* structures. All the operators of the algebra are constructed from a finite *set of defining generators* and all the commutation relations follow from a finite *system* of polynomial *conditions* imposed on the defining generators.

II. DEFINING GENERATORS. DEFINITIONS OF THE OPERATORS

It was shown[2] that all the Virasoro operators ($r=1$) can be constructed starting from a *minimal set* by a sequence of definitions. The simplest minimal set consists of just two generators; for example,

$$D_{r=1} \equiv \{L_3,L_{-2}\}. \tag{2.1a}$$

Then

$$\begin{aligned}
&L_1 = \tfrac{1}{5}[L_3,L_{-2}],\\
&L_{-1} = \tfrac{1}{3}[L_1,L_{-2}],\\
&L_2 = \tfrac{1}{4}[L_3,L_{-1}],\\
&L_0 = \tfrac{1}{2}[L_1,L_{-1}],\\
&L_{i+1} = [1/(i-1)][L_i,L_1], \quad i\geqslant 3,\\
&L_{-i-1} = [1/(1-i)][L_{-i},L_{-1}], \quad i\geqslant 2,
\end{aligned} \tag{2.1b}$$

is a sequence of definitions.

To generalize the minimal set to the arbitrary r term constant KN algebras, the cases $r=2$ and $r\geqslant 3$ should be treated separately.

For $r=2$, where the constant KN algebra reads

$$[N_m,N_n] = (m-n)(N_{m+n-1}+N_{m+n+1}), \tag{2.2}$$

the set

$$D_{r=2} \equiv \{N_1,N_{-2}\} \tag{2.3a}$$

is minimal with the sequential definitions

$$\begin{aligned}
&(\text{Def. }1)\quad N_0 = \tfrac{1}{3}[N_1,N_{-2}] - N_{-2},\\
&(\text{Def. }2)\quad N_2 = [N_1,N_0] - N_0,\\
&(\text{Def. }3)\quad N_{-1} = \tfrac{1}{4}[N_2,N_{-2}] - N_1,\\
&(\text{Def. }+)\quad N_{i+1} = (1/i)[N_i,N_0] - N_{i-1}, \quad i\geqslant 2,\\
&(\text{Def. }-)\quad N_{-i-1} = (1/i)[N_0,N_{-i}] - N_{-i+1}, \quad i\geqslant 2.
\end{aligned} \tag{2.3b}$$

For $r\geqslant 3$, in order to define all the N's, a minimal defining set contains at least $2(r-1)$ generators. Indeed, if a specific commutator is used to define one N only, $(r-1)$ N's from the r N's in the right-hand side have to be known already. But, since the index increases by step 2, there should be $(r-1)$ even N's and $(r-1)$ odd N's to start with.

The defining set

$$\begin{aligned}
D_r &\equiv \{N_{-r+2},N_{-r+3},\dots,N_{r-1}\}\\
&\equiv \{N_{-r+k+1},\quad 1\leqslant k\leqslant 2(r-1)\}
\end{aligned} \tag{2.4a}$$

has this minimal number of generators and hence is a minimal set. The sequential definitions are

$$\begin{aligned}
&(\text{Def. }1)\quad N_{-r+1} = [1/(4-2r)][N_{-r+2},N_{r-2}] - \sum_{k=0}^{r-2} N_{-r+3+2k},\\
&(\text{Def. }+)\quad N_{i+1} = [1/(2-r-i)][N_{-r+2},N_i] - \sum_{k=0}^{r-2} N_{i-2r+3+2k}, \quad i\geqslant r-1,\\
&(\text{Def. }-)\quad N_{-i-1} = [1/(r+i-2)][N_{r-2},N_{-i}] - \sum_{k=0}^{r-2} N_{-i+1+2k}, \quad i\geqslant r-1.
\end{aligned} \tag{2.4b}$$

With the minimal sets (2.1a), (2.3a), and (2.4a) and the sequential definitions (2.1b), (2.3b), and (2.4b) all the operators of the r-term constant KN algebras are obtained.

For each r, we now must find the conditions to be imposed on the minimal set so that the commutation relations are fulfilled for all the operators.

 0022-2488/90/030527-06$03.00

III. CONDITIONS

Clearly the minimal set, together with the sequential definitions, is not enough to guarantee the commutation relations of the algebra. Conditions on the defining generators have to be imposed. For any r, we construct a system of conditions for our definitions (2.1), (2.3), and (2.4); their number is given in the following proposition.

Proposition: The cardinal C_r of systems of conditions for arbitrary r is

$$C_{r=1} = 6, \quad C_{r=2} = 6, \quad C_{r\geqslant 3} = (r-1)(2r-1). \tag{3.1}$$

The proof of the case $r = 1$ can be found by slight changes (removal of the central charge) from the Virasoro presentation.[2] The proofs of the cases $r = 2$ and $r = 3$, which are also exceptional, can be found in an article written in honor of Professor R. Brout for his sixtieth birthday.[3]

Here, we limit ourselves to a detailed description of the generic case ($r \geqslant 4$) and, in Fig. 1, Fig. 2, and Fig. 3, to a résumé of the results for the cases $r = 1$, 2, and 3, respectively.

It should be noted that the proof of the proposition is not simple. The passage from a system of conditions to the commutation relations requires the construction of a rather subtle and detailed path, using what we will call *allowed Jacobi identities.* At every step in the path, the known commutators make certain Jacobi identities allowed which prove new commutation relations and, in turn, at the next step, new allowed Jacobi identities, etc.

The rest of the section will be organized as follows. We will first (Sec. III A) discuss the tools (*Jacobi identities*) that allow the determination of unknown commutators from known ones. The full system of conditions is given and discussed (Sec. III C). The commutation relations are separated in four ensembles (Sec. III B), a finite one around the conditions and three infinite ones. The determination proceeds in each region in turn (Secs. III D–III G).

A. Jacobi identities

To prove the commutation relations from the conditions, the tools are the Jacobi identities. Let us immediately stress that a Jacobi identity

$$[[N_m,N_n],N_p] + [[N_n,N_p],N_m] + [[N_p,N_m],N_n] = 0 \tag{3.2}$$

can be used to provide useful information if and only if the

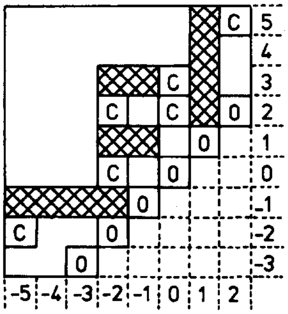

FIG. 1. Conditions for $r = 1$ in the $m \leqslant n$ lattice corresponding to $[N_m,N_n]$. The crosses are the commutators known by our definitions (2.1b). They extend to infinity upwards in the column 1 and at the left in row -1. The boxes marked C form a system of conditions. All the other boxes, extending to the full half-plane, are proved by sequences of Jacobi identities.

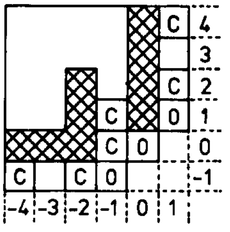

FIG. 2. Conditions for $r = 2$ in the $m \leqslant n$ lattice corresponding to $[N_m,N_n]$. The crosses are the commutators known by our definitions (2.2b). They extend to infinity upwards in the column 0 and at the left in row 0. The boxes marked C form a system of conditions. All the other boxes, extending to the full half-plane, are proved by sequences of Jacobi identities.

inside commutators $[N_m,N_n]$, $[N_n,N_p]$, and $[N_p,N_m]$ are already known by some preceding step in the argument. Once the inside commutators are known, the Jacobi identity will be called *allowed* and be represented by $J(m,n,p)$. It reduces to a linear relation between $3r$ commutators. Introducing $Y_{a,b}$ as a shorthand for

$$Y_{a,b} \equiv \sum_{k=0}^{r-1} [N_{a-r+1+2k},N_b] \tag{3.3}$$

and the corresponding *grade* $\Sigma = a + b$, the allowed $J(m,n,p)$,

$$(m-n)Y_{m+n,p} + (n-p)Y_{n+p,m} + (p-m)Y_{p+m,n} = 0, \tag{3.4}$$

relates three Y's of the same grade. It can be shown that repeated use of (3.4), when the Jacobi identities are allowed, enables one to determine all the Y's of given grade when three well-chosen ones are known.

In general, a Jacobi identity will prove an unknown commutation relation if it is allowed and if two of the Y's in (3.4) are fully known as well as $r-1$ of the r commutators of the third Y, the remaining one being the unknown. Obviously, other but less frequent situations may occur when the unknown commutator appears in more than one Y.

B. Regions

In order to extend the proof to all the commutators (1.3), it is useful to associate to the commutator $[N_m,N_n]$ a box in a (m,n) lattice with, taking the antisymmetry into account, $m \leqslant n$ (the boxes with $m = n$ will be labeled by zero). The lattice (m,n) is conveniently divided into four regions roughly sketched in Fig. 4:

region (α), the central region containing, in particular, the conditions;

region (β), the (extended) positive–positive region;

region (γ), the (extended) negative–negative region;

region (δ), the (restricted) positive–negative region.

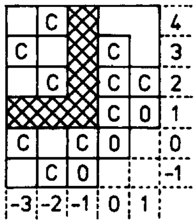

FIG. 3. Conditions for $r = 3$ in the $m \leqslant n$ lattice corresponding to $[N_m,N_n]$. The crosses are the commutators known by our definitions (2.3b). They extend to infinity upwards in the column -1 and at the left in row 1. The boxes marked C form a system of conditions. All the other boxes, extending to the full half-plane, are proved by sequences of Jacobi identities.

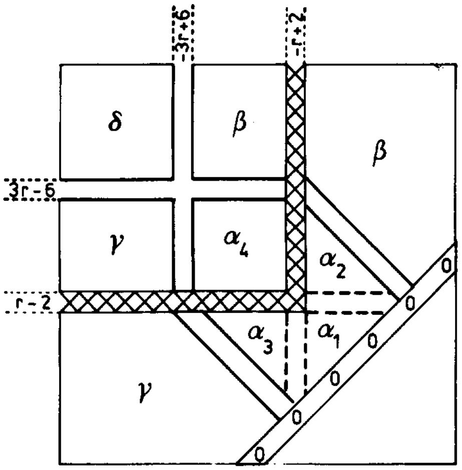

FIG. 4. For r general. A rough sketch of the $m \leqslant n$ lattice divided into the four regions: region $\alpha (= \cup_i \alpha_i)$, finite central region where the conditions are located; region β, extended positive–positive region; region γ, extended negative–negative region; and region δ, reduced positive–negative region. Crosses indicate commutators known by definition.

The precise limits of these regions will be explained in the next sections. A detailed study, commutator by commutator, will be needed in the central region. The other regions will be treated by inductive arguments.

C. Conditions

In Fig. 5, the latttice corresponding to the subregion α_1 of α is drawn. The commutators corresponding to the definitions (2.4b) and hence known are cross-ruled.

Our system of conditions is labeled by the letter C.

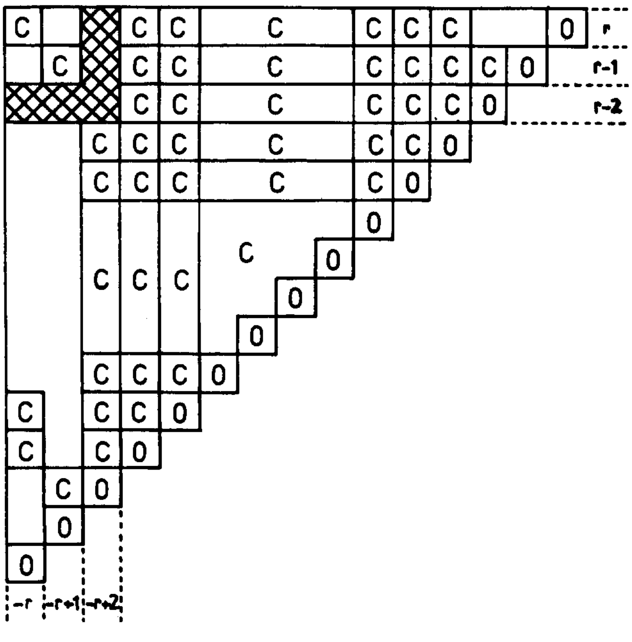

FIG. 5. Conditions for general r in the $m \leqslant n$ lattice corresponding to $[N_m, N_n]$. The crosses are the commutators known by our definitions (2.4b). They extend to infinity upwards in the column $-r+2$ and at the left in row $r-2$. The boxes marked C form a system of conditions. All the other boxes, extending to the full half-plane, are proved by sequences of Jacobi identities.

Since Jacobi identities are written in terms of $Y_{a,b}$'s, to obtain a commutator from a $Y_{a,b}$ that involves r commutators, $r-1$ commutators of given b should be known. This remark shows that there is not relation between the boxes inside the lower right triangle limited by the vertical line $-r+2$ and the horizontal line $r-2$. Remember the step increase 2 when checking that there are always less than $r-1$ commutators of given b. The related commutators are hence independent and can be taken as conditions.

As soon as one leaves this triangle, Jacobi identities start playing a role and a detailed study has to be performed. Extra conditions then appear in the vertical lines $-r$ and $-r+1$ and in the horizontal lines $r-1$ and r, as shown in Sec. III D.

D. Region α. The central region

As will be seen later (Secs. III E–III G), if all the commutators are known in a finite region $\bar{\alpha}$, which will be defined soon, an inductive argument enables one to prove all the commutators in the regions β, γ, and δ of infinite extent. The region $\bar{\alpha}$ is represented in Fig. 6. It is bounded by the oblique lines $m+n=-2r+3$ and $m+n=2r-3$, which are partially included in it, by the vertical lines $m=-3r+5$ when $n=3r-2y$ $(y=3,...,r)$ and $m=-3r+6$ when $n=3r-2y-1$ $(y=3,...,r)$, and finally by the mirror horizontal lines $n=3r-5$ when $m=-3r+2y$ $(y=3,...,r)$ and $n=3r-6$ when $m=-3r+2y+1$ $(y=3,...,r)$. The limiting boxes are again included in region $\bar{\alpha}$. Region α, where we first prove the commutation relations, is composed of region $\bar{\alpha}$, to which are added the oblique lines $m+n=-2r+2$ $(m \geqslant -3r+5)$ and $m+n=2r-2$ $(m \geqslant -r+3)$, and the four boxes $(m,n)=(-r+1,3r-5)$, $(-r+1,3r-4)$, $(-3r+5,r-1)$, $(-3r+4,r-1)$.

Region of the conditions, region α_1: All the conditions are summarized in Fig. 5. Under the next three subheadings we discuss the remaining (determined) white boxes.

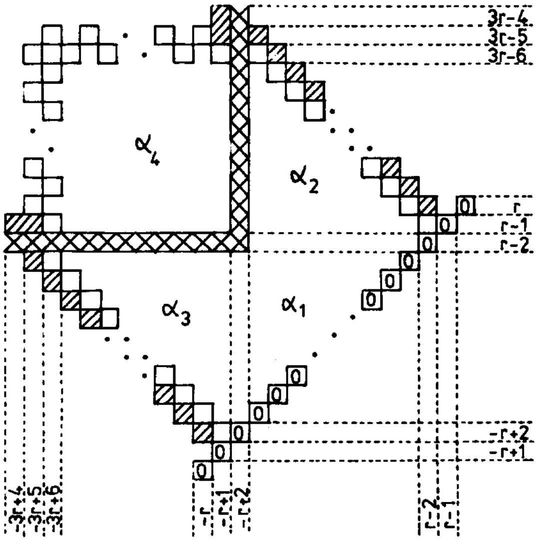

FIG. 6. Region α. In the $m<n$ lattice the precise limits of region α and of its subregion $\bar{\alpha}$. The region $\alpha - \bar{\alpha}$ is hatched.

The $m=-r+1(-r+2\leqslant n<r-2)$ vertical and the $n=r-1(-r+2<m\leqslant r-2)$ horizontal lines: Let us first consider the vertical line $-r+1$ below the definitions $(-r+2\leqslant n<r-2)$ and the horizontal line $r-1$ at the right of the definitions $(-r+2<m\leqslant r-2)$.

The Jacobi identities

$$J(-r+2,i,r-2), \quad -r+2<i<r-2, \tag{3.5}$$

are allowed as a result of the conditions of Fig. 5, as can be checked by inspection. Hence

$$(-r+2-i)Y_{i-r+2,r-2}+(i-r+2)Y_{i+r-2,-r+2} + (2r-4)Y_{0,i}=0. \tag{3.6}$$

The two first Y's in (3.6) are known. All the commutators in $Y_{0,i}$ are known except $(i,r-1)$ and $(1-r,i)$. If $(i,r-1)$ is taken as a condition, $(1-r,i)$ is determined.

By the limitation on i in (3.5), the extra boxes $(-r+1,-r+2)$ and $(r-2,r-1)$ which cannot be reached, both have to be added as conditions.

The $m=-r(-r+3\leqslant n\leqslant r-3)$ vertical and the $n=r(-r+3\leqslant m\leqslant r-3)$ horizontal lines: To the unknown $(-r,a)$ $(-r+5\leqslant a\leqslant r-3)$ box, let us associate $Y_{-1,a}$, which is entirely known except for this very commutator $(-r,a)$. Analogously, we associate $Y_{1,b}$ to (b,r) $(r+3\leqslant b\leqslant r-5)$. The grades are $\Sigma=a-1$ and $\Sigma=b+1$, respectively, with $(-r+4\leqslant\Sigma\leqslant r-4)$.

With the same grade, $Y_{\Sigma-r+2,r-2}$, $Y_{\Sigma+r-2,-r+2}$, and the Jacobi related $Y_{0,\Sigma}$ are known. These three Y's are not enough to determine all the Y's of the same grade (they are not independent).

If $Y_{1,b}(-r+3\leqslant b\leqslant r-5)$ is known, the sequence of allowed Jacobi identities,

$$\begin{aligned}&J(-r+2,b,r-1)\rightarrow Y_{b-r+2,r-1},\\ &J(-r+1,b+1,r-1)\rightarrow Y_{b+r,-r+1},\\ &J(-r+1,b+2,r-2)\rightarrow Y_{-1,b+2}\equiv Y_{-1,a},\end{aligned} \tag{3.7}$$

determines $Y_{-1,a}$ and hence the result: if (b,r) is taken as a condition, $(-r,b+2)$ is determined.

No such argument holds for the commutators $(-r,-r+3)$, $(-r,-r+4)$, $(r-4,r)$, and $(r-3,r)$, which have to be taken as conditions. They correspond to $|\Sigma|=r-3$ or $|\Sigma|=r-2$.

The commutators $(-r+1,r-1)$ and $(-r,r)$, $(|\Sigma|=r-1)$: Let us now consider the grade $\Sigma=r-1$. The $Y_{2r-3,-r+2}$ is known. By our general arguments, at least two more Y's of the same grade have to be fixed in order to determine all of them. If $(-r+1,r-1)$ is added as a condition, $Y_{0,r-1}$ becomes known. The allowed Jacobi identity $J(-r+2,r-2,r-1)$ determines $Y_{1,r-2}$ and hence $(r,r-2)$. Finally, the addition of $(-r,r)$ as a condition completes the knowledge of $Y_{-1,r}$.

This argument ends the determination of our system of conditions. All the missing commutators of the region α and of all the other regions can be inferred from this system.

The remaining part of α is again divided into three regions, sketched in Fig. 4 and drawn more precisely in Fig. 6: the upper right triangle (region α_2), the lower left triangle (region α_3), and the upper left approximate square (region α_4).

The upper right triangle of region α, region α_2: The upper right triangle is filled by oblique lines with fixed level $\Lambda=m+n$ inductively by letting Λ increase from 4 to $2r-2$. For this fixed Λ, the boxes $(-r+\Lambda-p,r+p)$ $(1\leqslant p\leqslant\Lambda-3)$ are associated to $Y_{p+1,-r+\Lambda-p}$ known up to this commutator.

For the corresponding grade $\Sigma=\Lambda-r+1$ $(5-r\leqslant\Sigma\leqslant r-1)$, three Y's are fully known: $Y_{1,\Sigma-1}$, $Y_{\Sigma+r-2,-r+2}$, and $Y_{\Sigma-r+2,r-2}$. The $\Sigma+r-4$ unknowns $Y_{p+1,\Sigma-p-1}$, i.e., to be specific $Y_{2,\Sigma-2}$, $Y_{3,\Sigma-3},\ldots,Y_{\Sigma+r-3,-r+3}$, are then obtained by the following set of allowed Jacobi identities:

$$\begin{aligned}&J(-r+2+a,\Sigma-1-a,r-1),\\ &J(-r+3+a,\Sigma-1-a,r-2), \quad a=0,\ldots,\end{aligned} \tag{3.8}$$

which, starting with $a=0$, determines first $Y_{\Sigma-r+1,r-1}$, then sequentially $Y_{\Sigma+r-3,-r+3}, Y_{2,\Sigma-2}, Y_{\Sigma+r-4,-r+4}$, $Y_{3,\Sigma-3},\ldots$. Remark that the interval is filled starting from both ends in turn.

The lower left triangle of region α, region α_3: This reasoning can be applied by symmetry, *mutatis mutandis*, to the lower left triangle.

The upper left square of region α, region α_4: The upper left square has again to be divided into the $-r+1$ and the $-r$ vertical lines and in the $r-1$ and r horizontal lines. The remaining of region α is then scanned sequentially in vertical and horizontal lines crossing on boxes on the oblique line $(-a,a)$.

The $m=-r+1(r\leqslant n\leqslant 3r-4)$ vertical and the $n=r-1(-3r+4\leqslant m\leqslant -r)$ horizontal lines:

To the boxes $(-r+1,r+j)$ $(0\leqslant j\leqslant 2r-6)$ in the $m=-r+1$ vertical line, let us associate $Y_{j+1,-r+1}$, which will be sequentially determined and hence fix the desired commutator. For this grade $\Sigma=j-r+2$, both $Y_{j-2r+4,r-2}$ and $Y_{-1,-r+3+j}$ are known. Hence the allowed Jacobi identity $J(-r+1,-r+3+j,r-2)$ gives the expected result.

For the box $(-r+1,3r-5)$ associated to $Y_{2r-4,-r+1}$, the allowed Jacobi identity $J(-r+1,r-3,r-1)$ provides the wanted commutator since $Y_{0,r-3}$ and $Y_{-2,r-1}$ are known by inspection.

For the last box $(-r+1,3r-4)$ associated to $Y_{2r-3,-r+1}$, the relevant allowed $J(-r+1,r-2,r-1)$ gives the result since $Y_{0,r-2}$ and $Y_{-1,r-1}$ are known.

The $n=r-1$ horizontal line follows by the usual symmetry.

The $m=-r(r+1\leqslant n\leqslant 3r-5)$ vertical and the $n=r(-3r+5\leqslant m\leqslant -r-1)$ horizontal lines: These two lines are the last ones for which the intersection box $(-r,r)$ on the diagonal line $(-a,a)$ is a condition.

To the box $(-r,n)$ $(r+1\leqslant n\leqslant 3r-5)$ on the vertical line, one associates $Y_{n-r+1,-r}$, which will sequentially determine the associated commutator through the allowed $J(-r,n-2r+3,r-2)$ since $Y_{n-3r+3,r-2}$ and $Y_{-2,n-2r+3}$ are already known.

The horizontal line follows again by symmetry.

The remaining region α_4: To obtain the commutation relations in the remaining of region α_4, we will prove sequentially the commutator $(-p,p)$ $(r+1\leqslant p\leqslant 3r-6)$ together with the lines (within α) above and on the left of the $(-p,p)$ box.

Suppose we know these commutators up to $p=a-1$; we want to prove them for $p=a$.

First to the box $(-a,a)$ we associated $Y_{r-a-1,a}$ for which $(-a,a)$ is the only unknown. The allowed Jacobi identity $J(-r+2,2r-a-3,a)$ fixes it since $Y_{2r-3,-r+2}$ is known as well as $Y_{a-r+2,2r-a-3}$.

Consider now the boxes $(-a,n)$ $(a+1\leqslant n\leqslant 3r-6)$ on the vertical line above $(-a,a)$. Associate to it $Y_{n-r+1,-a}$ for which sequentially in n, $(-a,n)$ is the only unknown commutator. The allowed Jacobi identity $J(-a,n-2r+3,r-2)$ fixes it since $Y_{n-2r-a+3,r-2}$ and $Y_{r-a-2,n-2r+3}$ are known.

Finally consider the limiting box $(-a,3r-5)$, $a=r+2i$, $0\leqslant i\leqslant r-3$. Another Jacobi identity $J(-a,r-3,r-1)$ can be used since $Y_{r-a-1,r-3}$ and $Y_{r-a-3,r-1}$ are known. It should be remarked that it is precisely for this last Y that we need to extend the proof to the box $(-3r+4,r-1)$ in the $n=r-1$ horizontal line treated before.

The horizontal line (m,a) $(-3r+5\leqslant m\leqslant -a-1)$ is treated by the usual symmetry argument.

This finishes the proof of the commutators within the finite region α and hence within subregion $\bar{\alpha}$ needed to complete inductively the full determination in the other infinite regions.

E. Region β. Extended positive–positive region

In Fig. 7 and Fig. 8, corresponding to an even and to an odd case, respectively, a set of boxes are drawn. To prove the commutation relations for the black boxes of these figures one needs to know the dotted boxes to allow certain Jacobi identities and the hatched boxes to use these Jacobi identities.

Even case (level $2p$, $p>r-2$): The $(p+3r-7)$ allowed Jacobi identities

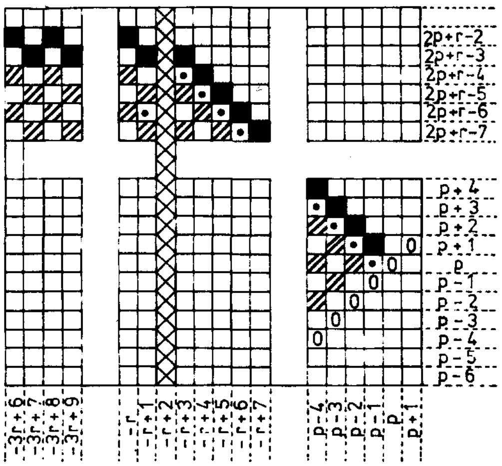

FIG. 7. Region β. Even case. In the $m<n$ lattice of region β, the black commutators are obtained in terms of the hatched ones. The dotted commutators are also necessary to allow the Jacobi identities (3.9).

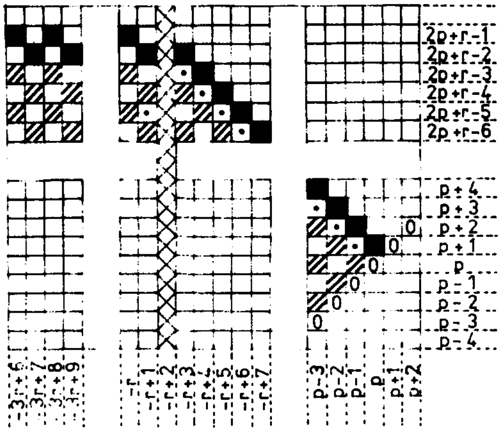

FIG. 8. Region β. Odd case. In the $m<n$ lattice of region β, the black commutators are obtained in terms of the hatched ones. The dotted commutators are also necessary to allow the Jacobi identities (3.11) and (3.13).

$$J(-r+2,p-1-s,p+s),\quad 0\leqslant s\leqslant p+r-4, \quad (3.9a)$$

$$J(1-r-2s,4-r+2s,2p+r-6-2s),\quad 0\leqslant s\leqslant r-3, \quad (3.9b)$$

$$J(-r-2s,5-r+2s,2p+r-6-2s),\quad 0\leqslant s\leqslant r-3, \quad (3.9c)$$

give, respectively, the $(p+3r-7)$ commutators

$$[N_{p-s-1},N_{p+s+1}],\quad 0\leqslant s\leqslant p+r-4, \quad (3.10a)$$

$$[N_{1-r-2s},N_{2p+r-3}],\quad 0\leqslant s\leqslant r-3, \quad (3.10b)$$

$$[N_{-r-2s},N_{2p+r-2}],\quad 0\leqslant s\leqslant r-3, \quad (3.10c)$$

i.e., those of the black boxes of Fig. 7 in terms of known commutators of lower level.

Odd case (level $2p+1$, $p>r-2$): The $(p+3r-7)$ allowed Jacobi identities

$$J(-r+2,p-1-s,p+1+s),\quad 0\leqslant s\leqslant p+r-4, \quad (3.11a)$$

$$J(1-r-2s,4-r+2s,2p+r-5-2s),\quad 0\leqslant s\leqslant r-3, \quad (3.11b)$$

$$J(-r-2s,5-r+2s,2p+r-5-2s),\quad 0\leqslant s\leqslant r-3, \quad (3.11c)$$

are not enough to prove the $(p+3r-6)$ commutators

$$[N_{p-s},N_{p+1+s}],\quad 0\leqslant s\leqslant p+r-3, \quad (3.12a)$$

$$[N_{r-2s},N_{2p+r-2}],\quad 0\leqslant s\leqslant r-3, \quad (3.12b)$$

$$[N_{-r-2s},N_{2p+r-1}],\quad 0\leqslant s\leqslant r-3. \quad (3.12c)$$

The set (3.11a) has to be supplemented by one Jacobi identity

$$J(-r+3,-r+4,2p+r-5). \quad (3.13)$$

In order for (3.11a) and (3.13) to prove (3.12a), a determinant has to be different from zero:

$$\det M=(-)^{p+r-1}[(p+r-3)!/3!]2(p+r-4)\times(2p+2r-3)(2p+2r-1). \quad (3.14)$$

F. Region γ. Extended negative–negative region

An analogous discussion holds obviously for region γ.

G. Region δ. Reduced positive–negative region

Once regions α, β, and γ are completed, region δ can be filled by using a sliding structure based on the allowed Jacobi identity $J(-r+2,-n,m-1)$:

$$\begin{aligned} J(-r+2,-n,m-1) & \\ \equiv (n-r+2)&[N_{-2r-n+3}+\cdots+N_{1-n},N_{m-1}] \\ &+(1-n-m)[N_{m-n-r}+\cdots \\ &+N_{m-n-2+r},N_{-r+2}] \\ &+(m+r-3)[N_{m-2r+2}+\cdots+N_m,N_{-n}]=0. \end{aligned} \tag{3.15}$$

It determines $[N_m,N_{-n}]$ when some (*a fortiori* all) commutators below the line m are known.

The sliding structure is given in Fig. 9. The black commutator is deduced from the dotted commutator allowing the preceding Jacobi identity and from the hatched ones. (Remember that all the commutators of N_{-r+2} are in the region supposed to be known.)

By the arguments of the preceding pages, we have proved that all the commutators can be deduced once our systems of conditions are imposed on our defining sets.

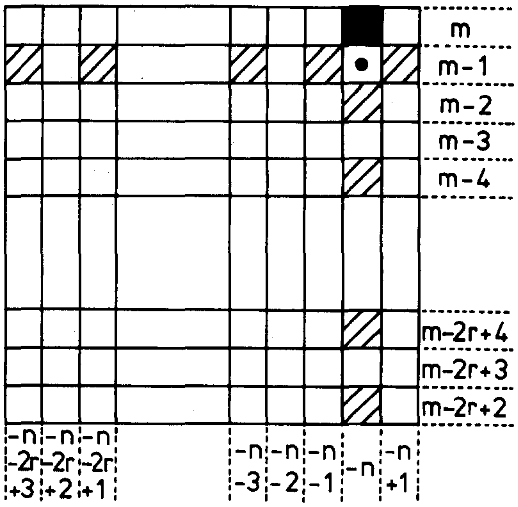

FIG. 9. The sliding structure. The black box commutator $(-n,m)$ is determined from the hatched commutators by a Jacobi identity allowed once the dotted commutator is known.

IV. CONCLUSIONS AND REMARKS

By constructing definite sets of defining operators and definite systems of conditions, we have shown that the constant r-term Krichever–Novikov-type algebras are finite structures.

The sets of defining generators [with two elements for $r=1$ and $2(r-1)$ elements for $r\geqslant 2$] we have proposed have the smallest possible number of operators. However, it is easy to see that there are minimal sets, i.e., sets for which no subset is already defining, with higher number of operators. This poses the problem of the classification of minimal sets.

On the other hand, the systems of conditions we have obtained again involve (see the Proposition) a finite number of conditions to guarantee the commutation relations. Once written in terms of the elements of the defining sets, these conditions become polynomials of high degree, in the form of nested multiple commutators.

If other minimal sets are chosen, with different number of elements N_D, we expect the number of conditions C to vary. It would be interesting to see whether there exists some invariant function of N_D and C which would hence be a characteristic parameter of the algebra.

Needless to say, the constant ones are only one example of the general [see (1.1)] r-term Krichever–Novikov-type algebras. The full classification of these algebras seems still to be lacking as the restriction on the $C^k(m,n)$ imposed by the Jacobi identities has not yet been solved in general.[1,4] The problem of the presentation may also be asked for these other algebras.

[1]I. Krichever and S. Novikov, Funct. Anal. Appl. **21**, 126 (1987) [Funct. Anal. Ego Prohozh. **21**, 46 (1987)].
[2]D. B. Fairlie, C. K. Zachos, and J. Nuyts, Commun. Math. Phys. **117**, 595 (1988); J. L. Uretsky, *ibid.* **122**, 1971 (1989); C. K. Zachos, "A presentation for Virasoro algebras," ANL-HEP-CP-88-49, 1988.
[3]J. Nuyts and I. Platten, "Presentation of r- terms constant Krichever–Novikov Algebras for r up to five," presented in honor of Professor R. Brout for his sixtieth birthday, Brussels, 1989.
[4]D. B. Fairlie, P. Fletcher, and J. Nuyts, J. Math. Phys. **30**, 957 (1989).

Fundamental representations of the m-principal realization of $\mathbf{gl}_\infty$

R. K. Dodd[a)]
Research Institute for Theoretical Physics, Hiroshima University, Takehara, Hiroshima 725, Japan

(Received 31 May 1988; accepted for publication 11 October 1989)

The m-principal realization of $\mathbf{gl}_\infty$ is introduced and its fundamental representations are defined on the representation space associated with the homogeneous Heisenberg subalgebra of $\mathbf{a}_m^{(1)}$ introduced by Frenkel and Kac. The Hirota equations associated with the representations are briefly reviewed. The main results concern the reduction of the algebras to the finite rank affine Lie algebras $\mathbf{gl}_n^{(1)}$ and $\mathbf{a}_{n-1}^{(1)}$ where it is shown that a complete nonredundant set of realizations can be obtained. This means that the associated Hirota equations can be obtained directly from the reduction of the Hirota equations for the fundamental m-principal representations of $\mathbf{gl}_\infty$ without requiring the complete nonredundant set of fundamental representations of the Lie algebra $\mathbf{a}_{n-1}^{(1)}$ itself.

I. INTRODUCTION

Recently some papers have appeared[1,2] that attempt to find all the solvable (integrable) equations that have a certain form. Since the integrability in two independent variables is related to the representation theory of affine Lie algebras an alternative approach might be to attempt a classification of the equations in this way. The papers by the Kyoto school[3–9] contain many results of this type, but as far as the author is aware no exhaustive classification has been attempted.

The most suitable algebra to start with is $\mathbf{a}_{n-1}^{(1)}$ since it has the simplest representation theory. However, even in this case, the set of canonical representations of the fundamental modules, defined by the maximal Heisenberg subalgebras, is not very convenient to use because it depends implicitly on elements $w\in W$, the Weyl group of $\mathbf{a}_{n-1}$.[10] This approach also omits equations such as the Kadomtsev–Petviashvili and the two-dimensional nonlinear Schrödinger equation[3] that are related to the representation theory of infinite rank affine Lie algebras.

For this reason we study the representations of the m-principal realization of the infinite rank affine Lie algebra $\mathbf{gl}_\infty$. The J reductions introduced by Date *et al.*[4] are then shown to produce a canonical set of realizations of $\mathbf{gl}_n^{(1)}$ and $\mathbf{a}_{n-1}^{(1)}$ by identifying $p(n)$ distinct maximal Heisenberg subalgebras as required by a Kac–Peterson theorem.[10]

The representation space for the fundamental modules of this realization is identical to that which arises in the homogeneous representation of the fundamental modules of $\mathbf{a}_m^{(1)}$.[11] However the lattice that occurs is not the root lattice. By using the wedge representation[12] the action of the groups associated with the m-principal realization of $\mathbf{gl}_\infty$ can be studied. This enables an interpretation to be given to the lattice that arises in the vertex representations and also results in explicit formulae defining the isomorphism between the two representations.

In Sec. IV besides considering the fundamental representations of the m-principal realization of $\mathbf{gl}_\infty$ we consider the fundamental representations of $\mathbf{gl}_n^{(1)}$ and $\mathbf{a}_{n-1}^{(1)}$ which can be obtained by their J reduction. The complete nonredundant set of fundamental representations of $\mathbf{a}_{n-1}^{(1)}$ is given in Sec. V by relating our work to the results of Lepowsky.[13]

In the last section the wedge representation is used to derive a restricted class of solvable equations in their Hirota form.[14] These were first derived in Ref. 7 using the spinor representation.

The main import of the results in Sec. IV is that the Hirota equations associated with a complete nonredundant set of fundamental representations of $\mathbf{a}_{n-1}^{(1)}$ can be obtained directly from the J reductions of the Hirota equations of the m-principal realizations of $\mathbf{gl}_\infty$. This analysis is given in Ref. 15.

II. THE m-PRINCIPAL REALIZATION OF $\widehat{\mathbf{gl}}_\infty$

The infinite rank affine Lie algebras $\widehat{\mathbf{gl}}_\infty$ and $\widehat{\mathbf{a}}_\infty$ have, as is well known,[3,16] the Dynkin diagram

$$\cdots\!-\!\!\circ\!\!-\!\!\circ\!\!-\!\!\circ\!\!-\!\!\circ\cdots.$$

Realizations of the Lie algebras can be given in terms of a complex matrix Lie algebra $\mathbf{gl}(m\infty)$. Let

$$E_{i,j}^{a,b}=(\delta_{r,i}^{h,a}\delta_{j,s}^{b,l})_{\substack{r,s\in Z\\ i\leqslant h,l\leqslant m}},$$

$$E_{i,j}^{a,b}E_{r,s}^{c,d}=\delta_{j,r}^{b,c}E_{i,s}^{a,c}\quad(\delta_{j,r}^{b,c}=\delta^{b,c}\delta_{j,r}),$$

then a basis for $\mathbf{gl}(m\infty)$ is given by $\{E_{i,j}^{a,b}\}$, $1\leqslant a,b\leqslant m$, $i,j\in Z$. It is convenient to adopt a dummy summation convention so that $g=g_{i,j}^{a,b}E_{i,j}^{a,b}\in\mathbf{gl}(m\infty)$ and all but a finite number of the entries $g_{i,j}^{a,b}$ are zero. Let $[\cdot,\cdot]_0$ denote the matrix commutator. The algebra $\hat{\mathbf{a}}(m\infty)\subset\mathbf{gl}(m\infty)$ is the subalgebra of traceless matrices.

The algebra $\mathbf{gl}(m\infty)$ acts on the space $V=\oplus_{a=1}^m V^a$, where $V^a\approx\mathbb{C}^\infty$. Let $\{u_i^a\}$ denote the standard basis for V^a, that is, u_i^a is the column vector with 1 in the ith entry of the

[a)] Permanent address: School of Mathematics, Trinity College, Dublin 2, Ireland.

ath copy of C^∞ and zero elsewhere. Then $E^{a,b}_{i,j}u^b_j = u^a_i$ and $v \in V$ has only a finite number of nonzero components.

We choose for the Chevalley generators $\{e_i, f_i\} i \in Z$ and simple coroots $\Pi^\vee = \{\bar{\alpha}^\vee_j\}$ the elements

$$\begin{aligned} &e_{m(r-1)+a} = E^{a,a+1}_{r,r}, \quad f_{m(r-1)+a} = E^{a+1,a}_{r,r}, \\ &a = 1,\dots,m-1, \\ &e_{m(r-1)+m} = E^{m,1}_{r,r+1}, \quad f_{m(r-1)+m} = E^{1,m}_{r+1,r}, \\ &\bar{\alpha}^\vee_r = [e_r, f_r]_0, \quad r \in Z. \end{aligned} \tag{2.1}$$

A completion of the Lie algebra $\mathbf{gl}(m\infty)$, denoted $\overline{\mathbf{gl}}(m\infty)$, is defined by requiring that in any block a,b, the submatrix $(g^{a,b}_{ij})_{i>h,\, j<h}$ (a,b fixed) has only a finite number of nonzero entries in any row or column. More formally define $\overline{V}$ as a completion of V by $\overline{V} = \{c^a_i u^a_i : c^a_i = 0 \ i \gg 0\}$ and its dual $\overline{V}^* = \{c^{a*}_i u^{a*}_i : c^{a*}_i = 0 \ i \ll 0\}$. Then the pairing between $\overline{V}^*$ and $\overline{V}$ is well defined, $\langle x, y^* \rangle < \infty, x \in \overline{V}, y^* \in \overline{V}^*$. Therefore the family of seminorms given by $\overline{V}^*$ define a weak topology on $\overline{V}$ and $\overline{V}$ becomes a locally convex topological vector space. Let A be the associative algebra of continuous endomorphism of $\overline{V}$. Then $\overline{\mathbf{gl}}(m\infty)$ is the Lie algebra associated with A and $\overline{\mathrm{GL}}(m\infty)$ is the Lie group of invertible mappings. We denote by $\mathrm{GL}(m\infty)$ the Lie group associated with $\widehat{\mathbf{gl}}(m\infty)$; that is the group of invertible matrices that have only a finite number of off-diagonal entries.

The Lie algebra $\mathbf{gl}(m\infty)$ is the central extension of $\overline{\mathbf{gl}}(m\infty)$ defined by the $\mathbb{C}$-valued cocycle $\Psi(\cdot,\cdot)$,

$$\Psi(E^{a,b}_{i,j}, E^{c,d}_{r,s}) = \delta^{b,c}_{j,r}\delta^{a,d}_{i,s}(\theta(i) - \theta(j))$$

$$\theta(i) = \begin{cases} 1, & i \leqslant 0, \\ 0, & i > 0. \end{cases}$$

Explicitly we have

$$\mathbf{gl}(m\infty) = \overline{\mathbf{gl}}(m\infty) \oplus \mathbb{C}z,$$

with the bracket on $\mathbf{gl}(m\infty)$ $[\cdot,\cdot]$, defined by

$$[g_1 + \lambda_1 z, g_2 + \lambda_2 z] = [g_1, g_2]_0 + \Psi(g_1, g_2)z,$$
$$\lambda_i \in \mathbb{C}, \quad g_i \in \overline{\mathbf{gl}}(m\infty). \tag{2.2}$$

The element $z \in \mathbf{gl}(m\infty)$ is the canonical central element since the Chevalley generators, viewed as elements of $\mathbf{gl}(m\infty)$, give

$$\alpha^\vee_i = [e_i, f_i] = \bar{\alpha}^\vee_i + \delta_{i,0} z,$$

and

$$z = \sum_{i \in Z} \alpha^\vee_i,$$

where $\Pi^\vee = \{\alpha^\vee_i\}$ are the simple coroots of $\mathbf{gl}(m\infty)$. The Cartan subalgebra $\underline{h} \subset \mathbf{gl}(m\infty)$ is spanned by $\Pi^\vee$ and z.

The algebra $\mathbf{gl}(m\infty)$ contains the subalgebra $\overset{m}{\oplus}\mathbf{gl}(m\infty)$ which is the direct sum of m copies of $\mathbf{gl}(\infty)$; let $\mathbf{gl}^a(\infty)$ denote the ath copy of $\mathbf{gl}(\infty)$ in the sum. The Chevalley generators of $\mathbf{gl}^a(\infty)$ can be taken as $e^a_i = E^{a,a}_{i,i+1}, f^a_i = E^{a,a}_{i+1,i}$. Observe that

$$\alpha^{a\vee}_i = \sum_{k=0}^{m-1} \alpha^\vee_{mj+a-1+k}, \quad a = 1,\dots,m,$$

where $\Pi^{a\vee} = \{\alpha^{a\vee}_j = [e^a_j, f^a_j]\}$. The subalgebra $\overset{m}{\oplus}\mathbf{gl}(\infty) \subset \mathbf{gl}(m\infty)$ has a natural gradation, the principal gradation, defined by

$$\deg e^a_1 = 1 = -\deg f^a_i.$$

Put $s^a(j) := \sum_{i \in Z} E^{a,a}_{i,i+j}$ then $\{s^a(j), z\}$ span a subalgebra of $\overset{m}{\oplus}\mathbf{gl}(\infty)$, the principal Heisenberg subalgebra, which has the induced principal gradation.

Proposition 2.1: (a) The Lie algebra $\overset{m}{\oplus}\mathbf{gl}(\infty)$ contains the principal Heisenberg subalgebra $\mathbf{s}_m = \overset{m}{\underset{a=1}{\oplus}} \bar{\mathbf{s}}^a \oplus \mathbb{C}z$ and $\mathbf{s}^a = \bar{\mathbf{s}}^a \oplus \mathbb{C}z$ can be identified with the principal Heisenberg subalgebra of $\mathbf{gl}^a(\infty)$. The algebra $\mathbf{s}_m$ is generated by $\{s^a(j), z: a = 1,\dots,m, j \in Z\}$,

$$[s^a(j), s^b(l)] = j\delta^{a,b}_{j,-l} z.$$

(b) The Lie algebra $\mathbf{gl}(m\infty)$ contains a Heisenberg subalgebra $\mathbf{s}_m$ which can be identified with the principal Heisenberg subalgebra of $\overset{m}{\oplus}\mathbf{gl}(\infty) \subset \mathbf{gl}(m\infty)$.

We call $\mathbf{s}_m \subset \mathbf{gl}(m\infty)$ the maximal Heisenberg subalgebra of $\mathbf{gl}(m\infty)$; its canonical projection onto $\overline{\mathbf{gl}}(m\infty)$ determines a Cartan subalgebra.

In the case of $\mathbf{a}(m\infty)$, $s^a(0) \notin \mathbf{a}(m\infty)$, however it does contain elements of the form,

$$a^c(0) = \sum_{b=1}^m \lambda^c_b s^b(0), \quad \sum_{b=1}^m \lambda^c_b = 0, \quad c = 1,\dots,m-1.$$

Fix a definite choice for $a^c(0)$, then the maximal Heisenberg subalgebra $\mathbf{s}_m \subset \mathbf{a}(m\infty)$ is generated by $\{s^b(j), a^c(0), z: j \in Z \setminus \{0\}\}$.

The principal gradation of $\overset{m}{\oplus}\mathbf{gl}(\infty)$ induces the gradation

$$\deg E^{a,b}_{r,s} = (s - r), \quad \deg z = 0,$$

of $\mathbf{gl}(m\infty)$ which we call the m-principal gradation of $\mathbf{gl}(m\infty)$. The associated generating functions are $\mathbf{gl}_m(\mathbf{u},\mathbf{v}) := (\mathbf{gl}^{a,b}(u_a, v_b))_{1 < a,b < m}$, where

$$\mathbf{gl}^{ab}(u_a, v_b) = \sum_{i,j \in z} E^{a,b}_{i,j} u^i_a v^{-j}_b + \delta^{a,b} \frac{u_a}{u_a - v_b} z. \tag{2.3}$$

The indices a, b are not summed in this expression. The function $u_a/(u_a - v_b)$ is always interpreted in the paper as the formal series $(1 - v_b/u_a)^{-1}$. This and the formal delta function which we also use, are defined by

$$\delta(k) = \sum_{i \in Z} k^i, \quad (1-k)^{-1} = \sum_{i \in Z_+} k^i.$$

Let

$$s_m(\mathbf{j}) = \sum_{a=1}^m s^a(j_a), \quad j_a \in Z,$$

$$w^{\mathbf{j}} = \mathrm{diag}((w_1)^{j_1},\dots,(w_m)^{j_m}),$$

where $s_m(\mathbf{j})\in \mathbf{s}_m \subset \mathbf{gl}(m\infty)$. The adjoint action of $s_m(\mathbf{j})$ on $\mathrm{gl}_m(\mathbf{u},\mathbf{v})$ is well defined,

$$[s_m(\mathbf{j}),\mathrm{gl}_m(\mathbf{u},\mathbf{v})] = u^{\mathbf{j}}\mathrm{gl}_m(\mathbf{u},\mathbf{v}) - \mathrm{gl}_m(\mathbf{u},\mathbf{v})v^{\mathbf{j}}, \quad j\in Z. \tag{2.4}$$

We define $\mathbf{gl}(m\infty)$ with the m-principal gradation as the m-principal realization of $\mathbf{gl}_\infty$. The generating functions satisfy the commutation relations

$$[\mathrm{gl}^{b,c}(p_1,p_2),\mathrm{gl}^{d,e}(q_1,q_2)] = \delta^{c,d}\delta(q_1/p_2)\mathrm{gl}^{b,e}(p_1,q_2) - \delta^{e,b} \times\delta(p_1/q_2)\mathrm{gl}^{d,c}(q_1,p_2). \tag{2.5}$$

III. *J* REDUCTIONS

The J reductions were introduced by the Japanese mathematicians Date, Jimbo, Kashiwara, and Miwa.[3,4] An automorphism is introduced that translates the basis of $\mathbf{gl}(m\infty)$. The fixed points of this automorphism define a finite rank affine Lie algebra $\mathbf{gl}_n^{(1)}$, or rather, a completion of it. The principal affect of this type of reduction on the Hirota equations associated with $\mathbf{gl}(m\infty)$ is to reduce the number of independent variables. There are other automorphisms of the algebra that affect the Hirota equations, but we do not consider them in this paper.

Define the automorphism $\rho^{\mathbf{n}}$, $\mathbf{n} = (n_1,...,n_m)\in Z_+^m$ by

$$\rho^{\mathbf{n}}: E_{r,s}^{a,b}\to E_{r+n_a,s+n_b}^{a,b}, \quad z\to z.$$

Let $\mathbf{gl}^{\mathbf{n}}(m\infty)$ denote the fixed point set of $\mathbf{gl}(m\infty)$ under $\rho^{\mathbf{n}}$ and introduce $\mathrm{Mat}_n(\mathbb{C})$, the algebra of $n\times n$ matrices over $\mathbb{C}$, where $n = l(\mathbf{n}) := \sum_{a=1}^m n_a$. The algebra $\overline{\mathbf{gl}}(m\infty)$ acts on V and through the correspondence $u^a_{n_as+j} \to v_j^a(k_a)^s := v_j^a\otimes(k_a)^s$, V can be identified with $\oplus_{a=1}^m \mathbb{C}^{n_a}\otimes\mathbb{C}[k_a,k_a^{-1}]$, where v_j^a is the jth standard basis element of $\mathbb{C}^{n_a}$. If $g = g^{a,b}_{r,s}E^{a,b}_{r,s}\in \overline{\mathbf{gl}}^{\mathbf{n}}(m\infty)$ put

$$g_{s-r} := (g^{a,b}_{n_as+i,n_br+j}F^{a,b}_{i,j})\in\mathrm{Mat}_n(\mathbb{C}),$$

$$F^{a,b}_{i,j} = (\delta_{pi}\delta_{jq})_{\substack{0\leqslant p\leqslant n_a-1,\\ 0\leqslant q\leqslant n_b-1}}$$

and it follows from the identification of vector spaces just introduced that without loss of generality we can put $k = k_a$, $a = 1,...,m$ and deduce that $\overline{\mathbf{gl}}^{\mathbf{n}}(m\infty)$ is isomorphic to a subalgebra of $\mathrm{Mat}_n(\mathbb{C}[k,k^{-1}])$ with the correspondence given by

$$g\leftrightarrow g(k) := \sum_{s\in Z} g_s\otimes k^{-s}\in\mathrm{Mat}_n(\mathbb{C}[k,k^{-1}]).$$

For $f,g\in\overline{\mathbf{gl}}^{\mathbf{n}}(m\infty)$ it is easy to show that

$$\Psi(f,g)\leftrightarrow\sum_{s\in Z} -s\,\mathrm{Tr}(f_sg_{-s}) = \mathrm{Res}\,\mathrm{Tr}\left\{\left(\frac{d}{dk}f(k)\right)g(k)\right\}.$$

The J reduction of $\mathbf{gl}^{\mathbf{n}}(m\infty)$ can therefore by realized as a completion of the finite rank affine Lie algebra

$$\mathbf{gl}_n^{(1)} := \mathbf{gl}(n)\otimes\mathbb{C}[k^{-1},k]\oplus\mathbb{C}z.$$

From the definition of $\mathbf{gl}(m\infty)$, elements belonging to the required completion of $\mathbf{gl}_n^{(1)}$ have the form

$$g(k) = \sum_{s\in Z_+\cup q} g_s(k)\otimes k^{-s} + \lambda z,$$

$$q\subset Z_-, \quad \mathrm{Card}\, q<\infty.$$

The bracket for $\mathbf{gl}_n^{(1)}$ is

$$[g_1(k)+\lambda_1z, g_2(k)+\lambda_2z] = [g_1(k),g_2(k)]_0 + \mathrm{Res}\,\mathrm{Tr}\left\{\frac{d}{dk}(g_1(k))g_2(k)\right\}. \tag{3.1}$$

The generating functions for $\overline{\mathbf{gl}}_n^{(1)}$ are obtained by the J reduction of the maximal Heisenberg subalgebra $\mathbf{s}_m$ and the generating functions $\overline{\mathbf{gl}}_m(\mathbf{u},\mathbf{v})$ for $\overline{\mathbf{gl}}(m\infty)$.

Lemma 3.1: Let $g\in\overline{\mathbf{gl}}(m\infty)$, $g\neq0$, and define $s_m(\mathbf{j}) = \sum_{a=1}^m s^a(j_a)$, $j_a\in Z$. Then $g\in\overline{\mathbf{gl}}^{\mathbf{n}}(m\infty)$ if and only if

$$[s_m(\mathbf{j}),g] = 0, \quad \mathbf{j}\equiv 0 \bmod \mathbf{n}.$$

The proof is straightforward. Applied to (1.4) the lemma shows that the corresponding generating functions for $\overline{\mathbf{gl}}^{\mathbf{n}}(m\infty)$ are obtained by imposing the restrictions

$$(u_a)^{n_a} = (v_b)^{n_b}, \quad a,b = 1,...,m.$$

The solutions to the equations are

$$u_a = \omega_a p^{\tilde{n}_a}, \quad v_a = \lambda_a p^{\tilde{n}_a},$$

where ω_a and λ_a are any n_ath roots of unity. The indeterminate p is normalized by putting $d := [n_1, n_2,...,n_m]$, the least common multiple of $\{n_1,...,n_m\}$ and defining $\tilde{n}_a := d/n_a$.

If we introduce the notation $\bar{i} = i \bmod n_a$, $[i] = i-\bar{i}$ then the J reductions of the generating functions $s^a(j)$, $\overline{\mathrm{gl}}^{b,c}_m(u_b,v_c)$ correspond to the following elements of $\mathbf{gl}_n^{(1)}$ or its completion

$$\overline{\mathrm{gl}}^{b,c}_m(\omega p^{\tilde{n}_b},\lambda p^{\tilde{n}_c}) \leftrightarrow\sum_{s\in z} p^{sd+(h\tilde{n}_b-l\tilde{n}_c)}(\omega^h\lambda^{-l})F^{b,c}_{h,l}\otimes k^{-s},$$

$$s^a(0)\leftrightarrow r^a(0) := \sum_{i=0}^{n_a-1} F^{a,a}_{i,i},$$

$$s^a(1)\leftrightarrow r^a(1) := \sum_{i=0}^{n_a-1} F^{a,a}_{i,\overline{(i+1)}}\otimes k^{[i+1]},$$

$$s^a(j)\leftrightarrow r^a(j) := (r^a(1))^j, \quad j\neq0. \tag{3.2}$$

The indices h and l are summed in (3.2) (but not b,c), $0\leqslant h\leqslant n_b-1$, $0\leqslant l\leqslant n_c-1$ and ω,λ are n_bth and n_cth roots of unity, respectively.

Under J reduction $\mathbf{gl}(m\infty)$ is equivalent to the complete Lie algebra $\mathbf{gl}_n^{(1)}$, $n = l(\mathbf{n})$ and the subalgebra $\overset{m}{\oplus}\mathbf{gl}(\infty)\subset\mathbf{gl}(m\infty)$ becomes the subalgebra $\oplus_{a=1}^m\mathbf{gl}_{n_a}^{(1)}\subset\mathbf{gl}_n^{(1)}$. The algebra $\mathbf{gl}_n^{(1)}$ contains the Heisenberg subalgebra

$\mathbf{s}_n$ generated by $\{r^a(j), z: j\in Z, a=1,...,m\}$. From (3.2) it is clear that $\mathbf{s}_n$ has the "root space" gradation since

$$\deg e_i^a = 0 = -\deg f_i^a,$$

$$\deg e_0^a = 1 = -\deg f_0^a, \quad 1\leqslant i\leqslant n_a-1, \quad a=1,...,m,$$

where

$$e_i^a = F^{a,a}_{\overline{(i-1)},i} \otimes k^{[i-1]} f_i^a = (e_i^a)^\dagger, \quad i=0,...,n_a-1.$$

However this gradation is not the one defined by the generating function in (3.2). In fact the gradation is fixed by the underlying finite-dimensional Lie algebra.

Let π: $\mathbf{gl}_n^{(1)} \to \mathbf{gl}(n)$ defined by $g\otimes k^s \to g$, $z\to 0$, $g\in\mathbf{gl}(n)$ denote the covering homomorphism. A theorem of Kac's[16] states that a class of dth-order automorphisms of $\mathbf{gl}(n)$ $(\mathbf{a}_{n-1})$ can be obtained up to conjugation in the following way. Put $d=\Sigma_{i=0}^{n-1} a_i s_i$ where for $\mathbf{gl}(n)$, $a_i=1$ and $s_i\in Z_+$. Let ω be a dth root of unity and define the automorphism σ of type $(s_0,...,s_{n-1};1)$ for $\mathbf{gl}(n)$ by

$$\sigma(e_i^0) = \omega^{s_i} e_i^0, \quad i=0,...,n-1,$$

where π: $e_i \to e_i^0$ and $\{e_i, f_i\}$ are the Chevalley generators of $\mathbf{gl}_n^{(1)}$. The associated Z_d gradation of $\mathbf{gl}(n)$, determined by the σ-eigenspace decomposition, is called the 1-s gradation of $\mathbf{gl}(n)$. Let $\mathbf{g}(j)$, $j\in Z_d$ be the eigenspace of σ with eigenvalue ω^j in this decomposition. The automorphism also defines a Z gradation, the $\mathbf{s}$ gradation of $\mathbf{gl}_n^{(1)}$,

$$\mathbf{gl}_n = \bigoplus_{j\in Z} \mathbf{g}_{(j \bmod Z_d)} \otimes k^j \oplus \mathbb{C}z, \quad \deg z = 0. \tag{3.3}$$

Consider now the situation that arises in the J reduction of $\mathbf{gl}(m\infty)$. Apply the covering homomorphism to the generating functions defined in (3.2). This fixes a Z_d gradation of $\mathbf{gl}(n_a)$ given by $\deg e_i^{0a} = \tilde{n}_a$, $i=0,...,n_a-1$, that is the $1-\tilde{n}_a 1_{n_a}$ gradation of $\mathbf{gl}(n_a)$, where $1_{n_a} := (1,...,1)\in Z^{n_a}$. Let $X_a := \mathrm{diag}(1,\omega^{\tilde{n}_a},...,\omega^{\tilde{n}_a(n_a-1)})$, where ω is the dth root of unity, then the $1-\tilde{n}_a 1_{n_a}$ gradation is determined by $\sigma_a(g) = X_a^{-1} g X_a$, $g\in\mathbf{gl}(n_a)$. The automorphisms $\{\sigma_a: a=1,...,m\}$ induce a Z_d gradation of $\mathbf{gl}(n)$ defined by $\sigma(g) = X^{-1}gX$, $g\in\mathbf{gl}(n)$, where $X := \mathrm{diag}(X_1,...,X_m)$. This Z_d gradation of $\mathbf{gl}(n)$ is induced by the $1-\tilde{\mathbf{n}}1_{\mathbf{n}}$ gradation of $\oplus_{a=1}^m \mathbf{gl}(n_a)$ where $\tilde{\mathbf{n}}1_{\mathbf{n}} := (\tilde{n}_1 1_{n_1},...,\tilde{n}_m 1_{n_m})$, and we shall call it the $1-\tilde{\mathbf{n}}1_{\mathbf{n}}$ gradation of $\mathbf{gl}(n)$. It is defined on basis elements by

$$\deg F_{h,l}^{a,b} = l\tilde{n}_b - h\tilde{n}_a \bmod d.$$

The corresponding Z gradation of $\mathbf{gl}_n^{(1)}$ defined by (3.2) will be referred to as the $\tilde{\mathbf{n}}1_{\mathbf{n}}$ gradation of $\mathbf{gl}_n^{(1)}$. The generating functions and the generators of the Heisenberg subalgebra $\mathbf{s}_n \subset \mathbf{gl}_n^{(1)}$ with the $\tilde{\mathbf{n}}1_{\mathbf{n}}$ gradation are

$$\overline{\mathbf{gl}}_{\mathbf{n}}^{b,c}(\omega^i p, \omega^j p) = \sum_{s\in z} p^{sd+h\tilde{n}_b - l\tilde{n}_c}\omega^{hi\tilde{n}_b - lj\tilde{n}_c} \times F_{h,l}^{b,c} \otimes k^{-(sd+h\tilde{n}_b - l\tilde{n}_c)},$$

$$r^a(0) = \sum_{s=0}^{n_a-1} F_{s,s}^{a,a}, \quad r^a(1) = \sum_{s=0}^{n_a-1} F^{a,a}_{\overline{(s-1)},s} \otimes k^{\tilde{n}_a},$$

$$r^a(s) = (r^a(1))^s, \quad s\in Z\setminus\{0\}. \tag{3.4}$$

If $n_a=1$ then $\overline{\mathbf{gl}}_{\mathbf{n}}^{a,a}(\omega^i p, \omega^j p)$ is excluded from the list of generating functions.

It is appropriate to rescale the central extension in (3.1),

$$[g_1(k), g_2(k)] = [g_1(k), g_2(k)]_0 + \frac{1}{d}\,\mathrm{Res}\,\mathrm{Tr}\left\{\left(\frac{d}{dk} g_1(k)\right) g_2(k)\right\} z,$$

so that the defining relations for the Heisenberg subalgebra are canonical,

$$[r^a(i), r^b(j)] = i\delta_{i,-j}^{a,b} z, \quad i,j\in Z.$$

The subalgebra $\mathbf{s}_n$ is a maximal Heisenberg subalgebra of $\mathbf{gl}_n^{(1)}$.

A set of generators for $\mathbf{gl}(n)$ is given by

$$e_0^0 = F^{m,1}_{n_m-1,0}, \quad e^0_{p_{a-1}+n_a} = F^{a,a+1}_{n_a-1,0}, \quad 1\leqslant a\leqslant m-1,$$

$$e^0_{p_{a-1}+i} = F^{a,a}_{i-1,i}, \quad 1\leqslant a\leqslant m, \quad 1\leqslant i\leqslant n_a-1,$$

where $p_a = n_1+n_2+\cdots+n_a$ and $p_0=0$. However the associated $\mathbf{s}$ gradation of $\mathbf{gl}_n^{(1)}$ is not the one required since $l(\mathbf{s}) = md$ in this case and the result is only correct for $m=1$.

The determination of the Chevalley generators of $\mathbf{gl}_n^{(1)}$ with the $\tilde{\mathbf{n}}1_{\mathbf{n}}$ gradation seems difficult except in the case when $n_a = N$, $a=1,...,m$. The Chevalley generators are then given by the J reduction of the Chevalley generators for $\mathbf{gl}(m\infty)$,

$$\begin{aligned} e_{m(N-1)} &= F^{m,1}_{N-1,0} \otimes k, \quad e_{m(r-1)+m} = F^{m,1}_{r,r+1} \otimes k, \\ &\qquad 0\leqslant r\leqslant N-2, \\ e_{m(r-1)+a} &= F^{a,a+1}_{r,r}, \quad a=1,...,m-1, \\ &\qquad 1\leqslant r\leqslant N-1. \end{aligned} \tag{3.5}$$

Then $\{e_i, f_i\}$, $f_i := e_i^\dagger$ are the required set. The difficulty attached to obtaining the generators in the general case is easily seen by studying the J reduction of $\mathbf{gl}(3\infty)$ with $\mathbf{n} = (5,2,4)$.

In the case of $\mathbf{a}(m\infty)$ the J reduction gives $\mathbf{a}_{n-1}^{(1)}$ with the $\tilde{\mathbf{n}}1_{\mathbf{n}}$ gradation. The generating functions are $\overline{\mathbf{gl}}_{\mathbf{n}}^{a,b}$ $(\omega^i p, \omega^j p)$ with the exclusion of $\overline{\mathbf{gl}}_{\mathbf{n}}^{a,a}$ $(\omega^i p, \omega^j p)$ if $n_a=1$. However the definition of the maximal Heisenberg subalgebra requires more care. Under J reduction we obtain $m-1$ linearly independent elements

$$a^b(0) \leftrightarrow h^b(0) := \sum_{c=1}^m \lambda_c^b r^c(0), \quad b=1,...,m-1.$$

Introduce the vectors $\boldsymbol{\lambda}^b := (\lambda_1^b,...,\lambda_m^b)$, $\boldsymbol{\gamma}^b := (n_1\lambda_1^b,...,n_m\lambda_m^b)$ then the linearly independent elements $h^b(0)$ are chosen so that

$$\boldsymbol{\lambda}^a\cdot\boldsymbol{\gamma}^b = \delta^{a,b}, \quad \boldsymbol{\lambda}^b\cdot\mathbf{n} = 0, \quad a,b=1,...,m-1.$$

Define $\mathbf{h}_n$ as the Heisenberg subalgebra of $\mathbf{a}_{n-1}^{(1)}$ generated by $\{h^a(j), z: j\in Z, a=1,...,m-1\}$, where $h^a(j) := h^a(0)\otimes k^{jd}$. Observe that if $n_a=1$ then $r^a(1) = F_{0,0}^{a,a}\otimes k^d \notin \mathbf{a}_{n-1}^{(1)}$. For $n_a\neq 1$ define $\mathbf{s}_{n_a}$ as the Heisenberg subalgebra of $\mathbf{a}_{n-1}^{(1)}$ generated by $\{r^a(j), z: j\in Z, j\not\equiv 0 \bmod n_a\}$. Then put

$$\mathbf{s_n} = \bigoplus_{a:n_a\neq 1} \mathbf{s}_{n_a} \oplus \mathbf{h_n}, \tag{3.6}$$

and $\mathbf{s_n}$ in (3.6) is a maximal Heisenberg subalgebra of $\mathbf{a}^{(1)}_{n-1}$.

It remains to specify the minimum number of generating functions for $\mathbf{gl}^{(1)}_n$ and $\mathbf{a}^{(1)}_{n-1}$. First we observe that $\overline{\mathbf{gl}}^{b,c}_{\mathbf{n}}(\omega^i p, \omega^j p)^\dagger = \overline{\mathbf{gl}}^{c,b}_{\mathbf{n}}(\omega^j p, \omega^i p)$ $b\neq c$ if $p^\dagger := p^{-1}$, $k^\dagger := k^{-1}$. An examination of (3.4) shows that in order to solve the problem we need only determine the number of elements from $\pi\{\overline{\mathbf{gl}}^{b,c}_{\mathbf{n}}(\omega^i p, \omega^j p)\mathbf{s_n}\}$ necessary to form a basis for $\mathbf{gl}(n)$ or $\mathbf{a}_{n-1}$. The result is given in the proposition below.

Proposition 3.2: (a) A basis for $\mathbf{gl}^{(1)}_n$, $n = l(\mathbf{n})$ with the $\tilde{\mathbf{n}}\mathbf{1_n}$ gradation is provided by the homogeneous components of the generating functions

$$\overline{\mathbf{gl}}^{c,c}_{\mathbf{n}}(\omega^i p,p), \quad i = 1,\ldots,n_c - 1, \quad c\neq b \text{ if } n_b = 1,$$

$$\overline{\mathbf{gl}}^{b,c}_{\mathbf{n}}(\omega^i p,p), \quad i = 1,\ldots,f_{bc}, \quad b<c \quad c = 1,\ldots,m,$$

and $\overline{\mathbf{gl}}^{b,c\dagger}_{\mathbf{n}}(\omega^i p,p)$ where ω is a nontrivial dth root of unity, $d = [n_1,\ldots,n_m]$ and f_{bc} is the greatest common division of $\{n_b,n_c\}$ together with the generators of the maximal Heisenberg subalgebra $\mathbf{s_n}$, $\{r^a(j), z: j\in Z, a = 1,\ldots,m\}$.

(b) A basis for $\mathbf{a}^{(1)}_{n-1}$ is provided by the generating functions $\overline{\mathbf{gl}}^{c,c}_{\mathbf{n}}(\omega^i p,p)$, $\overline{\mathbf{gl}}^{b,c}_{\mathbf{n}}(\omega^i p,p)$, $\overline{\mathbf{gl}}^{\dagger b,c}_{\mathbf{n}}(\omega^i p,p)$ listed in (a) together with the generators of the maximal Heisenberg subalgebra of $\mathbf{a}^{(1)}_{n-1}$, $\mathbf{s_n} = \bigoplus_{a:n_a\neq 1} \mathbf{s}_{n_a} \oplus \mathbf{h_n}$.

The generating function $\overline{\mathbf{gl}}^{a,b}_{\mathbf{n}}(\omega^i p,p)$ appears here rather than $\mathbf{gl}^{a,b}_{\mathbf{n}}(\omega^i p,p)$ because it satisfies the important relation obtained by the J reduction of (2.4),

$$[r^a(j), \overline{\mathbf{gl}}^{b,c}_{\mathbf{n}}(\omega p,p)] = (\omega^{j\tilde{n}_a}\delta^{a,b} - \delta^{a,c})p^{j\tilde{n}_a} \times \overline{\mathbf{gl}}^{b,c}_{\mathbf{n}}(\omega p,p). \tag{3.7}$$

This is the relationship which underpins the vertex representations of the next two sections.

The reduction given in this section is not unique since it depends on the enumeration of the basis $\{F^{a,b}_{h,l}\}$ (e.g., $1\leqslant h\leqslant n_a$, $1\leqslant l\leqslant n_b$ gives a different realization).

IV. FUNDAMENTAL m-PRINCIPAL VERTEX REPRESENTATIONS OF $\mathbf{gl}_\infty$ AND THEIR J REDUCTIONS

The m-principal realization of $\mathbf{gl}_\infty$ has a vertex representation on a space which can be formally identified with the representation space for the Heisenberg system associated with $\mathbf{a}^{(1)}_m$ introduced by Frenkel and Kac.[11] It decomposes into irreducible representations of the fundamental modules for $\mathbf{gl}_\infty$. The lattice associated with this representation is not a root lattice. An interpretation of it is given in the next section.

The representation is really the same as arises in Ref. 7, apart from the implied definition of the cocycle, which is obtained by reversing the inequalities in the definition overleaf. A similar approach is used in Ref. 17 for the representations of $\mathbf{gl}^{(1)}_n$ defined by three different maximal Heisenberg subalgebras.

In Ref. (8), Lepowsky and Wilson show that with every maximal Heisenberg subalgebra of $\mathbf{a}^{(1)}_{n-1}$ it is possible to associate a representation of a fundamental module. A maximal Heisenberg subalgebra of $\mathbf{a}^{(1)}_{n-1}$ is defined by an element of the Weyl group W of the associated finite-dimensional Lie algebra $\mathbf{a}_{n-1}$. Thus the homogeneous Heisenberg subalgebra corresponds to the identity element for example. Kac and Peterson[10] show that a complete nonredundant list of Heisenberg subalgebras of $\mathbf{a}^{(1)}_{n-1}$ is afforded by a set of representatives of the conjugacy classes of W. For $\mathbf{a}_{n-1}$ the number of conjugacy classes is $p(n)$, where p is the classical partition function. The main theorem of their paper can then be used to construct a representation of a fundamental module for each conjugacy class. We shall term these the canonical vertex representations of a fundamental module of $\mathbf{a}^{(1)}_{n-1}$.

In this section we obtain explicit formulas for some of the vertex representations of $\mathbf{a}^{(1)}_{n-1}$ by using J reductions. It is clear from the work of Kac and Peterson that we need only show that we obtain $p(n)$ distinct maximal Heisenberg subalgebras of $\mathbf{a}^{(1)}_{n-1}$ by this process in order to establish that in principle a canonical set of vertex representations might be obtained in this way.

Lemma 4.1: (a) Let $\{\mathbf{gl}^{\mathbf{q}}(m\infty): l(\mathbf{q}) = n, 1\leqslant m < \infty\}$ be the set of J reductions that define realizations of $\mathbf{gl}^{(1)}_n$. Then there are $p(n)$ distinct realizations of $\mathbf{gl}^{(1)}_n$ defined by distinct maximal Heisenberg subalgebras $\mathbf{s_q}\subset\mathbf{gl}^{(1)}_n$.

(b) Let $\{\mathbf{a}^{\mathbf{q}}(m\infty): l(\mathbf{q}) = n, 1\leqslant m<\infty\}$ be the set of J reductions that define realizations of $\mathbf{a}^{(1)}_{n-1}$. Then there are $p(n)$ distinct realizations of $\mathbf{a}^{(1)}_{n-1}$ defined by distinct maximal Heisenberg subalgebras $\mathbf{s_q}\subset\mathbf{a}^{(1)}_{n-1}$.

Proof: (a) Consider the set of vectors $Q = \{\mathbf{q}: \mathbf{q}\in Z^m_+, l(\mathbf{q}) = n, 1\leqslant m\leqslant n\}$. There are $p(n)$ distinct partitions of n and to each such partition we can associate a vector $\mathbf{q}\in Q$ and a J reduction of $\mathbf{gl}(m\infty)$ for some m. Thus if $\Sigma^m_{a=1} r_a = n$ is one such partition then $\mathbf{q} = (r_1,\ldots,r_m)$ defines the required J reduction of $\mathbf{gl}(m\infty)$. The corresponding maximal Heisenberg subalgebra of $\mathbf{gl}^{(1)}_n$ is $\mathbf{s_q} = \oplus^m_{a=1}\mathbf{s}_{n_a}$. The ordering is arbitrary since if σ is a permutation of $(1,\ldots,m)$, then σ: $F^{a,b}_{i,j}\otimes k^r \to F^{\sigma(a),\sigma(b)}_{i,j}\otimes k^r$, $z\to z$ is an automorphism of $\mathbf{gl}^{(1)}_n$. The argument for (b) is similar.

Theorem 4.2: A canonical set of realizations of $\mathbf{gl}^{(1)}_n$ ($\mathbf{a}^{(1)}_{n-1}$) can be obtained from the J reductions of $\{\mathbf{gl}(m\infty): 1\leqslant m\leqslant n\}$.

Let $\mathbf{s}_m$ denote the maximal Heisenberg subalgebra of $\mathbf{gl}(m\infty)$. Let $\mathbf{c}$ be the vector space spanned by $\{s^a\}$, $s^a := s^a(0)$, and $\{s_a\}$ the dual basis of $\mathbf{c}^*$, $(s_a, s^b) = \delta_{a,b}$. Introduce the Z lattice associated with $\mathbf{c}^*$, $M = \oplus^m_{a=1} Zs_a$ and let $\mathbb{C}[M]$ denote the group algebra of M. Put $\mathbf{s}_- = \oplus^m_{a=1}\mathbf{s}^a_-$, where $\mathbf{s}^a_\pm = \oplus_{\pm i>0}\mathbb{C}s^a(i)$ and let $S = \mathrm{Sym}(\mathbf{s}_-)$ denote the symmetric algebra on $\mathbf{s}_-$. S has the induced m- principal gradation and can be identified with the ring of polynomials $\mathbb{C}[x] := \mathbb{C}[x^{(1)},\ldots, x^{(m)}]$, where $x^{(a)} := (x^{(a)}_1, x^{(a)}_2,\ldots)$ with $\deg x^{(a)}_j = -j$.

The Frenkel–Kac cocycle $\epsilon: M\times M\to\{\pm 1\}$ is uniquely defined by

$$\epsilon(\alpha+\beta,\gamma) = \epsilon(\alpha,\gamma)\epsilon(\beta,\gamma),$$

$$\epsilon(\alpha,\beta+\gamma) = \epsilon(\alpha,\beta)\epsilon(\alpha,\gamma),$$

$$\epsilon(s_a,s_b)=\begin{cases}1, & a\leqslant b\\ -1, & a>b.\end{cases}$$

The irreducible representation of the Heisenberg system $\{\mathbf{s}_m,M\}$ on $U=\mathbb{C}[x]\otimes\mathbb{C}[M]$ is given by

$$\begin{aligned}
&s^a(l)\cdot w\otimes e^\alpha=\partial_{x_l^{(a)}}w\otimes e^\alpha,\\
&s^a(-l)\cdot w\otimes e^\alpha=lx_l^{(a)}w\otimes e^\alpha,\quad l>0,\\
&s^a(0)\cdot w\otimes e^\alpha=(\alpha,s^a)w\otimes e^\alpha,\\
&z\cdot w\otimes e^\alpha=w\otimes e^\alpha,\\
&e^\beta\cdot w\otimes e^\alpha=\epsilon(\beta,\alpha)w\otimes e^{\alpha+\beta},
\end{aligned}\tag{4.1}$$

where $\alpha,\beta\in M$, $w\otimes e^\alpha\in U$. The representation of $\mathbf{gl}(m\infty)$ on U is defined in terms of the vertex operator

$$X(\pm s_a,k)=\exp(\pm k\cdot x^{(a)})\exp(\mp k^{-1}\cdot D_{x^{(a)}})T_{\pm s_a}(k),\tag{4.2}$$

where

$$k\cdot x^{(a)}:=\sum_{i\geqslant 1}k^ix_i^{(a)}$$

and

$$D_{x^{(a)}}:=\left(\partial_{x_1^{(a)}},\ldots,\frac{1}{r}\partial_{x_r^{(a)}},\ldots\right).$$

The first two factors of the vertex operator act on $\mathbb{C}[x]$ whereas the last factor

$$T_\gamma(k)=\exp(\ln k\cdot\gamma(0)+\gamma),\quad \gamma\in M,$$

acts on $\mathbb{C}[M]$,

$$T_\gamma(k)\cdot e^\alpha=k^{(\gamma,\alpha)+(1/2)(\gamma,\gamma)}\epsilon(\gamma,\alpha)e^{\gamma+\alpha}.\tag{4.3}$$

The vertex operator maps U into its completion and has the formal expansion

$$X(\pm s_a,k)=\sum_{l\in Z+1/2}X_l(\pm s_a)k^l.$$

Then, as is well known,[17,18] the anticommutators of the components define a Clifford algebra,

$$[X_l(s_a),\,X_j(\pm s_b)]_+=\tfrac{1}{2}(1\mp 1)\delta^{a,b}_{i,-j}.\tag{4.4}$$

It follows that if we put

$$e^{a,b}_{i,j}:=X_{i-1/2}(s_a)X_{-(j-1/2)}(-s_b)-\delta^{a,b}_{ij}\theta(i),\tag{4.5}$$

then $\{e^{a,b}_{i,j},z\}$ generates a Lie algebra that is isomorphic to $\mathbf{gl}(m\infty)$ with the correspondence given explicitly by

$$E^{a,b}_{i,j}\leftrightarrow e^{a,b}_{i,j},\quad z\leftrightarrow 1.\tag{4.6}$$

The vertex operator corresponding to $\mathbf{gl}_m(\mathbf{u},\mathbf{v})$ can now be explicitly constructed. It is necessary to normally order the product of vertex operators that arises in this process so that it always makes sense. We then find that

$$:X(s_b,u)X(-s_c,v):=\epsilon(s_b,s_c)(u-v)^{-(s_b,s_c)}\times(uv)^{(1/2)(s_b,s_c)}X(s_b,-s_c,u,v),$$

where

$$\begin{aligned}
X(s_b,-s_c,u,v)&=\exp(u\cdot x^{(b)}-v\cdot x^{(c)})\exp[-(u^{-1}D_{x^{(b)}}\\
&\quad -v^{-1}D_{x^{(c)}})]S_{s_b,-s_c}(u,v),
\end{aligned}$$

and

$$S_{\gamma,u}(u,v)=\exp(\ln u\cdot\gamma(0)+\ln v\cdot u(0)+\gamma+u).\tag{4.7}$$

Let $\overline{\mathbf{gl}}_m(\mathbf{u},\mathbf{v})$ denote the restriction of $\mathbf{gl}_m(\mathbf{u},\mathbf{v})$ to the completion of $\overline{\mathbf{gl}}(m\infty)$. Then a representation (U,π^m) of $\mathbf{gl}(m\infty)$ is defined by the action of the Heisenberg subalgebra $\mathbf{s}_m$ given by (4.1) and the homogeneous components of the generating functions

$$\pi^m:\ \overline{\mathbf{gl}}^{b,c}_m(u,v)\rightarrow X^{b,c}(u,v),$$

where

$$X^{b,c}(u,v)=\epsilon(s_b,s_c)\begin{cases}[u/(u-v)](X(s_b,-s_c,u,v)-1), & b=c,\\ u^{1/2}v^{-1/2}X(s_b,-s_c,u,v), & b\neq c.\end{cases}\tag{4.8}$$

Let A denote the Clifford algebra generated by $\{X_l(\pm s_a)\}$ and observe that $c_0=1\otimes e^0$ satisfies

$$X_l(\pm s_a)\cdot c_0=0,\quad l<0.$$

Consequently c_0 is a vacuum vector of an A-module C generated by $\{X_l(\pm s_a)\}$ and this plays the central role in the representation theory of $\mathbf{gl}(m\infty)$ developed by Date, Jimbo, Kashiwara, and Miwa.[3–7]

The spaces C and U can be identified. Put $\deg e^\alpha=-\frac{1}{2}(\alpha,\alpha)$ and let $V_{k,\mu}=\{v\in V\colon s(0)\cdot v=(\mu,s)v,\ d\cdot v=-kv\}$, where V is C or U and d is the degree operator. A character for V is defined by

$$\operatorname{ch}V=\sum_{k\in(1/2)Z_+}\sum_{\mu\in M}V_{k,\mu}e^\mu q^k,$$

which gives in the two cases,

$$\begin{aligned}
\operatorname{ch}U&=\prod_{j=1}^{\infty}(1-q^j)^{-m}\sum_{\mu\in M}e^\mu q^{(1/2)(\mu,\mu)},\\
\operatorname{ch}C&=\prod_{a=1}^{m}\prod_{l\in Z+1/2}(1+q^le^{s_a})(1+q^le^{-s_a}).
\end{aligned}\tag{4.9}$$

The identification follows from the Jacobi identity for θ functions.

A unique contravariant Hermitian form can be defined on U. If $u=p_1,\ldots,p_m$, $w=q_1,\ldots,q_m$, $p_a,q_a\in\mathbb{C}[x^{(a)}]$ put

$$H_U(u\otimes e^\alpha,w\otimes e^\beta)=H_{\mathbb{C}[x]}(u,w)\langle e^\alpha,e^\beta\rangle$$

$$\langle e^\alpha,e^\beta\rangle=\delta_{\alpha,\beta},$$

$$H_{\mathbb{C}[x]}(u,w)=\prod_{a=1}^{m}H_{\mathbb{C}[x^{(a)}]}(p_a,q_a).$$

For $p,q\in\mathbb{C}[x^{(a)}]$ define

$$H_{\mathbb{C}[x^{(a)}]}(p,q)=p(D_{x^{(a)}})q^*(x^{(a)})|_{x^{(a)}=0}.\tag{4.10}$$

The form $H_U(\cdot,\cdot)$ is made unique by the normalization $H(1,1)=1$. The unique contravariant form on C is defined by

$$H_C(c_0,c_0) = 1,$$
$$H_C(X_i(s_a)c_0,\ X_j(s_b)c_0) = \delta_{i,j}^{a,b}\theta(1-i), \qquad (4.11)$$
$$H_C(X_i(-s_a)c_0,\ X_j(-s_b)c_0) = \delta_{i,j}^{a,b}\theta(1-i),$$

with all other brackets between elements $\{X_i(\pm s_a)c_0\}$ zero. Existence follows from the fact that the two forms agree if the completions of V and C are considered. Define the Schur polynomials $p_n(y)$, $y=(y_1,y_2,\ldots)$ by

$$\sum_{j>0} p_j(y)k^i = \exp \sum_{r>1} y_r k^r. \qquad (4.12)$$

The polynomials can be constructed from the determinantal expression,

$$n!p_n(y) = \begin{bmatrix} y_1 & 1 & & \\ 2y_2 & y_1 & -2 & \\ \vdots & \vdots & & \\ (n-1)y_{n-1} & (n-2)y_{n-2} & \cdots & -n+1 \\ ny_n & (n-1)y_{n-1} & \cdots & y_1 \end{bmatrix}. \qquad (4.13)$$

Then we have

$$X(s_a,k)\cdot c_0 = k^{1/2}X^a(1\otimes e^{s_a}),$$

where

$$X^a = \sum_{i>1} X_i^a k^i, \quad X_i^a = \sum_{j>0} p_{i+1}(x^{(a)})p_j(-D_{x^{(a)}}).$$

From the definition of (4.12) $\partial_{y_j} p_n(y) = p_{n-j}(y)$ and (4.13) gives

$$H_{\mathbb{C}[y]}(p_i(y),p_j(y)) = \delta_{i,j}p_i(1,\tfrac{1}{2},\tfrac{1}{3},\ldots) = \delta_{i,j}.$$

Consequently,

$$H_U(X_{i-1/2}(s_a)\cdot 1,\quad X_{j-1/2}(s_b)\cdot 1) = H_{\mathbb{C}[x]}(p_{i-1}(x^{(a)}),\ p_{j-1}(x^{(b)}))\langle e^{s_a},\ e^{s_b}\rangle = \delta_{i,j}^{a,b}\theta(1-i),$$

and the other relationships are established similarly.

The contravariant property of $H_U(\cdot,\cdot)$ can be established directly from its definition. Thus we find that $s^a(j)^\dagger = s^a(-j)$ and $(e^{s_a})^\dagger = e^{-s_a}$. It follows that $X_j(s_a)^\dagger = X_{-j}(-s_a)$ and that the adjoint of $g\in\mathbf{gl}(m\infty)$ with respect to $H_U(\cdot,\cdot)$ is given by $g^\dagger = (g^t)^*(E_{i,j}^{a,bt} = E_{j,i}^{b,a})$. If the formal variable k is constrained to the unit circle then

$$X(s_a,k)^\dagger = \sum_{j\in Z+1/2} X_j^+(s_a)k^{-j}.$$

The free fermions in Refs. 3 and 7 can be identified with $\psi_{i-1}^a \approx X_{i-1/2}(s_a)$, $\psi_{i-1}^{a*} \approx X_{i-1/2}^\dagger(s_a)$.

The representation $(U,\ \pi^m)$ is reducible. Consider the action of the algebra $\mathbf{Z}_U$ generated by the homogeneous components of

$$Z_{s_a,-s_b}(u,v) := \epsilon(s_a,s_b)(u/v)^{(1/4)(s_a-s_b,s_a-s_b)}(u/(u-v))^{(s_a,s_b)}S_{s_a,-s_b}(u,v),$$

$$Z_{s_a,-s_b}(u,v)\cdot e^\alpha = \epsilon(s_a,s_b)\begin{cases} u^{(s_a,\alpha)+1}v^{-(s_b,\alpha)}\epsilon(s_a-s_b,\alpha)e^{\alpha+s_a-s_b}, & a\neq b,\\ u^{(s_a,\alpha)}v^{-(s_a,\alpha)}(1-v/u)^{-1}e^\alpha, & a=b.\end{cases}$$

The vacuum space of $\mathbf{s}_+ = \oplus_{a=1}^m \mathbf{s}_+^a$, $\Omega_U := \mathbb{C}[M]$ decomposes into irreducible $\mathbf{Z}_U$ subspaces $\Omega_U = \oplus_{i\in Z}\Omega_U^i$, where Ω_U^i is spanned by $\{e^\alpha\in\mathbb{C}[M]:(\alpha,I)=i, I=\Sigma_{a=1}^m s^a\}$. It follows that $U_i := \mathbb{C}[x]\otimes\Omega_U^i$ is an irreducible $\mathbf{gl}(m\infty)$ module because Ω_U^i is $\mathbf{Z}_U$-irreducible.[18]

Since z acts as the identity operator these are necessarily level 1 $\mathbf{gl}_\infty$ modules and the Dynkin diagram of $\mathbf{gl}_\infty$ shows that they can be identified with the fundamental modules. These are the irreducible highest weight modules generated by the action of the universal enveloping algebra of $\mathbf{gl}(m\infty)$ on a highest weight vector v_Λ such that

$$h\cdot v_\Lambda = \langle\Lambda,h\rangle v_\Lambda,\quad \mathbf{n}_+\cdot v_\Lambda = 0,\quad h\in\mathbf{h},$$

where $\mathbf{n}_+$ is spanned by the Chevalley generators $\{e_i\}$. The pairing is the standard one $\langle\alpha_i,\ \alpha_j^\vee\rangle = a_{ij}$, where α_i is the root corresponding to the coroot $a_i^\vee$ and (a_{ij}) is the Cartan matrix.[16] The fundamental modules are defined by $\Lambda_i\in\mathbf{h}^*$ such that $\langle\Lambda_i,\ \alpha_j^\vee\rangle = \delta_{i,j}$, $i,j\in Z$.

Let $\Sigma_{i,j\in Z}\widetilde{E}_{i,j}^{b,c}u^iv^{-j} = \mathrm{gl}^{b,c}(u,v)$, where $\mathrm{gl}^{b,c}(u,v)$, is defined in (2.3). Define the action of $\widetilde{E}_{i,j}^{b,c}$ by

$$\tilde{E}^{b,b}_{j,j}\cdot v_{\Lambda_\mu} = \begin{cases} \left.\begin{matrix} 0, & j\leqslant i \\ -v_{\Lambda_p}, & j>i \end{matrix}\right\} b\leqslant a, \\ \left.\begin{matrix} 0, & j\leqslant i-1 \\ -v_{\Lambda_p}, & j>i-1 \end{matrix}\right\} b>a, \end{cases} \tag{4.14}$$

where $p = m(i-1)+a$. Then we find that

$$a_j^\vee\cdot v_{\Lambda_p} = \delta_{j,p} v_{\Lambda_p}.$$

Consequently we can identify the $\mathbf{gl}(m\infty)$ module U_p with the fundamental $\mathbf{gl}_\infty$ module for Λ_p. In fact $v_{\Lambda_p} = 1\otimes e^{\Lambda_p}$, where $\Lambda_p \in \mathbb{C}[M]$ is a particular element such that $(\Lambda_p, I) = p$ which we now calculate.

$$\left(\sum_{j\in Z} X_{j-1/2}(s_b)X_{-(j-1/2)}(-s_b u^j v^{-j})\right)(1\otimes e^{\Lambda_p})$$
$$= (u/v)^{(\Lambda_p,s_b)}(u/(u-v))(1\otimes e^{\Lambda_p}).$$

The corresponding relation from (4.14) is

$$\left(\sum_{j\in Z}\tilde{E}^{b,b}_{i,j} u^j v^{-j}\right)\cdot v_{\Lambda_p} = \begin{cases} (u/v)^i(u/(u-v))v_{\Lambda_p}, & b\leqslant a, \\ (u/v)^{i-1}(u/(u-v))v_{\Lambda_p}, & b>a, \end{cases}$$

so that $\Lambda_p\in\mathbb{C}[M]$ is given by

$$\Lambda_p = i\sum_{b=1}^{a} s_b + (i-1)\sum_{b=a+1}^{m} s_b. \tag{4.15}$$

With this expression for Λ_p it is straightforward to check that $\mathbf{n}_+\cdot v_{\Lambda_p} = 0$.

Proposition 4.3: The infinite rank affine Lie algebra $\mathbf{gl}_\infty$ has a representation (U, π^m) on $U = \mathbb{C}[x]\otimes\mathbb{C}[M]$ called the m-principal vertex representation given by the action of the Heisenberg subalgebra $\mathbf{s}_m$, (4.1) and the homogeneous components of the generating functions $X^{b,c}(u_b,u_c)$, (4.8). The representation decomposes into a direct sum of irreducible representations (U_p, π^m_p) $p\in Z$ which can be identified with the fundamental modules of $\mathbf{gl}_\infty$.

Since we have irreducible representations of $\mathbf{gl}(m\infty)$ we can use the J reductions of the previous section to obtain vertex representations of $\mathbf{gl}_n^{(1)}$.

Let ω be a dth root of unity where $d = [n_1,\dots,n_m]$ corresponds to the J reduction of $\mathbf{gl}(m\infty)$ defined by $\mathbf{n} = (n_1,\dots,n_m)$ and put $\tilde{n}_a = d/n_a$. The maximal Heisenberg subalgebra $\mathbf{s}_\mathbf{n}\subset\mathbf{gl}_n^{(1)}$ acts as in (4.1) on the representation space $U_{\rm gl} := \mathbb{C}[x]\otimes\mathbb{C}[M]$, where M can be identified with the previous lattice $(r_a \approx s_a)$. Thus the spaces U and $U_{\rm gl}$ can be identified as sets, but $U_{\rm gl}$ has the $\tilde{\mathbf{n}}\mathbf{1}_\mathbf{n}$ gradation.

$$\begin{aligned} &\deg(w(x)\otimes e^\alpha) = \deg w(x) - \tfrac{1}{2}\{\alpha,\alpha\}, \\ &\deg x_s^{(a)} = -s\,\tilde{n}_a, \quad \{r_a,r_b\} = \tilde{n}_a\delta_{a,b}. \end{aligned} \tag{4.16}$$

In this case the character formula gives

$$\mathrm{ch}\, U_{\rm gl} = \prod_{a=1}^{m}\prod_{j\geqslant 1}(1-q^{j\tilde{n}_a})^{-1}\sum_{\mu\in M} e^\mu q^{(1/2)\{\mu,\mu\}}. \tag{4.17}$$

It appears that the corresponding vertex operators for $\mathbf{gl}_n^{(1)}$ with the $\tilde{\mathbf{n}}\mathbf{1}_\mathbf{n}$ gradation should, in principle, be obtained from Lemma 3.1, (3.4) and Proposition 3.2(a),

$$\begin{aligned} &\overline{\mathbf{gl}}^{b,c}_\mathbf{n}(\omega p,p)\to X^{b,c}_{\rm gl}(\omega p,p) := \bar{X}^{b,c}((\omega p)^{\tilde{n}}b, p^{\tilde{n}}c), \\ &X^{b,c}_{\rm gl}(\omega p,\lambda p) = \exp(E^{b,c}_+(\omega p,\lambda p))\exp(-E^{b,c}_-(\omega p,\lambda p)) \\ &\qquad\times Z^{b,c}(\omega p,\lambda p). \end{aligned} \tag{4.18}$$

The explicit form of $E^{b,c}_\pm(\omega p,\lambda p)$ and $Z^{b,c}(\omega p,\lambda p)$ are easily obtained. The vertex operators (4.18) satisfy the relation (3.7),

$$\begin{aligned} [r^a(j),X^{b,c}_{\rm gl}(\omega p,\lambda p)] &= (\omega^{j\tilde{n}_a}\delta^{a,b} - \lambda^{j\tilde{n}_a}\delta^{a,c}) \\ &\quad\times p^{j\tilde{n}_a}X^{b,c}_{\rm gl}(\omega p,\lambda p), \end{aligned} \tag{4.19}$$

where $r^a(j)$ acts according to (4.1), $j\neq 0$. Then by a standard argument for a $\mathbf{gl}_n^{(1)}$ module $U_{\rm gl}$, $U_{\rm gl} = \mathbb{C}[x]\otimes\Omega_{U_{\rm gl}}$, with $(\mathbf{s}_\mathbf{n})_+\cdot\Omega_{U_{\rm gl}} = 0$, but it is not clear that $\Omega_{U_{\rm gl}}$ in general can be identified with $\mathbb{C}[M]$. Since the vertex operators defined in (4.18) satisfy the relationship (3.7) it follows that the required vertex operators will have a decomposition of the type given in (4.18) with

$$\begin{aligned} E^{a,b}_\pm(\omega p,\lambda p) = \sum_{j\geqslant 1} j^{-1}((\omega p)^{\pm j\tilde{n}_a} r^a(\mp j) \\ - (\lambda p)^{\pm j\tilde{n}_b} r^b(\mp j)). \end{aligned} \tag{4.20}$$

The operators defined in (4.20) act on $U_{\rm gl}$ according to (4.1). The generators of the $\mathbf{Z}_{U_{\rm gl}}$ algebra can be obtained by using the method developed in Ref. 18 to obtain the generalized commutation relationships.[15] Essentially the lattice has to be modified to include all the elements $\{\pi(r^a(j)): 1\leqslant a\leqslant m, j\in Z\setminus\{0\}\}$, where π is the covering homomorphism of the previous section. An alternative approach is to relate the realizations of Sec. III directly to the representations obtained by Lepowsky.[13] This is the method adopted in Sec. V.

The same considerations apply to the fundamental representations associated with $\mathbf{a}^{(1)}_{n-1}$. In addition, both the vertex operators and the representation spaces can be obtained by restrictions imposed on those for $\mathbf{gl}_n^{(1)}$. We now derive the constraints on $\mathbb{C}[x]$ and the action of $\mathbf{s}_\mathbf{n}\subset\mathbf{gl}_n^{(1)}$ so that they can be respectively identified with the level 1 module $\mathrm{Sym}(\tilde{\mathbf{s}}_\mathbf{n})_-$ and the action of $\tilde{\mathbf{s}}_\mathbf{n}\subset\mathbf{a}^{(1)}_{n-1}$ on it where $\tilde{\mathbf{s}}_\mathbf{n}$ is the maximal Heisenberg defined in (3.6). Order the Heisenberg subalgebra $\mathbf{s}_\mathbf{n}$ so that $n_a\neq 1$, $a = 1,\dots,q$ and $n_a = 1$, $a = q+1,\dots,m$ ($q = 0$ is also included and has the obvious meaning). Apply the corresponding automorphism to $\mathbf{gl}_n^{(1)}$.

Identify the generators of $\mathbf{s}_{n_a}$ in the decomposition (3.6) with the following operators on $\mathbb{C}[x]$:

$$r^a(j)\to\partial_{x_j}(a),\quad r^a(-j)\to jx_j^{(a)},\quad j>0,\quad j\not\equiv 0 \bmod n_a,$$
$$a = 1,\dots,q. \tag{4.21a}$$

and the generators of $\mathbf{h}_\mathbf{n}$ with

$$h^a(j)\to\sum_{b=1}^{m}\lambda^a_b\partial_{x^{(b)}_{jn_b}},\quad h^a(-j)\to j\sum_{b=1}^{m}\lambda^a_b x^{(b)}_{jn_b},$$
$$j>0,\quad a = 1,\dots,m-1. \tag{4.21b}$$

Put $h^m(0) := \sum_{a=1}^{m}\varphi_a r^a(0)$ so that $h^m(j) = h^m(0)\otimes k^{jd}$ and define $\boldsymbol{\varphi} := (\varphi_1,\dots,\varphi_m)$, $\boldsymbol{\psi} := (n_1\varphi_1,\dots,n_m\varphi_m)$. Introduce the additional Heisenberg subalgebra $\mathbf{h}_m$ generated by

$\{h^m(j),\ z{:}j\in Z\}$,

$$[h^m(i),h^a(j)] = i\delta^{m,z}z, \quad a=1,\ldots,m.$$

The action of $\mathbf{h}_m$ is given by

$$h^m(j)\to\sum_{b=1}^m \varphi_b\partial_{x^{(b)}_{jn_b}},\quad h^m(-j)\to j\sum_{b=1}^m \varphi_b x^{(b)}_{jn_b},\quad j>0, \tag{4.21c}$$

and φ is uniquely determined,

$$\varphi=\alpha\left(\mathbf{n}-\sum_{b=1}^{m-1}(\mathbf{n}\cdot\gamma^b)\lambda^b\right),\quad \varphi\cdot\psi=1,\quad \alpha\in\mathbb{R}.$$

This expression for φ shows that

$$h^m(j)=\alpha\left(J(j)-\sum_{b=1}^{m-1}(\mathbf{n}\cdot\gamma^b)h^b(j)\right),$$

$$J(j):=\sum_{a=1}^m n_a r^a(j). \tag{4.22}$$

Considered as elements of $\mathbf{gl}_n^{(1)}$ there is an invertible transformation connecting $\{r^a(jn_a){:}\ 1\leqslant a\leqslant m\}$ and $\{h^a(j){:}\ 1\leqslant a\leqslant m\}$ for fixed $j\neq 0$. However if we consider the restriction of the representation of $\mathbf{s}_n\subset\mathbf{gl}_n^{(1)}$ on $\mathbb{C}[x]$ to $\tilde{\mathbf{s}}_n\subset\mathbf{a}_{n-1}^{(1)}$ on $\mathbb{C}[\tilde{x}]\equiv\mathrm{Sym}(\tilde{\mathbf{s}}_n)_-\subset\mathbb{C}[x]$ then Eq. (4.22) shows that we require the action of $\{J(j){:}\ j\in Z\setminus\{0\}\}$ to be trivial. The action is trivial provided

$$\sum_{a=1}^m n_a x^{(a)}_{jn_a}=0,\quad \sum_{a=1}^m n_a\partial_{x^{(a)}_{jn_a}}=0,\quad j>0. \tag{4.23}$$

Since $h^m(0)\notin\mathbf{a}_{n-1}^{(1)}$ does not act on the Fock space $\mathrm{Sym}(\tilde{\mathbf{s}}_n)_-$ there is no need to ensure that action is trivial. It follows that

$$\mathbb{C}[\tilde{x}]:=\left\{x\in\mathbb{C}[x]{:}\ \sum_{a=1}^m n_a x^{(a)}_{jn_a}=0\right\}.$$

For the cases $n_a=N$, $a=1,\ldots,m$ the matrices $h^a(0)$ can be calculated by the Gram–Schmidt orthonormalization technique.

In two cases the restriction of the character formula (4.17) to the $\mathbf{a}_{n-1}^{(1)}$ module $U_a:=\mathbb{C}[\tilde{x}]\otimes C[M]$ coincides with the known expressions for a fundamental module determined by a maximal Heisenberg subalgebra. The corresponding vertex operators are determined by restricting (4.18),

$$X_a^{b,c}(\omega p,\lambda p):=X_{\mathrm{gl}}^{b,c}(\omega p,\lambda p)|_{\mathbb{C}[\tilde{x}]}. \tag{4.24}$$

The space U_{gl} decomposes into irreducible $[X_{\mathrm{gl}}^{b,c}(\omega p,\lambda p)]$ subspaces

$$U_{\mathrm{gl}}=\bigoplus_{s\in Z}U^s_{\mathrm{gl}},\quad U^s_{\mathrm{gl}}:=\mathbb{C}[x]\otimes e^{\Lambda_s}\mathbb{C}[\widetilde{M}],$$

where $\mathbb{C}[\widetilde{M}]:=\{\mu\in\mathbb{C}[M]{:}\ l(\mu)=0\}$ and $l(\Lambda_s)=s$. A similar decomposition holds for U_a.

As remarked in Sec. III the case $n_a=N$, $a=1,\ldots,m$ is the only case where the Chevalley generators of the J-reduced algebra can be easily determined. The corresponding coroots obtained from (3.5) are

$$\alpha^\vee_{m(N-1)}=F^{m,m}_{N-1,N-1}-F^{1,1}_{0,0}+(1/N)z,$$

$$\alpha^\vee_{m(r-1)+m}=F^{m,m}_{r,r}-F^{1,1}_{r+1,r+1}+(1/N)z,\quad 0\leqslant r\leqslant N-2,$$

$$\alpha^\vee_{m(r-1)+a}=F^{a,a}_{r,r}F^{a+1,a+1}_{r,r},\ a=1,\ldots,m-1,\ 1\leqslant r\leqslant N-2. \tag{4.25}$$

The two known cases are the principal ($m=1$, $n_1=N$) and the homogeneous representations ($n_a=1$, $1\leqslant a\leqslant m$).[11,16]

(i) $m=1, n_1=N$ (principal representation of $a^{(1)}_{N-1}$):
Heisenberg subalgebra

$$r^1(j)\to\partial_{x^{(1)}_j},\quad r^1(-j)\to jx^{(1)}_j\ j>0,\quad j\not\equiv 0 \bmod N,$$

vertex operators

$$X_a^{1,1}(\omega p,p)=\frac{\omega}{\omega-1}\exp\left(\sum_{j\not\equiv 0 \bmod N}p^j(\omega^j-1)x^{(1)}_j\right)\times\exp\left(\sum_{j\not\equiv 0 \bmod N}p^{-j}(\omega^{-j}-1)D_{x^{(1)}_j}\right).$$

The character formula gives

$$\mathrm{ch}\ U^s_a=e^{sr}1_{j\neq 0}\prod_{\bmod N}(1-q^j)^{-1},$$

where $s\in Z_N$ labels the fundamental representation. The vertex operators corresponding to the sth fundamental representation (U^s_a,π^s) are $c_sX_a^{1,1}(\omega p,p)$ and c_s is easily determined using (4.25). Since in this case v_{Λ_s} is a highest weight vector define

$$F^{1,1}_{i,i}v_{\Lambda_s}=\begin{cases}\left(\dfrac{i+1}{N}\right)v_{\Lambda_s}, & i\leqslant s,\\[2mm] \left(\dfrac{i+1}{N}-1\right)v_{\Lambda_s}, & i>s,\end{cases}$$

and then (4.25) gives $\alpha_j\cdot v_{\Lambda_s}=\delta_{j,s}v_{\Lambda_s}$. The identity obtained from π_s: $\overline{\mathrm{gl}}^{1,1}_N(\omega p,p)\to c_sX_a^{1,1}(\omega p,p)$ acting on $v_{\Lambda_s}\to 1\otimes e^{sr}1$ then fixes $c_s=\omega^s$.

(ii) $n_a=1$; $a=1,\ldots,m$ (homogeneous representation of $a^{(1)}_{m-1}$):
Heisenberg subalgebra

$$h^a(j)\to\sum_{b=1}^{m-1}(\lambda^a_b-\lambda^a_m)\partial_{x^{(b)}_j}h^a(-j)\to j\sum_{b=1}^{m-1}(\lambda^a_b-\lambda^a_m)x^{(b)}_j,\quad j>0,\quad a=1,\ldots,m-1.$$

vertex operators

$$X^{a,b}(p,p)=\epsilon(r_a,r_b)\exp\left(\sum_{j\geqslant 1}p^j(x^{(a)}-x^{(b)})\right)\times\exp\left(-\sum_{j\geqslant 1}p^{-j}(D_{x^{(a)}}-D_{x^{(b)}})\right)\times Z^{a,b}(p,p)\quad a\neq b,a,b=1,\ldots,m.$$

If is convenient to make the identification $\mathbb{C}[\tilde{x}]\equiv\mathbb{C}[y]:=\mathbb{C}[y^{(1)},\ldots,y^{(m-1)}]$ so that on $\mathbb{C}[y]$, $y^{(a)}=x^{(a)}-x^{(a+1)}$, $a=1,\ldots,(m-1)$ and $D_{x^{(1)}}=D_{y^{(1)}}$, $D_{x^{(m)}}=-D_{y^{(m-1)}}$, $D_{x^{(a)}}=-D_{y^{(a-1)}}+D_{y^{(a)}}$, $a=2,\ldots,(m-1)$. This is equivalent to the Frenkel–Kac representation[11]; the factor $\epsilon(r_a,r_b)$ in the vertex operator is usually absorbed into the corresponding root vector. The action of $F^{a,a}_{0,0}$ on v_{Λ_s}, a highest weight vector in $U^s_a, s\in Z_m$, is defined by

$$F^{a,a}_{0,0}\cdot v_\Lambda = \begin{cases}(1-\delta_{s,0})v_{\Lambda_s}, & a\leqslant s,\\ 0, & a>s,\end{cases}$$

so that from (4.25) we get $\alpha_j\cdot v_{\Lambda_s} = \delta_{j,s}v_{\Lambda_s}$. It follows that $U^s_a = \mathbb{C}[y]\otimes e^{\Lambda_s}C[\widetilde{M}]$ where $\Lambda_0=0$ and $\Lambda_s=\Sigma^s_{a=1}r_a$, $1\leqslant s\leqslant m-1$. Since the simple roots $\{\beta_a: a=1,...,m-1\}$ of a_{m-1} are $\beta_a = r_a - r_{a+1}$, the $\Lambda_s, s\neq 0$, are the miniscule weights of $\mathbf{a}_{m-1}$. A highest weight vector is $1\otimes e^{\Lambda_s}$ since $X^{a,b}_j\cdot(1\otimes e^{\Lambda_s})=0$, $j>0$, where $X^{a,b}_j$ is the coefficient of p^j in the formal expansion of $X^{a,b}_a(p,p)$. The character formula gives

$$\operatorname{ch} U^s_a = \prod_{i\geqslant 1}(1-q^i)^{-m}\sum_{\mu\in C[\widetilde{M}]+\Lambda_s} q^{\frac{1}{2}(\mu,\mu)}e^\mu.$$

V. THE CANONICAL FUNDAMENTAL REPRESENTATIONS OF $\mathbf{a}^{(1)}_{n-1}$

Let π: $\mathbf{gl}^{(1)}_n\to\mathbf{gl}(n,\mathbb{C})$ denote the covering homomorphism so that $\mathbf{gl}(n,\mathbb{C})$ has the $1-\tilde{\mathbf{n}}\mathbf{1}_n$ gradation determined by the d th-order automorphism σ defined earlier and as before let Λ, a d th root of unity, determine the eigenspace decomposition of $\mathbf{gl}(n,\mathbb{C})$ with respect to σ. Then $\mathbf{h}^0:=\pi(\mathbf{s}_n)$ is a Cartan subalgebra of $\mathbf{gl}(n,\mathbb{C})$. Define a trace form on $\mathbf{h}^0$ by

$$\langle r^a_i,r^b_j\rangle = n_a\delta^{a,\underline{b}}_{i,\underline{-j}},$$

where

$$\binom{b}{\underline{j}}\equiv\binom{b}{j \bmod n_b}.$$

As is well known there exists a unitary matrix U such that $\tilde{\mathbf{h}}^0:=U\mathbf{h}^0U^{-1}$ is the conjugate Cartan subalgebra of diagonal matrices. However since the bilinear form $\langle\cdot,\cdot\rangle$ is invariant under similarity transformations we consider the Abelian group A which is generated by elements $\alpha\in\mathbf{h}^0$ of the form [c.f. (4.20)]

$$\alpha = n_a^{-1}\sum_{i\in Z_{n_a}}\omega^{-i\tilde{n}_a}r^a_i - n_b^{-1}\sum_{j\in Z_{n_b}}\lambda^{-j\tilde{n}_b}r^b_j, \qquad (5.1)$$

where ω and λ are d th roots of unity (i.e., there exists i,j such that $\omega=\Lambda^i$, $\lambda=\Lambda^j$). Observe that the α defined in (5.1) corresponds to an equivalence class of representations defined by $\omega\to\omega\Lambda^{jn}a$, $\lambda\to\lambda\Lambda^{jn}b$. In calculations we always use the canonical representative such that corresponding to the pair (r^a_i,ω) in (5.1) we choose the element of the equivalence class $\omega\equiv\Lambda^j$ for which $\{n_a,j\}$ are relatively prime.

It is clear that for $\alpha,\beta,\in A,\alpha+\beta$ has the form (5.1) only if $\alpha+\beta=0$ or $\langle\alpha,\beta\rangle = -1$ when $\alpha+\beta\in A$. Let Δ be the set of such elements $\alpha\in A$. It is easy to check that the elements of A satisfy the conditions (2.1)–(2.5) for the Abelian group L introduced by Lepowsky,[15] i.e., $\langle\alpha,\alpha\rangle\in 2Z$, $\langle\sigma\alpha,\sigma\beta\rangle = \langle\alpha,\beta\rangle$ and $\langle\sigma^{d/2}\alpha,\alpha\rangle\in 2Z$ for even d.

Introduce a bimultiplicative function ϵ on $A\times A$ which takes values in $\{\pm 1\}$,

$$\epsilon(\alpha+\beta,\lambda)=\epsilon(\alpha,\lambda)(\beta,\lambda),\ \epsilon(\alpha,\beta+\lambda)=\epsilon(\alpha,\beta)\epsilon(\alpha,\lambda),$$
$$\epsilon(\alpha,\beta)\epsilon(\beta,\alpha) = (-1)^{\langle\alpha,\beta\rangle}. \qquad (5.2)$$

Define ϵ on A in the following way. Let

$$\epsilon_0(\Lambda^i r^a_0,\Lambda^j r^b_0):=\begin{cases}1, & a<b,\\ 1, & a=b,\quad i\leqslant j,\\ -1, & a=b,\quad i>j,\\ -1, & a>b.\end{cases} \qquad (5.3)$$

Note that the evaluation here is with respect to the canonical representative for the pairings (r^a_0,Λ^i) etc. Put

$$\epsilon\left(n_a^{-1}\sum_{i\in Z_{n_a}}\omega^{-i\tilde{n}_a}r^a_i,\ n_b^{-1}\sum_{j\in Z_{n_b}}\lambda^{-j\tilde{n}_b}r^b_j\right)$$
$$:=\epsilon_0(\omega r_0,\lambda r_0), \qquad (5.4)$$

then $\epsilon(\cdot,\cdot)$ so defined satisfies the conditions (5.2) on $A\times A$.

Let $x=\oplus_{i\in Z}x_{(i)}$ be the eigenspace decomposition of $x\in\mathbf{gl}(n,\mathbb{C})$ and set $x(\sigma):=\Sigma_{j\in Z}x_{(j)}\otimes k^j$ and $x(p):=\Sigma_{j\in Z}x_{(j)}\otimes k^jp^{-j}$. In particular define

$$x_\alpha(p):=(n_an_b)^{-1/2}\epsilon_0(\omega r_0,\lambda r_0)\pi(\overline{\mathrm{gl}}_n(\omega p,\lambda p)), \qquad (5.5)$$

where α is defined in (5.1). Then we have for $\alpha,\beta\in\Delta$ that

$$[x_\alpha,x_\beta]_0=\begin{cases}\epsilon(\alpha,-\alpha)\alpha, & \alpha+\beta=0,\\ \epsilon(\alpha,\beta)x_{\alpha+\beta}, & \langle\alpha,\beta\rangle=-1,\\ 0, & \langle\alpha,\beta\rangle\geqslant 0,\end{cases} \qquad (5.6)$$

and $[\mathbf{h}^0,\mathbf{h}^0]_0=0$, $[h,x_\alpha]_0=\langle h,\alpha\rangle x_\alpha = -[x_\alpha,h], h\in\mathbf{h}^0$. The form extends naturally to the whole of $\mathbf{gl}(n,\mathbb{C})$ by putting $\langle x_\alpha,x_\beta\rangle = \delta_{\alpha,-\beta}\epsilon(\alpha,-\alpha)$. Notice that $\sigma x_\alpha = x_{\sigma\alpha}$.

Put $Z(\alpha,p):=\epsilon_0(\omega r^a,\lambda r^b)(n_an_b)^{-1/2}\ Z^{a,b}(\omega p,\lambda p)$, where α is defined as in (5.1) and

$$Z^{a,b}(\omega p,\lambda p):=\exp(-E^{a,b}_+(\omega p,\lambda p))\,\overline{\mathrm{gl}}_n(\omega p,\lambda p)$$
$$\times\exp(E^{a,b}_-(\omega p,\lambda p)), \qquad (5.7)$$

which is obtained by inverting (4.18). The quantities $E^{a,b}_\pm(\omega p,\lambda p)$ are defined in (4.20). The realization of $Z(\alpha,p)$ is determined from its action on a level one module U_{gl} of $\mathbf{gl}^{(1)}_n$. The definition (5.7) is therefore to be understood as the representation of $Z(\alpha,p)$ in this module. Its action on U_{gl} is then as usual conveniently represented by $\cdot$.

The representation of $Z(\alpha,p)$ is determined from the generalized commutation relations for $\mathrm{gl}_n(\omega p,\lambda p)$. This is given in Ref. [12] but we shall omit the derivation and state the result. Set $I(n):=\{j\in Z_d:\langle\sigma^j\alpha,\beta\rangle = n\}$. Then we can prove directly the following result.[15]

Lemma 5.1: For α, β, $\in\Delta$,

$$\prod_{j\in Z_d}(1-\Lambda^{-j}q/p)^{\langle\sigma^j\alpha,\beta\rangle}Z(\alpha,p)Z(\beta,q) - \prod_{j\in Z_d}(1-\Lambda^{-j})^{\langle\alpha,\sigma^j\beta\rangle}Z(\beta,q)Z(\alpha,p)$$
$$= d^{-1}\sum_{j\in I(-1)}\epsilon(\sigma^j\alpha,\beta)Z(\sigma^j\alpha+\beta,q)\delta\left(\Lambda^{-j}\frac{p}{q}\right) - d^{-1}\sum_{j\in I(-2)}\epsilon(-\beta,\beta)\beta(q)\delta\left(\Lambda^{-j}\frac{p}{q}\right) + d^{-2}\sum_{j\in I(-2)}D\delta\left(\Lambda^{-j}\frac{p}{q}\right).$$

In this lemma D is the degree operator $Df(p) = p(d/dp)f(p)$. Lemma 5.1 is a special case of the general formula derived in Ref. 18 ($p\to p^{-1},q\to q^{-1}$ to get agreement with the formula of Lepowsky and Wilson).

It follows that we can write down the fundamental representations for a canonical set of maximal Heisenberg subalgebras of $\mathbf{gl}_n^{(1)}$ and $\mathbf{a}_{n-1}^{(1)}$ by using the results in Ref. 13. More directly we can obtain the representation by introducing the operators $\{p^\alpha,e^\alpha\colon \alpha\in A\}$ on U_{gl} such that

$$e^\alpha\cdot e^\beta=\epsilon_C(\alpha,\beta)e^{\alpha+\beta},\qquad p^\alpha\cdot e^\beta=e^\beta\cdot p^{\alpha+\langle\alpha,\beta\rangle}. \tag{5.8}$$

These operators are the obvious analogs to the operators defining the $T_\gamma(k)$ in (4.3). In this case however we do not require the bimultiplicative function $\epsilon_C(.,.)$ to take values in $\{\pm 1\}$. We can then show directly that

$$Z(\alpha,p):=d^{-\frac12\langle\alpha,\alpha\rangle}f(\alpha)e^\alpha p^{\Sigma\sigma^j\alpha+\frac12\langle\Sigma\sigma^j\alpha,\alpha\rangle} \tag{5.9}$$

defines an operator which when restricted to Δ satisfies Lemma 5.1. In (5.9) the function $f(\alpha)$ is defined by

$$f(\alpha)=\begin{cases} f'(\alpha)2^{\langle\sigma^{d/2}\alpha,\alpha\rangle/2}, & d\in 2Z,\\ f'(\alpha), & d\in 2Z+1,\quad \alpha\in A,\end{cases} \tag{5.10}$$

where

$$f'(\alpha)=\prod_{0<j<d/2}(1-\Lambda^{-j})^{\langle\sigma^j\alpha,\alpha\rangle},\quad \alpha\in A. \tag{5.11}$$

The function $\epsilon_C(.,.)$ has the representation

$$\epsilon_C(\alpha,\beta)=\epsilon(\alpha,\beta)/\epsilon'(\alpha,\beta),$$

where

$$\epsilon'(\alpha,\beta)=\prod_{-d/2<j<0}(-\Lambda^{-j})^{\langle\sigma^j\alpha,\beta\rangle},\quad \alpha\in A, \tag{5.12}$$

and satisfies,

$$\epsilon_C(\alpha,\beta)/\epsilon_C(\beta,\alpha)=C(\alpha,\beta),\qquad C(\alpha,\beta)=\prod_{j\in Z_d}(-\Lambda^j)^{\langle\sigma^j\alpha,\beta\rangle}. \tag{5.13}$$

This follows from the calculation involved in establishing the validity of the expression for $Z(\alpha,p)$.

The corresponding vertex operators for $\mathbf{gl}_n^{(1)}$ with the $\tilde{\mathbf{n}}\mathbf{1}_n$ gradation are defined by

$$x_\alpha(p)\to X(\alpha,p):=E_+(\alpha,p)E_-(\alpha,p)Z(\alpha,p),\quad \alpha\in\Delta, \tag{5.14}$$

and $E_\pm(\alpha,p):=\exp[\pm E_\pm^{a,b}(\omega p,\lambda p)]$ for α defined in (5.1).

A level 1 $\mathbf{gl}_n^{(1)}$ module is defined by $U_{\mathrm{gl}} = \mathrm{Sym}(\mathbf{s}_{\mathbf{n}-})\otimes\Omega_{\mathrm{gl}}$, where Ω_{gl} is the vacuum space of $\mathbf{s}_{\mathbf{n}+}$, $\mathbf{s}_{\mathbf{n}+}\cdot\Omega_{\mathrm{gl}}=0$. The action of $\mathbf{s}_\mathbf{n}$ on U_{gl} is the usual one given in the previous section. If U_{gl} is a highest weight module then Ω_{gl} is generated by the action of the $\mathbf{Z}_{U_{\mathrm{gl}}}$ algebra on a highest weight vector $v\in U_{\mathrm{gl}}$. In particular we have

$$\begin{aligned} &e^\alpha\cdot e^\beta\cdot v=\epsilon_C(\alpha,\beta)e^{\alpha+\beta}\cdot v,\\ &h\cdot e^\beta\cdot v=\langle h,\beta\rangle e^\beta\cdot v,\\ &p^\alpha\cdot e^\beta\cdot v=p^{\langle\alpha,\beta\rangle}e^\beta\cdot v,\quad \alpha,\beta\in A.\end{aligned} \tag{5.15}$$

U_{gl} is an irreducible $\mathbf{gl}_n^{(1)}$ module provided Ω_{gl} is an irreducible $S_\mathbf{n}$ module, where $S_\mathbf{n}$ is the group corresponding to $\mathbf{s}_\mathbf{n}$. Let P_0 denote the projection operator $P_0\colon \mathbf{h}^0\to\mathbf{h}^0_{(0)}$. Put $N:=(1-P_0)\mathbf{h}^0A$ and let $\hat{A}$ denote the unique up to equivalence central extension $1\to\langle\Lambda_0\rangle\to\hat{A}\to A\to 1$, where $\Lambda_0 = (-1)^d\Lambda$. Let $\hat{N}$ be the pullback of N to $\hat{A}$. Then Lepowsky shows that an irreducible $S_\mathbf{n}$ module has the structure $\Omega_{\mathrm{gl}}\approx\mathbb{C}[\hat{A}]\otimes_{\mathbb{C}[\hat{N}]}T$ where T is an irreducible $\hat{N}$ module. The construction of T is given in Ref. 13 and we refer the reader to this reference and also Ref. 10 for a detailed consideration of the fundamental modules.

To obtain the corresponding $\mathbf{a}_{n-1}$ fundamental representations we impose the restrictions (4.23) on the fundamental representations U_{gl} and $X_{\mathrm{gl}}(\alpha,p)$. In the notation of this section this requires

$$\sum_{a=1}^m n_a r^a(jn_a)=0,\quad j\neq 0. \tag{5.16}$$

The character formulas (4.17) agree with the formulas derived by Kac and Peterson.[10] They correspond to a gradation of the fundamental module defined by

$$\deg(w\otimes e^\alpha\otimes t)=\deg w-\frac12\left\langle\sum_{j\in Z_d}\sigma^j\alpha,\alpha\right\rangle,\qquad w\in\mathrm{Sym}(\mathbf{s}_{\mathbf{n}-}),\quad e^\alpha\otimes t\in\Omega_{\mathrm{gl}}. \tag{5.17}$$

We end this section with the first nontrivial example which arises from $n=3$ ($n=2$ has only the principal and homogeneous representations already considered). The possibilities are: (i) $\mathbf{n}=(3)$, (ii) $\mathbf{n}=(1,1,1)$, and (iii) $\mathbf{n}=(2,1)$. The first two are respectively the principal and homogeneous representations which are easily seen to be in agreement with those given in the previous section [for (i) $\dim\Omega_{\mathrm{gl}}=1$]. Case (iii) corresponds to a new representation. If we use $\mathbb{C}[x]\,\mathrm{Sym}(\mathbf{s}_{\mathbf{n}-})$ which was introduced in the last section we get $\pm\alpha_1$, $\pm\alpha_2$, $\pm\alpha_3\in\Delta$ and $\alpha_1=\alpha_2-\alpha_3$, where

$$\alpha_1=r_1^1,\quad \alpha_2=\tfrac12(r_0^1+r_1^1)-r_0^2,\quad \alpha_3=\tfrac12(r_0^1-r_1^1)-r_0. \tag{5.18}$$

For this case only a single infinity of indeterminants x is required and the irreducible $\mathbf{a}_2^{(1)}$ module is $\mathbb{C}[x]\otimes\Omega_a,\Omega_a\approx\Omega_{\mathrm{gl}}$:

$$X_a(\alpha_1,p)=\frac14\exp\Big(\sum_{j\geqslant1}2x_{2j-1}p^{2j-1}\Big)\times\exp\Big(-\sum_{j\geqslant1}2D_{x_{2j-1}}p^{-2p+1}\Big)e^{\alpha_1},$$

$$X_a(\alpha_2,p)=2^{-\frac12}\exp\Big(\sum_{j\geqslant1}(x_{2j-1}+3p^{2j}x_{2j})\Big)\times\exp\Big(-\sum_{j\geqslant1}(p^{-2j+1}D_{x_{2j-1}}+3p^{-2j}D_{x_{2j}})\Big)\times e^{\alpha_2}p^{3/2}+2P_0\alpha_2,$$

$$X_a(\alpha_3,p)=2^{-\frac12}\exp\Big(\sum_{j\geqslant1}(-p^{2j-1}x_{2j-1}+3p^{2j}x_{2j})\Big)$$

$$\times \exp\Big(-\sum_{j\geq 1}(-p^{-2j+1}D_{x_{2j-1}}+3p^{-2j}D_{x_{2j}})\Big)$$
$$\times e^{\alpha_3}p^{3/2}+2P_0\alpha_3. \tag{5.19}$$

VI. THE WEDGE REPRESENTATION AND GROUP ACTIONS

The groups $\mathrm{GL}(m\infty)$ and $\overline{\mathrm{GL}}(m\infty)$ were defined in Sec. III. The group $\mathrm{GL}(m\infty)$ is generated by the action of the exponential map on $\mathbf{gl}(m\infty)$. Although $\overline{\mathbf{gl}}(m\infty)$ is the Lie algebra of $\overline{\mathrm{GL}}(\infty)$, conversely the exponential map is not defined on all elements of $\overline{\mathbf{gl}}(m\infty)$. However from (4.12) we have

$$\exp \sum_{a=1}^{m}\sum_{i\geq 1} x_i^{(a)}s^a(i) = \mathop{\times}_{a=1}^{m} (p_{s-t}(x^{(a)}))_{s,t\in Z}\in \overline{\mathrm{GL}}(m\infty). \tag{6.1}$$

This follows from the fact that the subgroups $\exp x_i^a s^a(i)$, $i>0$, are commutative subgroups of $\overline{\mathrm{GL}}(m\infty)$, since for $i,j>0$, $[s^a(i),s^b(j)]_0=0$. Define $\overline{S}_m$ as the centralizer of $\overline{\mathbf{s}}_m$ in $\overline{\mathrm{GL}}(m\infty)$. Thus the element (6.1) belongs to $\overline{S}_m$.

Let

$$\overline{V}^{(m(i-1)+a)}=\Big\{\sum_{b=1}^{m}\sum_{j<1}c_j^b u_j^b+\sum_{b=1}^{a}c_i^b u_i^b; c_i^b\in\mathbb{C}\Big\}\subset\overline{V},$$

so that $\subset\overline{V}^{(n)}\subset\overline{V}^{(n+1)}\subset$ and $\dim\overline{V}^{(n+1)}/\overline{V}^{(n)}=1$. The Grassmannian of $\overline{V}$, Gr $(\overline{V})$ is the set of closed subspaces $W\subset\overline{V}$ such that there exists an n for which $\overline{V}^{(n)}\subset W$ and $\dim W/\overline{V}^{(n)}<\infty$. This means that the subspaces W are "comparable" to $\overline{V}^{(0)}$ in the sense that the projection pr_W: $W\to\overline{V}^{(0)}$ is Fredholm (finite kernel and cokernel).[8,19,20] The index associated with pr_W is called the virtual dimension of W:

$$\text{virt.dim } W = \text{index pr}_W := \dim(\ker \mathrm{pr}_W)-\dim(\text{coker pr}_W).$$

Since $\widehat{\mathrm{GL}}(m\infty)$ consists of the identity component $\mathrm{Gr}(\overline{V})$ decomposes into the orbits of $\widehat{\mathrm{GL}}(m\infty)$,

$$\mathrm{Gr}(\overline{V})=\mathop{\mathrm{Gr}_p}_{p\in Z}(\overline{V}),$$

where $\mathrm{Gr}_p(\overline{V})=\{W\in\mathrm{Gr}(\overline{V})\text{: virt.dim}W=p\}$. However the action of $\overline{\mathrm{GL}}(m\infty)$ is transitive on $\mathrm{Gr}(\overline{V})$; we denote the identity component by $\overline{\mathrm{GL}}_0(m\infty)$.

Let $\Omega:=\wedge\overline{V}=\bigwedge_{a=1}^{m}\overline{V}^a$ denote the exterior algebra on $\overline{V}$. It is convenient to order the basis elements of Ω. Thus let $\alpha^a=u_{j_1}^a\wedge u_{j_2}^a\wedge\cdots\in\overline{V}^a$ then any basis element of Ω can be written in the form $\alpha=\alpha^1\wedge\alpha^2\wedge\cdots\wedge\alpha^m\in\Omega$. Introduce the definitions

$$\theta_j^a:=u_j^a\wedge u_{j-1}^a\wedge\cdots\in\overline{V}^a,$$
$$\theta_{m(i-1)+a}=\theta_i^1\wedge\cdots\wedge\theta_i^a\wedge\theta_{i-1}^{a+1}\wedge\cdots\wedge\theta_{i-1}^m\in\Omega.$$

The group $\widehat{\mathrm{GL}}(m\infty)$ acts on Ω in the usual way,

$$G\cdot(\alpha^1\wedge\cdots\wedge\alpha^a\wedge\cdots\wedge\alpha^m) = (G\cdot\alpha^1)\wedge\cdots\wedge(G\cdot\alpha^a)\wedge\cdots\wedge(G\cdot\alpha^m)$$
$$G\in\mathrm{GL}(m\infty),$$

It follows that Ω decomposes into the orbits of $\widehat{\mathrm{GL}}(m\infty)$, $\Omega=\oplus_{p\in Z}\Omega_p$ with $\theta_p\in\Omega_p$. If $\lambda=v_1\wedge v_2\wedge\cdots\in\Omega_p$, put $\lambda_w=\{\Sigma_{i\geq 1}c_i v_i; c_i\in\mathbb{C}\}$ so that $\lambda_w\in\mathrm{Gr}_p(\overline{V})$. Therefore define

$$\text{virt.deg }\lambda=\text{virt.dim }\lambda_w,$$

and $\alpha\in\Omega_p$ if virt.deg $\alpha=p$.

The algebra $\widehat{\mathbf{gl}}(m\infty)$ acts on Ω according to

$$g\cdot(\alpha^1\wedge\cdots\wedge\alpha^a\wedge\cdots\wedge\alpha^m) = (g\cdot\alpha^1)\wedge\cdots\wedge\alpha^a\wedge\cdots\wedge\alpha^m + \alpha^1\wedge\cdots\wedge(g\cdot\alpha^a)\wedge\cdots\wedge\alpha^m\wedge+\cdots + \alpha^1\wedge\cdots\wedge\alpha^a\wedge\cdots\wedge(g\cdot\alpha^m),\quad g\in\mathbf{gl}(m\infty). \tag{6.2}$$

Therefore we have

$$\mathbf{n}_+\cdot\theta_p=0,\quad \mathbf{n}\cdot\theta_p\subset\Omega_p,\qquad \breve{\alpha}_i^v\cdot\theta_p=\delta_{i,p}\theta_p, \tag{6.3}$$

where $\mathbf{n}_+(\mathbf{n}_-)$ are the subalgebras of $\widehat{\mathbf{gl}}(m\infty)$ spanned by the Chevalley generators $\{e_i\}(\{f_i\})$. Consequently (6.2) identifies Ω_p, with the action of $\widehat{\mathbf{gl}}(m\infty)$ given by (6.2), as the pth fundamental module of $\widehat{\mathbf{gl}}_\infty$ and θ_p is a highest weight vector. Denote this representation by $(\Omega_p,\hat{\pi}_p^m)$.

The representation extends by linearity to a projective representation $(\Omega_p,\overline{\pi}_p^m)$ of $\overline{\mathbf{gl}}(m\infty)$ if we use the normal ordering defined by the Clifford structure in Sec. III,

$$\overline{\pi}_p^m(E_{r,s}^{a,b})=E_{r,s}^{a,b}-\delta_{r,s}^{a,b}\theta(r)I_{\Omega_p}, \tag{6.4}$$

where I_{Ω_p} is the identity operator on Ω_p. A linear representation (Ω_p,π_p^m) of $\mathbf{gl}(m\infty)$ is obtained by letting z act as I_{Ω_p} on Ω_p.

We define virt.deg α^a, where $\alpha=\alpha^1\wedge\cdots\wedge\alpha^a\wedge\cdots\wedge\alpha^m$, as the restriction of virt.deg to the algebra $\overline{V}^a$ associated with the Grassmannian $\mathrm{Gr}(\overline{V}^a)$.

It is now possible to give an interpretation of $\mathbb{C}[M]$. From (6.4) we have

$$s^a(0)\cdot\alpha=\alpha^1\wedge\alpha^2\cdots\wedge s^a(0)\cdot\alpha^a\wedge\cdots\wedge\alpha^m = (\text{virt.deg }\alpha^a)\alpha,$$

and it follows that

$$s(0)\cdot\alpha=(\mu,s)\alpha,$$

where $s\in c$ and $\mu=\mu_a s_a\in M,\mu_a=$ virt.deg α^a. In particular since virt.deg $\alpha=p$ we find that $\mu=\Lambda_p+\lambda$ where $s(0)\cdot\theta_p=(\Lambda_p,s)\theta_p$ and

$$l(\lambda)=0\Big(\Lambda_p=i\sum_{b=1}^{a}s_b+(i-1)\sum_{b=a+1}^{m}s_b\Big).$$

The space Ω_p is an irreducible $\mathbf{gl}(m\infty)$ module and we can assign an m-principal gradation to it:

$$\lambda_{\mathbf{r,s}}^{\mathbf{a,b}}:=E_{r_0,s}^{a_0,b_0}E_{r_1,s-1}^{a_1,b_1}\cdots E_{r_{l-1},s-(l-1)}^{a_{l-1},b_{l-1}}\cdot\theta_p,$$

$$\deg \lambda^{\mathbf{a},\mathbf{b}}_{\mathbf{r},\mathbf{s}} = \begin{cases} \sum_{u=0}^{l-1} (s-u-r_u) - \frac{1}{2}(\mu,\mu), & b_0 \leqslant a, \\ 0 \ (\text{Note } \lambda^{\mathbf{a},\mathbf{b}}_{\mathbf{r},\mathbf{s}} = 0), & b_0 > a, \end{cases}$$

for $r_0 > r_1 > \cdots > r_{l-1} > s - (l-1)$,

where $s(0)\cdot\lambda^{\mathbf{a},\mathbf{b}}_{\mathbf{r},\mathbf{s}} = (\mu,s)\lambda^{\mathbf{a},\mathbf{b}}_{\mathbf{r},\mathbf{s}}$. (6.5)

If the character for Ω is defined as in Sec. III then it can be evaluated by using the properties of the character function $\mathrm{ch}\,\Omega = \mathrm{ch}(\Lambda^m_{a=1}\overline{V}^a)$. This gives

$$\mathrm{ch}\,\Omega = \left[1 + \sum_{l \geqslant 1} \frac{q^l}{(1-q)(1-q^2)\cdots(1-q^l)}\right]^m \times \sum_{\mu \in M} q^{(1/2)(\mu,\mu)} e^\mu. \tag{6.6}$$

Comparison with the character formula (4.9) for U gives Euler's identity. Then $\mathrm{ch}\,\Omega_p$ is obtained from (6.6) by restricting $\mu \in M$ so that $l(\mu) = p$. Conversely Euler's identity shows that the $\mathbf{gl}(m\infty)$ modules U and Ω can be identified. Let η: $\Omega \to U$ denote the isomorphism defined by $\theta_0 \to 1 \otimes e^0$.

Introduce the operators $\{u^a_i, u^{b\dagger}_j\}$ by the formulas,[20]

$$u^a_l \cdot (u^{a_1}_{i_1} \wedge u^{a_2}_{i_2} \wedge \cdots) = u^a_l \wedge u^{a_1}_{i_1} \wedge u^{a_2}_{i_2} \cdots,$$

$$u^{b\dagger}_l \cdot (u^{a_1}_{i_1} \wedge u^{a_2}_{i_2} \wedge \cdots) = u^{b*}_l (u^{a_1}_{i_1}) u^{a_2}_{i_2} \wedge \cdots - u^{b*}_l (u^{a_2}_{i_2}) u^{a_1}_{i_1} \wedge \cdots + \cdots,$$

where $u^{b*}_j(u^a_i) = \delta^{a,b}_{i,j}\ u^{b*}_j \in \overline{V}^*$. Then $\{u^a_i, u^{b\dagger}_j\}$ generate a Clifford algebra A_Ω which is isomorphic to the Clifford algebra A generated by $\{X_{i-1/2}(s_a), X_{-(j-1/2)}(-s_b)\}$ introduced in the previous section. Furthermore the A_Ω module C_Ω generated by the action of A_Ω on the vacuum vector θ_0 which satisfies

$$u^a_l \cdot \theta_0 = 0, \quad l \leqslant 0, \quad u^{a\dagger}_l \cdot \theta_0 = 0, \quad l > 0,$$

can be identified with the A module C. Let β: $C_\Omega \to C$ be the isomorphism defined by β: $\theta_0 \to v_0$ so that the transported operators are β: $u^a_i \to X_{i-1/2}(s_a), u^{a\dagger}_i \to X_{-(i-1/2)}(-s_a)$. The relationships between the different representations of the fundamental $\mathbf{gl}_\infty$ modules can be summed up in the commutative diagram

$$\begin{array}{ccc} (C,\pi^m) & \xrightarrow{\beta} & (C_\Omega,\pi^m) \\ \downarrow\gamma & & \uparrow\alpha \\ (U,\pi^m) & \xleftarrow{\eta} & (\Omega,\pi^m), \end{array}$$

where α and γ are the isomorphisms which identify the vacuum vector with the basic highest weight vector.

The Hermitian contravariant form on Ω is defined on basis elements $\alpha^a = u^a_{i_1} \wedge u^a_{i_2} \wedge \cdots, \beta^a = u^a_{j_1} \wedge u^a_{j_2} \wedge \cdots$ by

$$H_\Omega(\alpha^1 \wedge \cdots \wedge \alpha^m, \beta^1 \wedge \cdots \wedge \beta^m) = \prod_{a=1}^m H_\Omega(\alpha^a, \beta^a),$$

$$H_\Omega(\alpha^a,\beta^a) = \det(u^{a*}_{i_l}(u^a_{j_k}))_{l,k \geqslant 1}. \tag{6.7}$$

It is made unique by the normalization $H_\Omega(\theta_0,\theta_0) = 1$. The virtual degree decomposes the space Ω since a necessary condition for $H_\Omega(\alpha,\beta) \neq 0$ is that if $s(0)\cdot\alpha = (\mu,s)\alpha$, $s(0)\cdot\beta = (\lambda,s)\beta$ then $\mu = \lambda$.

For the group $\overline{\mathrm{GL}}(m\infty)$ the action on $(\Omega,\bar{\pi}^m)$ is projective and irreducible. If $W \in \mathrm{Gr}_p(\overline{V})$ let $\{w_1,\dots,w_l\}$ be a basis for $W \bmod V^{(n)}, W \subset V^{(n)}$ for some n. Then put $\alpha_w = w_1 \wedge \cdots \wedge w_l \wedge \theta_n$ so that $\alpha_w \in \Omega_p$ and up to a constant factor is independent of n and the basis $W \bmod \overline{V}^{(n)}$. For $G \in \overline{\mathrm{GL}}(m\infty)$ the representation on $(\Omega,\bar{\pi}^m)$ is given by

$$\begin{aligned} G\cdot\theta_0 &= \alpha_{G\cdot V^{(0)}}, \\ G\cdot u \cdot G^{-1} &= (G\cdot u), \\ G\cdot u^\dagger \cdot G^{-1} &= (u\cdot G^{-1})^\dagger, \end{aligned} \tag{6.8}$$

and the representation is made unique by fixing the constant in $\alpha_{G\cdot V^{(0)}}$ for each $G \in \overline{\mathrm{GL}}(m\infty)$ (cf. Refs. 3 and 12).

The isomorphism η can be made explicit by evaluating the matrix coefficients associated with a particular evolution operator on $\mathrm{Gr}(\overline{V})$ (cf. Refs. 3 and 12). Let

$$\Gamma_+(x) = \mathop{\times}_{a=1}^m (p_{i-j}(x^{(a)}))_{i,j \in Z} \in \overline{\mathrm{GL}}_0(m\infty),$$

and denote by Γ_+ the commutative subgroup $\{\Gamma_+(x)\} \subset \overline{\mathrm{GL}}_0(m\infty)$. The action of the transported operator on U is given by

$$\Gamma_+(x) = \exp \sum_{a=1}^m \sum_{j \geqslant 1} x^{(a)}_j s^a(j),$$

where $s^a(j)$ acts according to (4.1). Let $v = w(x) \otimes e^\mu \in V_p$ so that $l(\mu) = p$. Then if $\eta^{-1}(v) = \alpha \in \Omega_p$ we also have $s(0)\cdot\alpha = (\mu,s)\alpha$. Put $\eta^{-1}(1 \otimes e^\mu) = \alpha_\mu$ so that in particular $\alpha_{\Lambda_p} = \theta_p$ then the isomorphism η implies that

$$H_U(\Gamma_+\cdot v, 1 \otimes e^\mu) = H_\Omega(\Gamma_+\cdot\alpha, \alpha_\mu). \tag{6.9}$$

The isomorphism η: $\Omega \to U$ is naturally expressed in terms of an orthonormal basis for $\mathbb{C}[x]$, with respect to $H_{\mathbb{C}[x]}(\cdot,\cdot)$ called the basis of Schur functions.[21] Let Par consist of finite sequences of nonincreasing positive integers $f = (f_j,\dots,f_1)$, $0 \leqslant f_1 \leqslant \cdots \leqslant f_j$. Alternatively interpret $f \in$ Par as a Young diagram consisting of j rows of length $f_j,\dots,f_1$. Then the basis of Schur functions for $\mathbb{C}[x]$ is given by

$$S_f(x^{(a)}) = \det(p_{f_h - h + l}(x^{(a)}))_{0 < h,l \leqslant j},$$

where f ranges over Par (or all possible Young diagrams).

If $u^a_{i_l} \wedge u^a_{i_{l-1}} \wedge \cdots \in \overline{V}^a$, $i_l > i_{l-1} > \cdots$ put $\sigma = (i_l, i_{l-1} \cdots)$ and write $u^a_\sigma := u^a_{i_l} \wedge u^a_{i_{l-1}} \wedge \cdots$. If virt.deg $u^a_\sigma = l$ then σ is called a set of virtual cardinal l.[19,20] This means that for $q \ll 0$, $i_q = q$. Let $\Sigma_>$ be the family of all such strictly decreasing sequences σ. Then

$$\sum_> = \coprod_{u \in Z} \sum_>^j,$$

where $\Sigma^j_>$ is the set of sequences which have virtual cardinal j. To any element $\sigma \in \Sigma^j_>$ we can naturally assign a unique element $f_\sigma \in$ Par,

$$f_\sigma = (i_j - j, i_{j-1} - (j-1), \dots).$$

Theorem 6.1: Let $\sigma_a \in \Sigma^{j_a}_>$, $a = 1,\dots,m$ such that $l(\mu) = p$ where $\mu = j_a s_a \in M$. Then under the isomorphism η: $\Omega \to U$, $\eta = \oplus_{p \in Z} \eta_p$,

$$\eta_p(u^1_{\sigma_1} \wedge \cdots \wedge u^m_{\sigma_m}) = \left\{ \prod_{a=1}^m S_{f_{\sigma_a}}(x^{(a)}) \right\} \otimes e^\mu .$$

Proof: Let η_p^{-1}: $v(x) = w(x) \otimes e^\mu \to u^1_{\sigma_1} \cdots u^m_{\sigma_m}$, where virt.card. $(\sigma_a) = j_a$ and put $\mu = j_a s_a$. If η^{-1}: $1 \otimes e^\mu \to \alpha_\mu$ it follows that $\alpha_\mu = \theta^1_{j_1} \wedge \cdots \wedge \theta^m_{j_m}$. It is convenient to let $j_a \in \Sigma^{j_a}_>$ also represent the sequence $(j_a, j_a - 1, \ldots)$; the distinction is clear from the context. We have

$$H_U\left(\left(\exp \sum_{a=1}^m \sum_{j \geq 1} x_j^{(a)} \partial^{(a)}_{y_j}\right) v(y), 1 \otimes e^\mu\right) = w(x).$$

Put $\Gamma_+(x^{(a)}) = (p_{i-j}(x^{(a)}))_{i,j \in Z}$, then,

$$H_\Omega(\Gamma_+(x^{(1)}) u^1_{\sigma_1} \cdots \Gamma_+(x^{(m)}) u^m_{\sigma_m}, \alpha_\mu)$$
$$= \prod_{a=1}^m \det(\Gamma_+(x^{(a)})^{\sigma_a}_{j_a}) = \prod_{a=1}^m S_{f_{\sigma_a}}(x^{(a)}) ,$$

where $\Gamma_+(x^{(a)})^{\sigma_a}_{j_a} := (p_{l_i - (j_a - k)}(x^{(a)}))_{0 < i,k}$,

$$\sigma_a = (l_0, l_1, \ldots) , \quad j_a = (j_a, j_a - 1, \ldots) .$$

The relationship (6.9) and the definition $v(x) = w(x) \otimes e^\mu$ establishes the result.

Note that the action of $\Gamma_+(x^{(a)})$ is well defined and that the determinants can be explicitly calculated. This result is a generalization of that due to Kac and Peterson for the $m = 1$ case[12]; the result for $m = 1$ in the spin representation was established by the Japanese authors.[3]

It is clear from Theorem 6.1 that the basis for Ω_p (or U_p) can be defined in terms of $\overset{m}{\times} \Sigma_>$ (or $\overset{m}{\times}$ Par). Let $\Delta = (\sigma_1, \ldots, \sigma_m) \in \overset{m}{\times} \Sigma_>$ and let $F_\Delta = (f_{\sigma_1}, \ldots, f_{\sigma_m}) \in \overset{m}{\times}$Par be the corresponding partition (or sequence of Young diagrams). Put $u_\Delta := u^1_{\sigma_1} \cdots u^m_{\sigma_m}, S_{F_\Delta}(x) := \Pi^m_{a=1} S_{f_{\sigma_a}}(x^{(a)})$ and set $\mu_\Delta = h_a s_a \in M$, where h_a virt.card. (σ_a) (= virt.deg. $u^a_{\sigma_a}$) and define virt.card $(\Delta) = l(\mu_\Delta)$.

Corollary 6.1.1: (a) An orthonormal basis of Ω_p with respect to $H_\Omega(\cdot,\cdot)$ is given by $\{u_\Delta$: $\Delta \in \overset{m}{\times} \Sigma_>$ and virt.card. $\Delta = p\}$.

(b) An orthonormal basis of U_p with respect to $H_U(\cdot,\cdot)$ is given by $\{S_{F_\Delta}(x) \otimes e^{\mu_\Delta}$: $\Delta \in \overset{m}{\times} \Sigma_>$ and virt.card. $\Delta = p\}$.

Notice in part (b) that $\{S_F(x) \otimes e^\mu$: $F \in \overset{m}{\times}Par, l(\mu) = p\}$ is an equivalent formulation of the basis. From the theorem and the corollary it is possible to give a concise expression for the action of $G \in \widehat{\mathrm{GL}}(m\infty)$ on Ω_p or U_p.

$$G \cdot (S_{F_{\tilde\Delta}}(x) \otimes e^{\mu_{\tilde\Delta}}) \overset{\eta^{-1}}{\mapsto} G \cdot u_{\tilde\Delta} = \sum_{\substack{\Delta \in \overset{m}{\times} \text{ Par} \\ :\text{virt.card. } \Delta = p}} \det(G^\Delta_{\tilde\Delta}) u_\Delta$$

$$\overset{\eta}{\mapsto} \sum_{\substack{\Delta \in \overset{m}{\times} \text{ Par} \\ :\text{virt.card.} \Delta = p}} \det(G^\Delta_{\tilde\Delta})(S_{F_\Delta}(x) \otimes e^{\mu_\Delta}) .$$

In this expression if $\sigma_a = (k_1, k_2, \ldots)$ and $\tilde\sigma_b = (l_1, l_2, \ldots)$ then $(G^\Delta_{\tilde\Delta})^{a,b}_{i,j} := G^{a,b}_{k_i, l_j}$.

VII. A RESTRICTED CLASS OF SOLVABLE EQUATIONS

In this section we briefly review a restricted class of solvable equations associated with $\mathbf{gl}_\infty$; a detailed classification will be given elsewhere.[15] The equations are given in their Hirota form[14] and were first obtained in Refs. 8 and 7.

The equations are restricted because (a) only fundamental representations of $\mathbf{gl}_\infty$ are considered, (b) the general evolution operators on $\mathrm{Gr}(\overline{V})$ are not treated. The second point refers to the specific map used to define the isomorphism $\Omega \to U$. In the previous section we used an element of Γ_+ to define the isomorphism η. However different evolution operators lead to different equations.[3,9] This is because each "distinct" isomorphism defines a new vertex representation of $\mathbf{gl}_\infty$ on U.

We consider the restricted class of equations associated with Γ_+ for which η: $\Omega \to U$ is defined by $\theta_0 \to 1 \otimes e^0$. In this case $u^a_i \to X_{i-1/2}(s_a)$, $u^{a\dagger}_i \to X_{-(i-1/2)}(-s_a)$ as a result of uniqueness. However it is instructive to derive this result directly.

The group $\Gamma_+ \subset \bar S_m \subset \overline{\mathrm{GL}}(m\infty)$ acts on the vector space $\overset{m}{\oplus} \overline{V}^a\{k\}$ and its dual, where $\overline{V}^a\{k\}$ is the vector space of formal Laurent series in k with coefficients in $\overline{V}^a$. In particular let $u^a(k) = u^a_i k^i \in \overline{V}^a\{k\}$, $u^{a*}(k) = u^{a*}_i k^{-i} \in \overline{V}^{a*}(k)$ and put $k \cdot x^{(a)} := \sum_{i \geq 1} k^i x_i(a)$. Then we obtain either directly or from (3.8),

$$\Gamma_+(x) \cdot u^a(k) = (\exp k \cdot x^{(a)}) u^a(k),$$

$$\Gamma_+(x) \cdot u^{a\dagger}(k) = (\exp - k \cdot x^{(a)}) u^{a\dagger}(k).$$

Define $\epsilon(k) := (k, \frac{1}{2} k^2, \ldots, (1/j) k^j, \ldots)$, write u_j for u^a_j, and put $\alpha = \sum_{j \geq 1} k^{j-1} u_{l+j}$, then we have

$$u^a(k) \cdot \theta^a_l = k^{l+1} \alpha \wedge \theta^a_l$$
$$= k^{l+1} \alpha \wedge (u_l + k\alpha) \wedge (u_{l-1} + k u_l + k^2 \alpha) \cdots$$
$$= k^{l+1} (\exp \epsilon(k^{-1}) \cdot s^a)^\dagger \cdot \theta^a_{l+1},$$
$$u^{a\dagger}(k) \cdot \theta^a_l = k^{-l} (u_{l-1} - k u_l)(u_{l-2} - k u_{l-1}) \cdots$$
$$= k^{-l} (\exp - \epsilon(k^{-1}) \cdot s^a)^\dagger \cdot \theta^a_{l-1} . \qquad (7.1)$$

These are the relationships which yield the fundamental formulae in the Japanese papers.[2,3,7] Thus for $\alpha_\mu = \theta^1_{\mu_1} \wedge \cdots \wedge \theta^m_{\mu_m}$, $\mu = \mu_a s_a$, we get

$$u^a(k) \cdot \alpha_\mu = (-1)^\gamma k^{\mu_a + 1} \{\Gamma_+(\epsilon^{(a)}(k^{-1}))\}^\dagger \cdot \alpha_\mu + s_a ,$$
$$u^{a\dagger}(k) \cdot \alpha_\mu = (-1)^\gamma k^{-\mu_a} \{\Gamma_+(-\epsilon^{(a)}(k^{-1}))\}^\dagger \cdot \alpha_{\mu - s_a} , \qquad (7.2)$$

where $\gamma = \Sigma^{a-1}_{b=1} \mu_b$, $\epsilon^{(a)}(k) := \epsilon(k)$, and $\Gamma_+(x^{(a)}) := \exp \Sigma_{i \geq 1} x^{(a)}_i s^a(i)$, a usage introduced in the previous section. It is easy to show by considering the actions of $u^a(k)$ and $u^{a\dagger}(k)$ on θ_0 and the actions of the transported operators on v_0 that

$$u^a(k) \overset{\eta}{\mapsto} k^{1/2}X(s_a,k)\,, \quad u^{a\dagger}(k) \overset{\eta}{\mapsto} k^{-1/2}X(-s_a,k)\,.$$

Remark: There is a difference between the cocycle used in this paper and the (implied) cocycle used in the Japanese papers. This is due to a different choice of ordering for the elements defining the vacuum space of $\mathbf{s}_m$ in the spin representation. In terms of the wedge representation their choice corresponds to $\tilde{\alpha}_\mu = (-1)^p \theta^1_{\mu_1} \wedge \cdots \wedge \theta^m_{\mu_m}$, $\tilde{\alpha}_\mu \in \Omega_p$. The corresponding cocycle is obtained by reversing the inequality signs in the definition of $\epsilon(\cdot,\cdot)$ in Sec. III.

The matrix coefficients[3,4,5]

$$\begin{aligned}\tau^p_{\mu,\nu}(\Gamma_+(x);G) &:= H_\Omega(\Gamma_+(x)\cdot G\cdot \alpha_\mu,\alpha_\nu)\\ &= H_U(\Gamma_+(x)\cdot G\cdot(1\otimes e^\mu),1\otimes e^\nu)\,,\end{aligned}$$

where $l(\mu)=p=l(\nu)$ and $G\in\mathrm{GL}(m\infty)$ are called τ-functions. Let $\alpha,\beta\in\Omega$, $G\in\mathrm{GL}(m\infty)$, then a calculation shows that

$$\begin{aligned}&\sum_{\substack{j\in Z\\1\leqslant a\leqslant m}} G\cdot u^a_j\cdot\alpha\otimes G\cdot u^{a\dagger}_j\cdot\beta\\ &= \sum_{\substack{j\in Z\\1\leqslant a\leqslant m}} (G\cdot u^a_j)\cdot G\cdot\alpha\otimes(u^a_j\cdot G)^\dagger\cdot G\cdot\beta\\ &= \sum_{\substack{j\in Z\\1\leqslant a\leqslant m}} u^a_j\cdot G\cdot\alpha\otimes u^{a\dagger}_j\cdot G\cdot\beta\in\Omega\otimes\Omega\,. \qquad (7.3)\end{aligned}$$

On $\Omega\otimes\Omega$ we have the induced contravariant Hermitian form

$$H_{\Omega^2}(\omega\otimes\omega',\lambda\otimes\lambda') = H_\Omega(\omega,\lambda)H_\Omega(\omega',\lambda')\,,$$

and a similar definition applies to $H_{U^2}(\cdot,\cdot)$. Define the operators J, $J^\dagger$ on $\Omega\otimes\Omega$, where $J^\dagger$ is the adjoint of J with respect to $H_{\Omega^2}(\cdot,\cdot)$ by

$$J := \sum_{\substack{i\in Z\\1\leqslant a\leqslant m}} u^a_i\otimes u^{a\dagger}_i\,.$$

Lemma 7.1: The operators J, $J^\dagger$ commute with the representations of $\mathbf{gl}(m\infty)$, $\widehat{\mathbf{gl}}(m\infty)$ and $\widehat{\mathrm{GL}}(m\infty)$ on $\Omega\otimes\Omega$.

The proof follows from (7.3) and the action of the basis elements of $\mathbf{gl}(m\infty)$, $(\mathbf{gl}(m\infty))$, defined in the previous section, on $\Omega\otimes\Omega$.

Let $b,\mu\in C[M]$ then

$$\begin{aligned}&J\cdot(\alpha_{\mu+b}\otimes\alpha_\mu)=0,\quad \mathbf{b}\in Z^m_+\,,\\ &J^\dagger\cdot(\alpha_{\mu+b}\otimes\alpha_\mu)=0, \qquad (7.4)\\ &\mathbf{b}\in Z^m_-\,.\end{aligned}$$

These relationships yield the Hirota equations associated with this restricted system. Since Lemma 7.1 is also valid for $\Gamma_+\subset\overline{\mathrm{GL}}_0(m\infty)$ it follows that (7.4) implies the orthogonality relationship

$$\begin{aligned}&H_{\Omega^2}(J\cdot\Gamma_+(x)\cdot G\cdot(\alpha_{\mu+b}\otimes\alpha_\mu),\alpha_{\lambda+c}\otimes\alpha_\lambda)=0\,,\\ &\mathbf{b},\mathbf{c}\in Z^m_+\,. \qquad (7.5)\end{aligned}$$

Conversely since $\Gamma_+(x)\in\Gamma_+$, $G\in\widehat{\mathrm{GL}}(m\infty)$ only preserve the virtual degree of $\alpha\in\Omega$ it follows that further conditions have to be imposed to ensure that (7.5) is nontrivial. A necessary condition is that $l(\lambda)=l(\mu)-1$ and $l(c)=l(b)+2$. The condition (7.5) will then be nontrivial for some $G\in\widehat{\mathrm{GL}}(m\infty)$. The Hirota equations are easily obtained from (7.5), but we proceed differently so that the Hirota polynomials of the system can be studied.

The transported operator $\eta\cdot J\cdot\eta^{-1}$ is easily obtained,

$$\eta\cdot J\cdot\eta^{-1} = \mathrm{Res}\left\{k^{-1}\sum_{a=1}^m X(s_a,k)\otimes X(-s_a,k)\right\}.$$

The decomposition $\Omega\otimes\Omega=\bigoplus_{p,q\in Z}\Omega_p\otimes\Omega_q$ implies that $\alpha_{\mu+b}\otimes\alpha_\mu\in\Omega_p\otimes\Omega_q$ where $p=l(\mu+b),q=l(\mu)$. Let $\Omega_{\mu+b,\mu}\subset\Omega_p\otimes\Omega_q$ denote the $\mathrm{GL}(m\infty)$ orbit of $\alpha_{\mu+b}\otimes\alpha_\mu$ and choose $\phi\in\Omega_{\mu+b,\mu}$. It is convenient to introduce the notation

$$\begin{aligned}\eta:\otimes^2\Omega &\to C[x',x'']\otimes C[M'\oplus M'']\\ &\approx\otimes^2(C[x]\otimes C[M])\,,\end{aligned}$$

where $M'\oplus M''$ is the Z lattice with basis vectors $\{s_a'\}$, $\{s_a''\}$. Then η: $\phi\to\phi(x',x'';p,q)$ where

$$\phi(x',x'';p,q):= \sum_{l(\gamma)=p,\,l(\omega)=q} v_\gamma(x')w_\omega(x'')e^{\gamma'+\omega''}\,,$$

and $\gamma'=\gamma_a s_a'$ etc. The image of ϕ will only have a finite number of terms if $G\in\mathrm{GL}(m\infty)$; however there will be an arbitrary number if $G\in\overline{\mathrm{GL}}_0(m\infty)$. Lemma 7.1 and (7.4) imply that

$$\begin{aligned}0=J\cdot\phi \overset{\eta}{\mapsto} \mathrm{Res}\{&k^{-1}X(s_a',k)X(-s_a'',k)\\ &\cdot\phi(x',x'';p,q)\}=0\,. \qquad (7.6)\end{aligned}$$

The change of variables $x=\frac{1}{2}(x'+x'')$, $y=-\frac{1}{2}(x'-x'')$, $\tilde{r}_a=\frac{1}{2}(s_a'-s_a'')$, $\tilde{\tilde{r}}_a=\frac{1}{2}(s_a'+s_a'')$

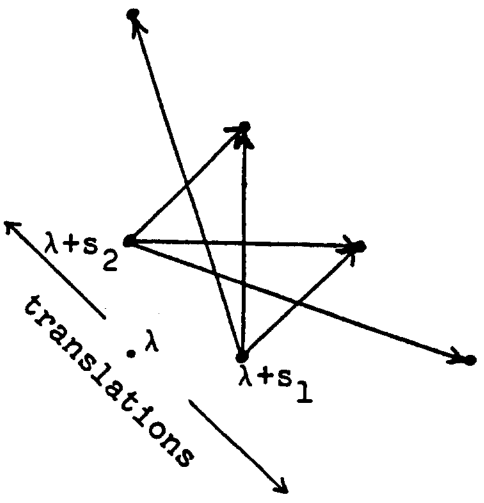

FIG. 1. Q^2_λ for the 2-($\mathbf{gl}(2\infty),\Gamma_+$) family.

gives a more natural representation of (7.6) on the space $\mathbb{C}[x,y]\otimes\mathbb{C}[\widetilde{M}\oplus\widetilde{\widetilde{M}}]$. It is important to note that the degree of an element of this module is given by

$$\deg(v(x)w(y)\otimes e^{\tilde{\gamma}+\tilde{\tilde{\omega}}})=\deg v(x)+\deg v(y)+\tfrac{1}{4}|\tilde{\gamma}|^2+\tfrac{1}{4}|\tilde{\tilde{\omega}}|^2.$$

The operator J acts on $\mathbb{C}[y]\otimes\mathbb{C}[\widetilde{M}]$,

$$J\cdot v(y)\otimes e^{\tilde{h}}=\left[\sum_{a=1}^{m}\epsilon(\tilde{r}_a,\tilde{h})\sum_{j\geqslant 0}p_j(-2y^{(a)})p_{j+h_a+1}\times(D_{y^{(a)}})\right]v(y)\otimes e^{\tilde{h}+2\tilde{r}_a}. \tag{7.7}$$

Equations (7.6) therefore have the form

$$\sum_{l(\gamma)=p,l(\omega)=q}J\cdot v_\gamma(x-y)w_\omega(x+y)\otimes\exp[(\widetilde{\gamma-\omega})+(\widetilde{\widetilde{\gamma+\omega}})]=0,$$

with J acting as in (7.7). If this relationship is projected onto $\mathbb{C}[x,y]\otimes\exp(\tilde{c}+2\lambda+c)$, where λ and c satisfy the conditions defined below (7.5), and then Taylor's formula is applied, Hirota's equations are obtained in the form,[3,7]

$$\sum_{a=1}^{m}E_{c_a-2}[D_z;y](f(x-z)g(x+z))\Big|_{z=0}=0,$$

where

$$E_{h_a}[z;y]:=(-1)^{h_1+\cdots+h_{a-1}}\sum_{j\geqslant 0}p_j(-2y^{(a)})\times p_{j+h_a+1}(z^{(a)})\exp(\hat{y}\cdot z)$$

and

$$\hat{y}\cdot z=\sum_{a=1}^{m}\hat{y}^{(a)}\cdot z^{(a)},\qquad \hat{y}^{(a)}=(y_1^{(a)},2y_2^{(a)},\ldots,ry_r^{(a)},\ldots). \tag{7.8}$$

Equations (7.8) were obtained with the functions $f(x)=v_{c+\lambda-s_a}(x)=\tau^{p}_{\mu+b,c+\lambda-s_a}(\Gamma_+(x);G)$, $g(x)=w_{\lambda+s_a}(x)=\tau^{q}_{\mu,\lambda+s_a}(\Gamma_+(x);G)$, however their functional form only depends upon $c\in Z^m_+$ such that $l(c)-2=l(b)\geqslant 0$. For a given $s\geqslant 0$ the set of equations (7.8) such that $c\in Z^m_+$ satisfies $l(c)=s+2$ we shall refer to as the s-$(\mathbf{gl}(m\infty),\Gamma_+)$ family of Hirota equations. In contrast to the $m=1$ case, (7.8) for a given s, furnishes a hierarchy of equations for each allowable c and the word family seems more appropriate in this case.

Define the totality of Hirota polynomials by the generating functions

$$\mathrm{Hir}[z;y]:=\sum_{s\geqslant 0}\mathrm{Hir}_s[z;y],$$

where

$$\mathrm{Hir}_s[z;y]:=\sum_{\substack{c\in Z^m_+\\ :l(c)=s+2}}\sum_{a=1}^{m}E_{c_a-2}[z;y]\otimes e^{\tilde{c}-2\tilde{r}_a}. \tag{7.9}$$

Let $\gamma\in\mathrm{Par}$ then $\mathrm{Hir}_s[z;y]$ consists of sums of the form $(y^{(a)})^\gamma P_{\gamma;c_a-1}(z)\otimes e^{\tilde{c}-2\tilde{r}_a}$ over all possible $\gamma\in\mathrm{Par}$ and c such that $l(c)=s+2$. We also have $\deg P_{\gamma;c_a-1}(z)\otimes e^{\tilde{c}-2\tilde{r}_a}=l(\gamma)+\tfrac{1}{4}|\tilde{c}|^2$.

The s-$(\mathbf{gl}(m\infty),\Gamma_+)$ family of equations involves the dependent variables $v_{c+\lambda-s_a}$, $w_{\lambda+s_a}$, $c\in Z^m_+$, $l(c)=s+2$ and $\lambda\in M$ such that $l(\lambda)=l(\mu)-1$. Choose a $\lambda\in M$ which satisfies this condition then a representation of the dependent variables is given in the following way.

Let

$$Q^s_\lambda=\{\{((\lambda+s_a),(c+\lambda-s_a)):\ a=1,\ldots,m\}\ c\in M:\ l(c)=s+2\},$$

Q^s_λ can be considered as a directed graph by joining $(\lambda+s_a)$ to $(c+\lambda-s_a)$ with the arrow directed towards the second point.

Definition: A skeleton Q^s_λ labels the minimal set of dependent variables which define the s-$(\mathbf{gl}(m\infty),\Gamma_+)$ family.

In general Eqs. (7.8) are differential-difference equations. Fix $\lambda\in M$, $l(\lambda)=l(\mu)-1$ then all the other dependent variables can be obtained by translations of the skeleton

$$\mu:Q^s_\lambda\to Q^s_{\lambda+\mu},$$

where $\mu\in C[M]$ and $l(\mu)=0$. Figure 1 depicts the case for the 2-$(\mathbf{gl}(2\infty),\Gamma_+)$ family.

ACKNOWLEDGMENTS

This work was carried out during the tenure of a JSPS fellowship. It is a pleasure to thank colleagues, students, and staff at the Institute for their kindness and hospitality and in particular my host scientist Takashi Sasaki for easing and solving the many problems which this gaijin experienced in early dealings with the charming and generous people of Japan.

[1] A. V. Mikhailov and A. B. Shabat, "Integrability conditions for the system of two equations of the form $u_t=A(u)u_{xx}+F(u,u_x)$. I," Theor. Math. Phys. **62**, 107 (1985).

[2] A. V. Mikhailov and A. B. Shabat, "Integrability conditions for the system of two equations of the form $u_t=A(u)u_{xx}+F(u,u_x)$. II," Theor. Math. Phys. **66**, 31 (1986).

[3] M. Jimbo and T. Miwa, "Solitons and infinite dimensional Lie algebras," R.I.M.S., Kyoto Univ. **19**, 943 (1983).

[4] E. Date, M. Jimbo, M. Kashiwara, and T. Miwa, "Transformation groups for soliton equations—Euclidean Lie algebras and reduction of the KP hierarchy," R.I.M.S., Kyoto Univ. **18**, 1077 (1982).

[5] E. Date, M. Jimbo, M. Kashiwara, and T. Miwa, "Transformation groups for soliton equations," in *Nonlinear Integrable Systems—Classical Theory and Quantum Theory*, edited by M. Jimbo and T. Miwa (World Scientific, Singapore, 1983).

[6] E. Date, M. Kashiwara, and T. Miwa, "Vertex operators and functions. Transformation groups for soliton equations. II," Proc. Jpn. Acad. **57**, 387 (1981).

[7] E. Date, M. Jimbo, M. Kashiwara, and T. Miwa, "Operator approach to the Kadomtsev–Petviashvili equation," J. Phys. Soc. Jpn. **50**, 3806 (1981).

[8] M. Sato and Y. Sato, "Soliton equations as dynamical systems on infinite dimensional Grassmann manifolds." Lect. Notes Num. Appl. Anal. **5**, 259 (1982).

[9] E. Date, M. Jimbo, and T. Miwa, "Method for generating discrete soliton equations. I," J. Phys. Soc. Jpn. **51**, 4116 (1982).

[10] V. G. Kac and D. H. Peterson, "Constructions of the basic representation of the loop group of E_8" in *Symposium on Anomalies, Geometry and Topology*, edited by W. A. Bardeen and A. R. White (World Scientific, Singapore, 1985).

[11] I. B. Frenkel and V. G. Kac, "Basic representations of affine Lie algebras and dual resonance models," Invent. Math. **62**, 23 (1980).
[12] V. G. Kac and D. H. Peterson, "Lectures on the infinite wedge representation and the MKP hierarchy," Lecture notes at the summer school on completely integrable systems, Montreal, Canada (August 1985).
[13] J. Lepowsky, "Calculus of twisted vertex operators," Proc. Natl. Acad. Sci. U.S.A. **82**, 8295 (1985).
[14] R. Hirota, "Direct methods in soliton theory," in *Solitons*, edited by R. K. Bullough and P. G. Caudrey (Springer-Verlag, Berlin, 1980).
[15] R. K. Dodd, "Classification of a restricted class of solvable equations associated with $\mathbf{gl}_\infty$" (in preparation).
[16] V. Kac, *Infinite Dimensional Lie Algebras* (Cambridge U.P., Cambridge, England, 1985), 2nd ed.
[17] I. B. Frenkel, "Representations of affine Lie algebras, Hecke modular forms and Korteweg–de Vries type equation," *Springer Lecture Notes in Mathematics*, Vol. 933, edited by A. Dold and B. Echmann (Springer, Berlin, 1981), pp. 71–110.
[18] J. Lepowsky and R. L. Wilson, "The structure of standard modules, I: Universal algebras and the Rogers–Ramanujan identities," Invent. Math. **77**, 199 (1984).
[19] G. Segal and G. Wilson, "Loop groups and equations of KdV type," I.H.E.S. 5 (1985).
[20] A. Pressley and G. Segal, "Loop groups," in *Oxford Mathematics Monographs* (Oxford U.P., Oxford, 1985).
[21] I. G. MacDonald, *Symmetric Functions and Hall Polynomials* (Oxford U.P., Oxford, 1979).

Representations of the braid group obtained from quantum sl(3) enveloping algebra

Zhong-Qi Ma[a)]
CCAST (World Laboratory), P. O. Box 8730, Beijing, People's Republic of China and International Centre for Theoretical Physics, Trieste, Italy

(Received 20 July 1989; accepted for publication 11 October 1989)

The quantum Clebsch–Gordan (CG) coefficients for the coproduct 6×6 of the quantum sl(3) enveloping algebra are computed. Based on the representation 6, the representation of the braid group and the corresponding link polynomial are obtained. The link polynomials based on the representations of the quantum sl(3) enveloping algebra with a one-row Young tableau are discussed.

I. INTRODUCTION

Two years ago, Akutsu and Wadati[1] "found an unexpected close connection between physics and mathematics," that is, they found out the similarity between the quantum Yang–Baxter equation (QYBE) and the multiplication rules of the braid group.[2] Taking the spectral parameter u in QYBE to be infinity, they obtained a set of representations of the braid group from the Boltzmann weights of the N-state models, which are the solutions of QYBE, in terms of normalizing, the symmetry breaking transformation and the limit process of $u\to\infty$. Then, they successfully found a set of link polynomials. This method was discussed independently by Jones.

On the other hand, from the trigonomatric solution[3] of the classical Yang–Baxter equation (CYBE)

$$\begin{aligned} r &= -C_0-2C_- \\ &= -\sum_j H_j\times H_j - 2\sum_{\alpha\in\Delta_+} E_{-\alpha}\times E_\alpha, \end{aligned} \tag{1}$$

where H_j and E_α are the Cartan bases, and Δ_+ is the set of positive roots, we computed, based on the quantum sl(2) enveloping algebra (q-sl(2)), the explicit forms of $\check{R}_q$ matrix,[4,5] satisfying QYBE without the spectral parameter, which is the same as the multiplication rule of the braid group. We found out that the solutions $\check{R}_q$ we got are just the same, up to an unimportant factor, as the $R_{\rm AW}$ matrix of Akutsu and Wadati. Thus we have obtained the explicit form of $R_{\rm AW}$ with any j (or $N=2j+1$).

In our computation for q-sl(2),[4] we used the same notations as those in the theory of angular momentum, and computed the analogies of CG coefficients, Racah coefficients, $3j$ symbol, and $6j$ symbol, step by step. Unfortunately, we are not able to compute quantum CG coefficients (q-CG) for the quantum sl(3) enveloping algebra [$q-$ sl(3)] generally, because even in su(3) Lie algebra, CG coefficients were computed only for some definite representations.[6]

Encouraged by the success of computation on q-sl(2), we are interested in generalizing this method to q-sl(3), although we are able to compute only for some definite representations. Here, $\check{R}_q$ matrices based on the fundamental representations of q-sl(n) were calculated[7] and proved to satisfy the Hecke algebra. The fusion of the fundamental representations was discussed.[8] Recently, the operator form of $\check{R}_q$ matrix for $q-$ sl(n) was given.[9] But, in order to construct link polynomials, more explicit forms of $\check{R}_q$ matrices are needed.

In this paper, we compute the q-CG coefficients for the coproduct 6×6, then obtain the $\check{R}_q$ matrix and construct the corresponding link polynomial. Throughout this paper, we assume that q is not a root of one so that all the finite irreducible representations (IR) are nonsingular. In Sec. II, we review q-sl(3) briefly. In Sec. III, we make some conventions on enumerating the states of the irreducible representation, and their relative phases. For the convenience for q-sl(3), we have to change some conventions used in su(3) Lie algebra.[6] The q-CG coefficients and relevent representation matrices are computed in Sec. IV, and then, the $\check{R}_q$ matrix in Sec. V. In terms of the general methods,[1,10,11] we construct the link polynomials in Sec. VI. In Sec. VII, some discussions are given for the representations of $q-$ sl(3) with one row Young tableau. We left the computation on the coproduct 8×8 of q-sl(3) in the next paper.

II. QUANTUM sl (3) ENVELOPING ALGEBRA

Deform the generators λ_a of su(3) algebra to those of q-sl(3):

$$\begin{aligned} &2\lambda_3\to h_1,\quad \sqrt{3}\lambda_8-\lambda_3\to h_2,\quad \lambda_1+i\lambda_2\to e_1,\\ &\lambda_1-i\lambda_2\to f_1,\quad \lambda_6+i\lambda_7\to e_2,\quad \lambda_6-i\lambda_7\to f_2, \end{aligned} \tag{2}$$

which satisfy the following multiplication rules:

$$\begin{aligned} &k_a=q^{h_a/2},\\ &k_ae_a=qe_ak_a,\quad k_ae_b=q^{-1/2}e_bk_a,\\ &k_af_a=q^{-1}f_ak_a,\quad k_af_b=q^{1/2}f_bk_a,\\ &[k_1,k_2]=[e_1,f_2]=[e_2,f_1]=0,\\ &[e_a,f_a]=(k_a^2-k_a^{-2})/(q-q^{-1}),\\ &e_a^2e_b-(q+q^{-1})e_ae_be_a+e_be_a^2=0,\\ &f_a^2f_b-(q+q^{-1})f_af_bf_a+f_bf_a^2=0, \end{aligned} \tag{3}$$

where a, $b=1$, 2, and $a\neq b$.

The explicit forms of generators in the fundamental representation of q-sl(3) are given as follows:

[a)] On leave from the Institute of High Energy Physics, P. O. Box 918(4), Beijing 100039, People's Republic of China.

 0022-2488/90/030550-07$03.00

$$h_1 = \mathrm{diag}(1,-1,0), \quad h_2 = \mathrm{diag}(0,1,-1),$$
$$e_1 = \tilde{f}_1 = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad e_2 = \tilde{f}_2 = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{pmatrix}, \tag{4}$$

where and in the rest of this paper the tilde denotes transpose.

Following the nomenclature in the SU(3) theory, we call the subalgebra q-sl(2) spanned by h_1, e_1, and f_1 q-isospin and that spanned by h_2, e_2, and f_2 q-U-spin, and use q-isospin and q-superspin Y to assign the states in IR:

$$\begin{aligned} &I_3 = h_1/2, \quad Y = (h_1 + 2h_2)/3, \\ &I^2 = (k_1 - k_1^{-1})(qk_1 - q^{-1}k_1^{-1})/(q-q^{-1})^2 + f_1 e_1. \end{aligned} \tag{5}$$

III. CONVENTIONS FOR IRREDUCIBLE REPRESENTATIONS

We choose the bases of an IR of q-sl(3) so that h_1, h_2, and I^2 are diagonal, namely, each state is the common eigenstate of I^2, I_3, and Y:

$$\begin{aligned} &I^2|I,I_3,Y\rangle = [I][I+1]|I,I_3,Y\rangle, \\ &h_1|I,I_3,Y\rangle = 2I_3|I,I_3,Y\rangle, \\ &h_2|I,I_3,Y\rangle = (3Y/2 - I_3)|I,I_3,Y\rangle, \end{aligned} \tag{6}$$

where and throughout this paper

$$[m] = (q^m - q^{-m})/(q - q^{-1}). \tag{7}$$

Because q is not a root of one, we have

$$[m] \neq 0 \quad \text{if } m \neq 0.$$

Now, we enumerate the states of an IR from 1 to N as the following order.[6]

(a) The state with the highest weight, i.e., with the highest eigenvalue of Y, and the highest eigenvalue of I_3 among the states with the same highest eigenvalue of Y, is enumerated by one.

(b) The states with the same Y and I_3 are ordered so that I decreases.

(c) The group of states for the same Y but different I_3 are ordered such that I_3 decreases.

(d) The group of states for different Y are ordered such that Y decreases.

Sometimes, we denote an IR $[\lambda_1,\lambda_2]$ of q-sl(3) by its dimension N. For example, the fundamental representation [1,0] is denoted by 3, [1,1] by 3*, [2,0] by 6, [2,2] by 6*, [4,0] by 15, and [3,1] by 15′. Note not to confuse the notation of an IR $[\lambda_1,\lambda_2]$ with [m] in (7). The enumerations of the states of the relevant representations are listed in Fig. 1.

In this paper, we use some different notations to denote the states of IR. When we emphasize the eigenvalues (weight), for example in (6), we use the notation $|I,I_3,Y\rangle$, but in the usual case, in order to reduce the notation we use the enumerated number m to denote a state, for example, $|[2,0]m\rangle$ or $|6,m\rangle$ to denote a state in IR [2,0].

To be able to define later uniquely q-CG coefficients of q-sl(3), it is necessary first of all to define precisely the relative phases of the states within an IR. For convenience we adopt a little different conventions as those adopted in su(3).[6]

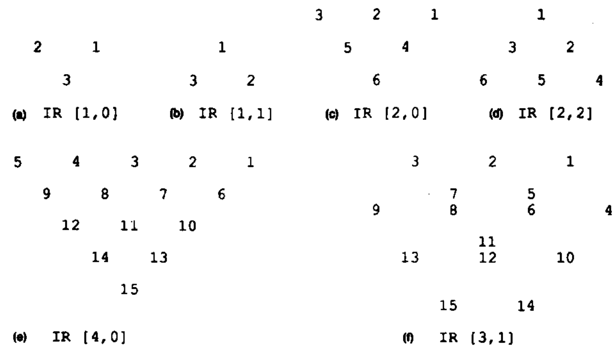

FIG. 1. The enumerations of the eigenstates of IR of q-sl(3).

(i) The relative phases within a definite q-isomultiplet are determined by q-sl(2) phase convention[4]:

$$\begin{aligned} &e_1|I,I_3,Y\rangle = \Gamma^I_{-I_3}(q)|I,I_3+1,Y\rangle, \\ &f_1|I,I_3,Y\rangle = \Gamma^I_{I_3}(q)|I,I_3-1,Y\rangle, \\ &\Gamma^j_m(q) = \{[j+m][j-m+1]\}^{1/2}. \end{aligned} \tag{8}$$

(ii) The relative phases between the different q-isomultiplets are determined so that the matrix elements of e_2 and f_2 are non-negative when q is positive real.

(iii) The coproduct of two IR's is defined as

$$\begin{aligned} k_a(|N_1,m_1\rangle|N_2,m_2\rangle) &= (k_a|N_1,m_1\rangle)(k_a|N_2,m_2\rangle), \\ e_a(|N_1,m_1\rangle|N_2,m_2\rangle) &= (e_a|N_1,m_1\rangle)(k_a^{-1}|N_2,m_2\rangle) \\ &\quad + (k_a|N_1,m_1\rangle)(e_a|N_2,m_2\rangle), \end{aligned} \tag{9}$$

and that replaced e_a by f_a.

(iv) The q-CG matrix is real orthogonal when q is positive real, and orthogonal for any q.

As an example, we calculate representation matrices of IR [2,0] of q-sl(3) and the q − CG coefficients for the decomposition of coproduct $[1,0]\times[1,0] = [2,0] + [1,1]$ ($3\times 3 = 6 + 3^*$). The states in IR's [1,0], [2,0], and [1,1] are denoted by $|m\rangle$, $|6,m\rangle$ and $|3^*,m\rangle$, respectively.

The state with the lowest weight of IR [2,0] is simple and can be expressed as a direct product of $|3\rangle$ of IR [1,0]:

$$|6,6\rangle = |3\rangle|3\rangle. \tag{10a}$$

According to the definition of coproduct (9) and the representation matrices (6) of q-sl(2), we have

$$\begin{aligned} |6,5\rangle &= [2]^{-1/2}e_2|6,6\rangle \\ &= [2]^{-1/2}\{q^{1/2}|2\rangle|3\rangle + q^{-1/2}|3\rangle|2\rangle\}, \\ |6,4\rangle &= e_1|6,5\rangle = [2]^{-1/2}\{q^{1/2}|1\rangle|3\rangle \\ &\quad + q^{-1/2}|3\rangle|1\rangle\}, \\ |6,3\rangle &= [2]^{-1/2}e_2|6,5\rangle = |2\rangle|2\rangle, \\ |6,2\rangle &= [2]^{-1/2}e_1|6,3\rangle \\ &= [2]^{-1/2}\{q^{1/2}|1\rangle|2\rangle + q^{-1/2}|2\rangle|1\rangle\}, \\ |6,1\rangle &= [2]^{-1/2}e_1|6,2\rangle = |1\rangle|1\rangle. \end{aligned} \tag{10b}$$

From the requirement of orthogonality of states, we get the state with the lowest weight of IR [1,1]

$$|3^*,3\rangle = [2]^{-1/2}\{q^{-1/2}|2\rangle|3\rangle - q^{1/2}|3\rangle|2\rangle\} \tag{11a}$$

then,

$$\begin{aligned} |3^*,2\rangle &= e_1|3^*,3\rangle = [2]^{-1/2}\{q^{-1/2}|1\rangle|3\rangle \\ &\qquad - q^{1/2}|3\rangle|1\rangle\}, \\ |3^*,1\rangle &= e_2|3^*,2\rangle = [2]^{-1/2}\{q^{-1/2}|1\rangle|2\rangle \\ &\qquad - q^{1/2}|2\rangle|1\rangle\}. \end{aligned} \tag{11b}$$

Equations (10) and (11) give both representation matrices of IR [2,0] and q-CG coefficients:

$$\begin{aligned} &D_q^6(k_1) = \mathrm{diag}(q,1,q^{-1},q^{1/2},q^{-1/2},1), \\ &D_q^6(k_2) = \mathrm{diag}(1,q^{1/2},q,q^{-1/2},1,q^{-1}), \\ &D_q^6(e_1)_{12} = D_q^6(e_1)_{23} = D_q^6(e_2)_{35} = D_q^6(e_2)_{56} = [2]^{1/2} \\ &D_q^6(e_1)_{45} = D_q^6(e_2)_{24} = 1, \end{aligned} \tag{12}$$

the rest of matrix elements are vanishing, and the matrices of f_a are the transpose of those of e_a:

$$(C_q^{33})_6 = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & A & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & A & 0 & 0 \\ 0 & B & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & A & 0 \\ 0 & 0 & 0 & B & 0 & 0 \\ 0 & 0 & 0 & 0 & B & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix},$$

$$(C_q^{33})_{3^*} = \begin{pmatrix} 0 & 0 & 0 \\ B & 0 & 0 \\ 0 & B & 0 \\ -A & 0 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & B \\ 0 & -A & 0 \\ 0 & 0 & -A \\ 0 & 0 & 0 \end{pmatrix}, \tag{13}$$

where $A = \{q/[2]\}^{1/2}$ and $B = \{q[2]\}^{-1/2}$. The rows of q-CG are ordered by $m_1 m_2$ of two fundamental representations.

IV. DECOMPOSITION OF COPRODUCT 6×6 IN q-sl(3)

The calculation for the decomposition of coproduct 6×6 in q-sl(3) is straightforward but tedious. The key is to determine the states with multiple weights, i.e., the states with the same Y and I_3, by the requirement of orthogonality. Most of the representation matrices can be obtained directly from the enumerations of the states and the representation matrices of q-sl(2). The matrices of f_a are the transpose of those of e_a. What is only needed to list is the matrix of e_2 in IR [3,1], because there are multiple weights in this IR. The nonvanishing matrix elements of e_2 in IR [3,1] are listed in Table I.

The rows of the q-CG matrix for 6×6 in q-sl(3) are ordered by m_1 and m_2 which both go from 1 to 6, and the columns are denoted by $(15,m)$, $(15',m)$ and $(6^*,m)$ for the different IR's, respectively. The q-CG matrix is a block matrix, and some submatrices are equal to each other. The nonvanishing matrix elements are listed in Table II. The equal submatrices are listed in the same table and distinguished by (a), (b), and so on.

V. REPRESENTATIONS OF BRAID GROUP

Now, we are going to compute the representation matrix of the generator b_i of the braid group

$$D(b_i) = \mathbb{1}\times\cdots\times\mathbb{1}\times\check{R}_q\times\mathbb{1}\cdots\times\mathbb{1}, \tag{14}$$

where $\check{R}_q$ is located in the ith and $(i+1)$th positions in the direct product. The $\check{R}_q$ matrix is defined as follows.[4]

(a) $\check{R}_q$ has r as its classical limit

$$\check{R}_q \sim \{\mathbb{1} + (q-1)r\}P,$$

where P is the transpose, and r is the solution of classical YBE:

$$\begin{aligned} r = &-(h_1\times h_1 + 3Y\times Y)/2 \\ &- 2(f_1\times e_1 + f_2\times e_2 + [f_2,f_1]\times[e_1,e_2]), \end{aligned} \tag{15}$$

where the quantum parameter $q = 1$.

(b) $\check{R}_q^{N_1N_2} D_q^{N_1N_2} = D_q^{N_2N_1}\check{R}_q^{N_1N_2}$,

TABLE I. The nonvanishing matrix elements of e_2 in IR [3,1].

Row	1	2	5	2	6	3	3	7
Column	4	5	10	6	10	7	8	11
$D_q(e_2)$	1	$\left\{\frac{[2]}{[3]}\right\}^{1/2}$	$\left\{\frac{[2]}{[3]}\right\}^{1/2}$	$\left\{\frac{[4]}{[3]}\right\}^{1/2}$	$\left\{\frac{[4]}{[3]}\right\}^{1/2}$	$\frac{1}{[3]^{1/2}}$	$\left\{\frac{[4][2]}{[3]}\right\}^{1/2}$	$\frac{[2]}{[3]^{1/2}}$
Row	8	8	11	12	9	13		
Column	11	12	14	14	13	15		
$D_q(e_2)$	$\left\{\frac{[4]}{[3][2]}\right\}^{1/2}$	$[3]^{1/2}$	1	$\left\{\frac{[4]}{[2]}\right\}^{1/2}$	$[2]^{1/2}$	$[2]^{1/2}$		

TABLE II. Nonvanishing matrix elements of $q-$ CG for 6×6 in $q-\mathrm{sl}(3)$.

Row \ Column	(a) (15,1)	(b) (15,5)	(c) (15,15)
(a) 1 1			
(b) 3 3		1	
(c) 6 6			

Row \ Column	(a) (15,2) (c) (15,4) (e) (15,13)	(b) (15,6) (d) (15,9) (f) (15,14)	(a) (15′,1) (c) (15′,3) (e) (15′,14)	(b) (15′,4) (d) (15′,9) (f) (15′,15)
(a) 12 (b) 14 (c) 23 (d) 35 (e) 46 (f) 56	$q\left\{\frac{[2]}{[4]}\right\}^{1/2}$		$q^{-1}\left\{\frac{[2]}{[4]}\right\}^{1/2}$	
(a) 21 (b) 41 (c) 32 (d) 53 (e) 64 (f) 65	$q^{-1}\left\{\frac{[2]}{[4]}\right\}^{1/2}$		$-q\left\{\frac{[2]}{[4]}\right\}^{1/2}$	

Row \ Column	(a) (15,3) (b) (15,10) (c) (15,12)	(a) (15′,2) (c) (15′,10) (c) (15′,13)	(a) (6*,1) (b) (6*,4) (c) (6*,6)
(a) 13 (b) 16 (c) 36	$q^2\left\{\frac{[2]}{[4][3]}\right\}^{1/2}$	$\left\{\frac{[2]}{[4]}\right\}^{1/2}$	$q^{-1}[3]^{-1/2}$
(a) 22 (b) 44 (c) 55	$[2]\left\{\frac{[2]}{[4][3]}\right\}^{1/2}$	$(q^{-1}-q)\left\{\frac{[2]}{[4]}\right\}^{1/2}$	$-[3]^{-1/2}$
(a) 31 (b) 61 (c) 63	$q^{-2}\left\{\frac{[2]}{[4][3]}\right\}^{1/2}$	$-\left\{\frac{[2]}{[4]}\right\}^{1/2}$	$q[3]^{-1/2}$

Row \ Column	(15,7)	(15′,5)	(15′,6)	(6*,2)
15	$q^2\left\{\frac{[2]}{[4][3]}\right\}^{1/2}$	$\left\{\frac{[2]}{[4][3]}\right\}^{1/2}$	$[3]^{-1/2}$	$q^{-1}[3]^{-1/2}$
24	$\frac{q^{1/2}[2]}{([4][3])^{1/2}}$	$\frac{q^{-3/2}[2]}{([4][3])^{1/2}}$	$\frac{-q^{3/2}}{([3][2])^{1/2}}$	$\frac{-q^{1/2}}{([3][2])^{1/2}}$
42	$\frac{q^{-1/2}[2]}{([4][3])^{1/2}}$	$\frac{-q^{3/2}[2]}{([4][3])^{1/2}}$	$\frac{q^{-3/2}}{([3][2])^{1/2}}$	$\frac{-q^{-1/2}}{([3][2])^{1/2}}$
51	$q^{-2}\left\{\frac{[2]}{[4][3]}\right\}^{1/2}$	$-\left\{\frac{[2]}{[4][3]}\right\}^{1/2}$	$-[3]^{-1/2}$	$q[3]^{-1/2}$

Row \ Column	(15,8)	(15′,7)	(15′,8)	(6*,3)
25	$\frac{q^{3/2}[2]}{([4][3])^{1/2}}$	$\frac{q^{-1/2}[2]}{([4][3])^{1/2}}$	$\frac{q^{-1/2}}{([3][2])^{1/2}}$	$\frac{q^{-3/2}}{([3][2])^{1/2}}$
34	$\left\{\frac{[2]}{[4][3]}\right\}^{1/2}$	$q^{-2}\left\{\frac{[2]}{[4][3]}\right\}^{1/2}$	$-q[3]^{-1/2}$	$-[3]^{-1/2}$
43	$\left\{\frac{[2]}{[4][3]}\right\}^{1/2}$	$-q^2\left\{\frac{[2]}{[4][3]}\right\}^{1/2}$	$q^{-1}[3]^{-1/2}$	$-[3]^{-1/2}$
52	$\frac{q^{-3/2}[2]}{([4][3])^{1/2}}$	$\frac{-q^{1/2}[2]}{([4][3])^{1/2}}$	$\frac{-q^{1/2}}{([3][2])^{1/2}}$	$\frac{q^{3/2}}{([3][2])^{1/2}}$

Row \ Column	(15,11)	(15′,11)	(15′,12)	(6*,5)
26	$q^2\left\{\frac{[2]}{[4][3]}\right\}^{1/2}$	$\left\{\frac{[2]}{[4]}\right\}^{1/2}$	0	$q^{-1}[3]^{-1/2}$
45	$\frac{q^{1/2}[2]}{([4][3])^{1/2}}$	$\frac{q^{-1/2}-q^{3/2}}{[4]^{1/2}}$	$(q[2])^{-1/2}$	$\frac{-q^{1/2}}{([3][2])^{1/2}}$
54	$\frac{q^{-1/2}[2]}{([4][3])^{1/2}}$	$\frac{q^{-3/2}-q^{1/2}}{[4]^{1/2}}$	$-\left\{\frac{q}{[2]}\right\}^{1/2}$	$\frac{-q^{-1/2}}{([3][2])^{1/2}}$
62	$q^{-2}\left\{\frac{[2]}{[4][3]}\right\}^{1/2}$	$-\left\{\frac{[2]}{[4]}\right\}^{1/2}$	0	$q[3]^{-1/2}$

where $D_q^{N_1N_2}$ is the coproduct of IR's N_1 and N_2, in our case $N_1=N_2=6$, and usually, we will omit the superscripts of $\check{R}_q$.

(c) $(\check{R}_q)^{-1}=P\check{R}_{q^{-1}}P.$

From the definition, it can be proven that $\check{R}_q$ satisfies the QYBE without the spectral parameter, and can be expressed as

$$\check{R}_q=\sum_N \xi_N q^{\eta(N_1,N_2,N)}(C_q^{N_2N_1})_N\widetilde{(C_q^{N_1N_2})}_N, \tag{16}$$

where $N=15$, $15'$, and 6^* when $N_1=N_2=6$, C_q is the q-CG matrix given in the previous section, ξ_N is the symmetry of the CG coefficients of su(3) for exchanging N_1 and N_2 to each other, and given as

$$\xi_{15}=\xi_{6^*}=1,\quad \xi_{15'}=-1, \tag{17}$$

and $\eta(N_1,N_2,N)$ is calculated in the classical level[4,11]

$$\begin{aligned}\eta(N_1,N_2,N)&=\widetilde{(C^{N_2N_1})}_N r(C^{N_2N_1})_N\\&=C_2(N_1)+C_2(N_2)-C_2(N),\end{aligned} \tag{18}$$

where C_2 is the Casimir operator and calculated for su(3) in the Appendix:

$$C_2([\lambda_1,\lambda_2])=(\lambda_1^2-\lambda_1\lambda_2+\lambda_2^2+3\lambda_1)/3. \tag{19}$$

In our case,

$$C_2(3)=C_2(3^*)=4/3,\quad C_2(6)=C_2(6^*)=10/3,$$
$$C_2(15)=28/3,\quad C_2(15')=16/3.$$

To make the expression more simple, we remove a factor of q from the definition of $\check{R}_q$ such that the term related to the IR, the first row of whose Young tableau is the longest and whose dimension is denoted by N_0, has factor one, i.e., define η' instead of η:

$$\eta'(N)=C_2(N_0)-C_2(N). \tag{18′}$$

Therefore, after removing a factor $q^{-2/3}$ for the fundamental representation, we have

$$\check{R}_q^{33}=(C_q)_6\widetilde{(C_q)}_6-q^2(C_q)_{3^*}\widetilde{(C_q)}_{3^*}, \tag{20}$$

and after removing a factor $q^{-8/3}$ for the IR [2,0], we have

$$\begin{aligned}\check{R}_q=(C_q)_{15}\widetilde{(C_q)}_{15}-q^4(C_q)_{15'}(C_q)_{15'}\\+q^6(C_q)_{6^*}\widetilde{(C_q)}_{6^*}.\end{aligned} \tag{21}$$

Obviously, $\check{R}_q^{33}$ satisfies the Hecke algebra

$$(\check{R}_q^{33}-1)(\check{R}_q^{33}+q^2)=0 \tag{22}$$

and $\check{R}_q$ satisfies

$$(\check{R}_q-1)(\check{R}_q+q^4)(\check{R}_q-q^6)=0. \tag{23}$$

Through the straightforward calculation, we obtain

$$\check{R}_q^{33}=\begin{pmatrix}1&0&0&0&0&0&0&0&0\\0&0&0&q&0&0&0&0&0\\0&0&0&0&0&0&q&0&0\\0&q&0&A&0&0&0&0&0\\0&0&0&0&1&0&0&0&0\\0&0&0&0&0&0&0&q&0\\0&0&q&0&0&0&A&0&0\\0&0&0&0&0&q&0&A&0\\0&0&0&0&0&0&0&0&1\end{pmatrix}, \tag{24}$$

where $A=1-q^2$. Obviously, this $\check{R}_q^{33}$ matrix coincides with the previous result.[7,8]

The $\check{R}_q$ matrix based on IR [2,0] of q-sl(3) is also a block matrix and some submatrices are equal to each other. The calculation results are listed in Table III. The equal submatrices are listed in the same table, and distinguished by (a), (b), and so on.

VI. LINK POLYNOMIALS

Substituting the $\check{R}_q$ matrix into (14), we obtain the representation of generators of the braid group, then the representation $D(B,n)$ of any element B of the braid group B_n with n strands. Define a direct product matrix V of n matrices v

$$V=v\times v\times\cdots\times v. \tag{25}$$

For the fundamental representation [1,0] v is a 3×3 diagonal matrix

TABLE III. $\check{R}_q$ matrix based on IR [2,0] of $q-\mathrm{sl}(3)$.

Row \ Column	(a) 11 (b) 33 (c) 66
(a) 11 (b) 33 (c) 66	1

Row \ Column	(a) 12 (b) 14 (c) 23 (d) 35 (e) 46 (f) 56	(a) 21 (b) 41 (c) 32 (d) 53 (e) 64 (f) 65
(a) 12 (b) 14 (c) 23 (d) 35 (e) 46 (f) 56	0	q^2
(a) 21 (b) 41 (c) 32 (d) 53 (e) 64 (f) 65	q^2	$1-q^4$

Row \ Column	(a) 13 (b) 16 (c) 36	(a) 22 (b) 44 (c) 55	(a) 31 (b) 61 (c) 63
(a) 13 (b) 16 (c) 36	0	0	q^4
(a) 22 (b) 44 (c) 55	0	q^2	$q(1-q^4)$
(a) 31 (b) 61 (c) 63	q^4	$q(1-q^4)$	$1-q^2-q^4+q^6$

Row \ Column	(a) 15 (b) 26	(a) 24 (b) 45	(a) 42 (b) 54	(a) 51 (b) 62
(a) 15 (b) 26	0	0	0	q^4
(a) 24 (b) 45	0	0	q^3	$q^2(1-q^2)(1+q^2)^{1/2}$
(a) 42 (b) 54	0	q^3	$q^2(1-q^2)$	$q(1-q^2)(1+q^2)^{1/2}$
(a) 51 (b) 62	q^4	$q^2(1-q^2)(1+q^2)^{1/2}$	$q(1-q^2)(1+q^2)^{1/2}$	$1-q^2-q^4+q^6$

Row \ Column	25	34	43	52
25	0	0	0	q^3
34	0	0	q^4	$q^2(1-q^2)(1+q^2)^{1/2}$
43	0	q^4	0	$q^2(1-q^2)(1+q^2)^{1/2}$
52	q^3	$q^2(1-q^2)(1+q^2)^{1/2}$	$q^2(1-q^2)(1+q^2)^{1/2}$	$1-2q^4+q^6$

$$v=\mathrm{diag}(q^{-2},1,q^2)/(q^{-2}+1+q^2), \tag{26a}$$

and for IR [2,0], v is a 6×6 diagonal matrix

$$v=\mathrm{diag}(q^{-4},q^{-2},1,1,q^2,q^4)/(q^{-4}+q^{-2}+2+q^2+q^4). \tag{26b}$$

It is easy to check that

$$\sum_n \{(1\times v)\check{R}_q\}_{mn,m'n}=\delta_{mm'}\tau,$$
$$\sum_n \{(1\times v)\check{R}_q^{-1}\}_{mn,m'n}=\delta_{mm'}\bar{\tau}, \tag{27}$$

where for the fundamental representation

$$\tau=(1+q^2+q^4)^{-1},\quad \bar{\tau}=q^4\tau \tag{28}$$

and for IR [2,0]

$$\tau=(1+q^2+2q^4+q^6+q^8)^{-1},\quad \bar{\tau}=q^8\tau. \tag{29}$$

Thus the link polynomials defined as follows are invariant under the Markov moves:

$$\alpha(B,n)=(\tau\bar{\tau})^{-(n-1)/2}(\bar{\tau}/\tau)^{e(B)/2}\,\mathrm{Tr}\{VD(B,n)\}, \tag{30}$$

where $e(B)$ is the exponential sum of the generators in B.[1]

For the fundamental representation [1,0], the Alexander–Conway relation (or called the Skein relation) is

$$\alpha(Ab^2B,n) = q^2(1-q^2)\alpha(AbB,n) + q^6\alpha(AB,n), \qquad A,B\in B_n \quad (31a)$$

and

$$\alpha(E,2) = q^{-2} + 1 + q^2. \quad (31b)$$

This link polynomial is a little different from the Jones polynomial. For comparison we give the analogy of the Jones polynomial:

$$\alpha_J(Ab^2B,n) = q(1-q^2)\alpha_J(AbB,n) + q^4\alpha_J(AB,n), \qquad A,B\in B_n, \quad (32a)$$

$$\alpha_J(E,2) = q^{-1} + q. \quad (32b)$$

For the IR [2,0], the Alexander–Conway relation is

$$\begin{aligned}\alpha(Ab^3B,n) &= q^4(1-q^4+q^6)\alpha(Ab^2B,n)\\ &\quad + q^{12}(1-q^2+q^6)\alpha(AbB,n)\\ &\quad - q^{22}\alpha(AB,n), \quad A,B\in B_n, \end{aligned} \quad (33a)$$

$$\alpha(E,2) = q^{-4} + q^{-2} + 2 + q^2 + q^4. \quad (33b)$$

This link polynomial is different from the Akutsu–Wadati polynomial.[1]

VII. DISCUSSION

We have computed q-CG coefficients and the $\check{R}_q$ matrix for IR [2,0] of q-sl(3). Now, we are going to discuss the general properties of those for IR $[\lambda,0]$ of q-sl(3) with a one-row Young tableau.

The decomposition of the coproduct $[\lambda,0]\times[\lambda,0]$ in q-sl(3) is

$$[2\lambda,0] + [2\lambda,-1,1] + [2\lambda-2,2] + \cdots + [\lambda,\lambda]. \quad (34)$$

The symmetry of the CG coefficients of SU(3) is

$$\xi_{[2\lambda-\mu,\mu]} = (-1)^\mu$$

and the difference of the Casimir operators of SU(3) is

$$\begin{aligned}\eta'([2\lambda-\mu,\mu]) &= C_2([2\lambda,0] - C_2([2\lambda-\mu,\mu])\\ &= 2\lambda\mu - \mu^2 + \mu. \end{aligned}\quad (35)$$

Both ξ and η' happen to coincide with those in the decomposition of the coproduct $[\lambda,0]\times[\lambda,0]$ in q-sl(2), namely, their $\check{R}_q$ matrices have the same eigenvalues and satisfy the same conditions:

$$\prod_N \{\check{R}_q - \xi_N q^{\eta'(N)}\} = 0. \quad (36)$$

Following the standard method to build the link polynomials,[1,10,11] we define

$$v = \operatorname{diag}(q^{-2\lambda}, q^{-2\lambda+2}, q^{-2\lambda+4},\ldots,q^0,q^{-2\lambda+4}, q^{-2\lambda+6},\ldots,q^2,\ldots,q^{2\lambda}), \quad (37)$$

then, we obtain

$$\tau = \left(\sum_{m=0}^{\lambda}\sum_{n=0}^{\lambda-m} q^{2n+4m}\right)^{-1}, \qquad \bar{\tau} = q^{4\lambda}\tau. \quad (38)$$

Therefore, the link polynomials (30) based on IR$[\lambda,0]$ or q-sl(3) are different from those on IR $[\lambda,0]$ of q-sl(2).

ACKNOWLEDGMENTS

The author would like to thank Professor Abdus Salam for hospitality at ICTP, Trieste, Italy. This work was partly supported by the Natural Science Foundation of China through the Nankai Institute of Mathematics.

APPENDIX: CASIMIR OPERATOR OF SU(3)

Define

$$T_2([\lambda_1,\lambda_2]) = \mathrm{Tr}(I_3^2),$$

where $[\lambda_1,\lambda_2]$ is an IR of SU(3). As a representation of SU(2) it can be decomposed as a direct sum of IR's $[\mu_1,\mu_2]$ of SU(2), for which we have

$$T_2^{(0)}([\mu_1,\mu_2]) = \tfrac{1}{12}(\mu_1-\mu_2)(\mu_1-\mu_2+1) \times(\mu_1-\mu_2+2).$$

Thus

$$\begin{aligned}T_2([\lambda_1,\lambda_2]) &= \sum_{\mu_2=0}^{\lambda_2}\sum_{\mu_1=\lambda_2}^{\lambda_1} T_2^{(0)}([\mu_1,\mu_2])\\ &= \tfrac{1}{48}(\lambda_1+2)(\lambda_2+1)(\lambda_1-\lambda_2+1)\\ &\quad\times(\lambda_1^2-\lambda_1\lambda_2+\lambda_2^2+3\lambda_1).\end{aligned}$$

Since the dimension of IR $[\lambda_1,\lambda_2]$ of SU(3) is $(\lambda_1+2)(\lambda_2+1)(\lambda_1-\lambda_2+1)/2$, we obtain (19).

[1] Y. Akutsu and M. Wadati, J. Phys. Soc. Jpn. **56**, 3039 (1987); Y. Akutsu, T. Deguchi, and M. Wadati, J. Phys. Soc. Jpn. **56**, 3464 (1987); M. Wadati and Y. Akutsu, Prog. Theor. Phys. Suppl. No. **94** (1988).

[2] Z. Q. Ma, Lecture given in the 13th John Hopkins Workshop on Current Problems in High Energy Particle Theory, Florence, Italy, 1989.

[3] D. I. Olive and N. Turok, Nucl. Phys. **B 220**, 491 (1983).

[4] B. Y. Hou, B. Y. Hou, and Z. Q. Ma, Beijing Institute of High Energy Physics preprints TH-89-7 and 8.

[5] A. M. Kirillov and N. Yu. Reshetikhin, LOMI preprints, E-9-88.

[6] J. J. de Swart, Rev. Mod. Phys. **35**, 916 (1963).

[7] O. Babelon, H. J. de Vega, and C. M. Viallet, Nucl. Phys. B **190**, 542 (1981). I. V. Cherednik, Theor. Math. Phys. **43**, 356 (1980). D. V. Chudnovsky and G. V. Chudnovsky, Phys. Lett. A **79**, 36 (1980). C. L. Schultz, Phys. Rev. Lett. **46**, 629 (1981).

[8] M. Jimbo, Comm. Math. Phys. **102**, 537 (1986). Lett. Math. Phys. **11**, 247 (1986).

[9] M. Rosso, Lecture given in the 13th John Hopkins Workshop on Current Problems in High Energy Particle Theory, Florence, Italy, 1989.

[10] V. F. R. Jones, Bull. Am. Math. Soc. **12**, 103 (1985).

[11] N. Yu. Reshetikhin, LOMI preprints, E-4-87, 4-17-87.

The potential group approach and hypergeometric differential equations

J. Wu[a)] and Y. Alhassid[b)]
Center for Theoretical Physics, Sloane Laboratory, Yale University, New Haven, Connecticut 06511

(Received for publication 31 August 1989; accepted for publication 1 November 1989)

This paper proposes a generalized realization of the potential groups SO(2,1) and SO(2,2) to describe the confluent hypergeometric and the hypergeometric equations, respectively. It implies that the classes of Schrödinger equations with solvable potentials whose analytical solutions are related to the confluent hypergeometric and the hypergeometric functions can be realized in terms of the above group structure.

I. INTRODUCTION

Group theoretic techniques are useful in many fields of physics.[1,2] However, most of the applications involving dynamical groups[3,4] were restricted to bound state problems. In the attempts to extend these techniques to describe scattering states, another kind of group, the potential group, was suggested.[5] This group connects states that have the same energy but belong to different potential strengths. Both bound and scattering states can then be realized within the same differential realization. This is made possible since the potential group, being noncompact, has both discrete and continuous representations. Such realizations were applied to various classes of solvable potentials.[6–8] In this paper we propose a way to generalize the realization of the potential groups SO(2,1) and SO(2,2) in order to describe the confluent hypergeometric equation and the hypergeometric equation, respectively. The realizations of these equations would imply a general proof that potential problems whose analytical solutions are given in terms of the confluent hypergeometric or hypergeometric functions could be described algebraically in terms of the group SO(2,1) or SO(2,2). In Ref. 9 we give examples of such solvable potentials, the Natanzon[10] and Ginocchio potentials,[11] and obtain their bound state spectra and scattering matrices by using purely algebraic techniques developed in Refs. 5, 12, and 13.

II. SO(2,1) AS A POTENTIAL GROUP

A. The algebra SO(2,1) and its realizations

The SO(2,1) algebra consists of three generators $J_\pm, J_3$ satisfying the commutation relations

$$
\begin{aligned}
&[J_+,J_-] = -2J_3,\\
&[J_3,J_\pm] = \pm J_\pm .
\end{aligned}
\tag{2.1}
$$

Consider the differential realization of the SO(2,1) algebra

$$
\begin{aligned}
J_\pm &= e^{\pm i\phi}\left[\pm\frac{\partial}{\partial x} + k_1(x)\left(i\frac{\partial}{\partial\phi}\mp\frac{1}{2}\right) + k_0(x)\right],\\
J_3 &= -i\frac{\partial}{\partial\phi},
\end{aligned}
\tag{2.2}
$$

where $k_1(x)$ and $k_0(x)$ are real functions to be determined and J_+ and J_- are Hermitian conjugate to each other, i.e., $J_- = (J_+)^\dagger$.

In order that $J_\pm, J_3$ form an SO(2,1) algebra, the commutation relations (2.1) have to be satisfied. This requirement provides the conditions which determine the functions $k_0(x)$ and $k_1(x)$, i.e.,

$$
\begin{aligned}
&\frac{dk_1(x)}{dx} + k_1^2(x) = 1,\\
&\frac{dk_0(x)}{dx} + k_0(x)k_1(x) = 0.
\end{aligned}
\tag{2.3}
$$

It can be shown that the most general solution for $k_1(x)$ is

$$
k_1(x) = \begin{cases} \tanh(x-c), & \text{for } k_1^2 < 1,\\ \pm 1, & \text{for } k_1^2 = 1,\\ \coth(x-c), & \text{for } k_1^2 > 1. \end{cases}
\tag{2.4}
$$

For any choice of $k_1(x)$, the function $k_0(x)$ can be found by solving the above first-order linear equation for $k_0(x)$.

The basis for an irreducible representation of SO(2,1) is characterized by

$$
\begin{aligned}
C_2|j,m\rangle &= j(j+1)|j,m\rangle,\\
J_3|j,m\rangle &= m|j,m\rangle,
\end{aligned}
\tag{2.5}
$$

where $C_2 = J_3^2 - \frac{1}{2}(J_+J_- + J_-J_+)$ is the SO(2,1) Casimir operator.

In the realization (2.2) the Casimir operator can be written as

$$
\begin{aligned}
C_2 = \frac{\partial^2}{\partial x^2} + (k_1^2(x)-1)\left(\frac{\partial^2}{\partial\phi^2}+\frac{1}{4}\right)\\
+2\frac{\partial k_0(x)}{\partial x}\, i\left(\frac{\partial}{\partial\phi}\right) - k_0^2(x) - \frac{1}{4},
\end{aligned}
\tag{2.6}
$$

and the basis (2.5) as

$$
|j,m\rangle = \Psi_{jm} = \psi_{jm}(x)\, e^{im\phi}.
\tag{2.7}
$$

The functions $\psi_{jm}(x)$ in (2.7) are the solutions of a one-dimensional Schrödinger equation with an m-dependent potential

$$
\left[-\frac{d^2}{dx^2} + V_m(x)\right]\psi_{jm}(x) = E\psi_{jm}(x),
\tag{2.8}
$$

where

[a)] Present address: Department of Physics, University of Tennessee, Knoxville, Tennessee 37996 and ORNL, Oak Ridge, Tennessee 37831.

[b)] Alfred P. Sloan Fellow.

 0022-2488/90/030557-06$03.00

$$V_m(x) = (k_1^2(x) - 1)\left(m^2 - \frac{1}{4}\right) + 2m\frac{dk_0(x)}{dx} + k_0^2(x), \tag{2.9}$$

and the energy is

$$E = -(j+\tfrac{1}{2})^2. \tag{2.10}$$

Due to (2.10) the Hamiltonian is related to the Casimir invariant of SO(2,1) by

$$H = -(C_2 + \tfrac{1}{4}). \tag{2.11}$$

Notice that in a given irreducible representation of SO(2,1) the energy is fixed since the Hamiltonian is a function of the SO(2,1) Casimir invariant C_2 only. States belonging to the same multiplet correspond to different potentials (2.9) characterized by the various values of m. For that reason we called the above SO(2,1) a potential algebra.[5]

B. Unitary representations of SO(2,1)

The realization (2.2) and (2.4) includes two classes of unitary representations[14] of SO(2,1).

(i) The discrete principal series D_j^+ for which

$$\begin{aligned} &j = -\tfrac{1}{2} - n/2 \quad (n = 0,1,2,\ldots), \\ &m = -j, -j+1, -j+2,\ldots . \end{aligned} \tag{2.12}$$

The corresponding states of the Schrödinger equation (2.8) describe bound states with energy $E = -(j+1/2)^2$. If the potential in (2.8) does not have bound states the discrete series will not appear in the corresponding realization.

(ii) The continuous principal series where

$$j = -\tfrac{1}{2} + ik \quad (0 < k < \infty),$$

$$m = 0, \pm 1, \pm 2, \ldots,$$

or

$$m = \pm\tfrac{1}{2}, \pm\tfrac{3}{2}, \pm\tfrac{5}{2},\ldots . \tag{2.13}$$

It describes the scattering states of (2.8) with energy $E = k^2 > 0$.

If in (2.12) we fix the potential strength m, then by varying j we obtain from the discrete series a finite number of bound states with energy

$$\langle H \rangle = -(j+\tfrac{1}{2})^2 = -(m-n-\tfrac{1}{2})^2,$$

$$n = 0,1,2,\ldots \{m - \tfrac{1}{2}\}, \tag{2.14}$$

where { } means the integer part. From the continuous series we obtain the continuous spectrum with

$$\langle H \rangle = +k^2 \quad (0 < k^2 < \infty). \tag{2.15}$$

The third class of the unitary representations of SO(2,1), the supplementary series,[14] is missing in (2.2).

C. Classes of SO(2,1) realizations

Depending on the choice of solutions to (2.3), Eq. (2.2) leads to three basic classes of SO(2,1) realizations:

(i) $k_1(x) = 1, \quad k_0(x) = e^{-x}$,
for which

$$J_\pm = e^{\pm i\phi}\left[\pm\frac{\partial}{\partial x} + \left(i\frac{\partial}{\partial\phi} \mp \frac{1}{2}\right) + e^{-x}\right], \tag{2.16}$$

(ii) $k_1(x) = \tanh x, \quad k_0(x) = 0$,
for which

$$J_\pm = e^{\pm i\phi}\left[\pm\frac{\partial}{\partial x} + \tanh x\left(i\frac{\partial}{\partial\phi} \mp \frac{1}{2}\right)\right], \tag{2.17}$$

(iii) $k_1(x) = \coth x, \quad k_0(x) = 0$,
for which

$$J_\pm = e^{\pm i\phi}\left[\pm\frac{\partial}{\partial x} + \coth x\left(i\frac{\partial}{\partial\phi} \mp \frac{1}{2}\right)\right]. \tag{2.18}$$

Any one of the above classes can be generalized by introducing a similarity transformation $J_\pm \to SJ_\pm S^{-1}$ with $S = e^{s(x)}$ or a coordinate transformation $x = x(z)$. The commutation relations will be preserved no matter how complicated these transformations are.

Each class describes a certain set of solvable SO(2,1) potentials. In Sec. III we shall show that the first class leads to the general confluent hypergeometric equation and thus to any potential related to it. The second and third classes lead to special cases of the hypergeometric equation and some solvable potentials related to it (Sec. IV). To obtain the general hypergeometric equation whose solution depends on three parameters we need to enlarge the group, since SO(2,1) provides only two parameters (j,m). This is accomplished in Sec. V and Sec. VI by using the algebra SO(2,2).

III. THE FIRST CLASS OF SO(2,1) REALIZATIONS

A. The Morse potential

Using the realization (2.16) we obtain in Eq. (2.8)

$$\left[-\frac{d^2}{dx^2} - 2me^{-x} + e^{-2x}\right]\psi_{jm}(x) = E\psi_{jm}(x). \tag{3.1}$$

A shift $x = x' - \log m$ would transform it to the one-dimensional Schrödinger equation with the Morse potential,[15] which after dropping the prime in x' is

$$\left[-\frac{d^2}{dx^2} + m^2(e^{-2x} - 2e^{-x})\right]\psi_{jm}(x) = E\psi_{jm}(x). \tag{3.2}$$

Both discrete series (2.12) and the continuous series (2.13) appear in this realization as explained in Sec. II. Figure 1 shows several SO(2,1) multiplets for the Morse potential.

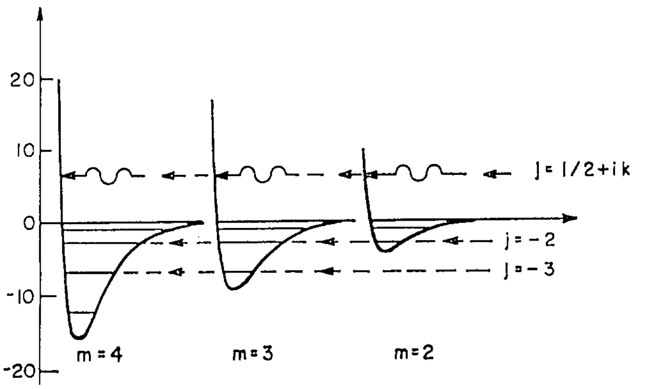

FIG. 1. Morse potential $V_m(x) = m^2(e^{-2x} - 2e^{-x})$ for $m = 2,3,4$ are plotted in solid lines. The SO(2,1) multiplets for $j = -2, -3, -1/2 + ik$ are shown by the horizontal dashed lines.

B. The confluent hypergeometric equation and SO(2,1)

We shall now show how to generalize the Morse realization of Sec. III A in order to realize the confluent hypergeometric equation with the potential group SO(2,1). This implies that all solvable potential problems whose solutions are related to the confluent hypergeometric function can be cast into differential realizations of the SO(2,1) algebra.

Consider the Morse SO(2,1) realization (2.16). After a coordinate transformation $y = 2e^{-x}$ we have

$$J_\pm = e^{\pm i\phi}\left[\mp y\frac{\partial}{\partial y} + \left(i\frac{\partial}{\partial\phi} \mp \frac{1}{2}\right) + \frac{y}{2}\right], \qquad J_3 = -i\frac{\partial}{\partial\phi}. \tag{3.3}$$

The Casimir operator in this realization is

$$C_2 = y^2\frac{\partial^2}{\partial y^2} + y\frac{\partial}{\partial y} - y\left(i\frac{\partial}{\partial\phi}\right) - \frac{y^2}{4} - \frac{1}{4}. \tag{3.4}$$

Next, we define a null operator

$$Q = (1/y)(C_2 - j(j+1)) = y\frac{\partial^2}{\partial y^2} + \frac{\partial}{\partial y} - i\frac{\partial}{\partial\phi} - \frac{y}{4} - \left(j+\frac{1}{2}\right)^2 y^{-1}, \tag{3.5}$$

where $j(j+1)$ is the eigenvalue of the Casimir operator. The action of Q on the simultaneous eigenfunction of C_2 and J_3

$$\Psi_{jm} = \psi_{jm}(y)\, e^{im\phi}, \tag{3.6}$$

leads to a differential equation for $\psi_{jm}(y)$, i.e.,

$$Q_c\psi(y) \equiv \left[y\frac{d^2}{dy^2} + \frac{d}{dy} + m - \frac{y}{4} - \left(j+\frac{1}{2}\right)^2 y^{-1}\right]\psi_{jm}(y) = 0. \tag{3.7}$$

Using a similarity transformation

$$S = e^{ay}y^b, \tag{3.8}$$

we can transform Q_c into the following operator:

$$H_{CH} = SQ_cS^{-1} = y\frac{d^2}{dy^2} + (1-2b-2ay)\frac{d}{dy} + \left(b^2 - \left(j+\frac{1}{2}\right)^2\right)\frac{1}{y} + (a^2 - \tfrac{1}{4})\,y + (2ab - a + m). \tag{3.9}$$

If in Eq. (3.9) we choose

$$a = \tfrac{1}{2}, \qquad b = j + \tfrac{1}{2}, \tag{3.10}$$

we obtain the confluent hypergeometric operator

$$H_{CH} = y\frac{d^2}{dy^2} + (\gamma - y)\frac{d}{dy} - \alpha, \tag{3.11}$$

where α and γ are related to the algebraic quantum numbers j and m by

$$\gamma = -2j, \qquad \alpha = -j - m. \tag{3.12}$$

It follows that the confluent hypergeometric functions ${}_1F_1(-j-m, -2j;y)$ are related to a basis $|j,m\rangle$ of SO(2,1). By comparing (3.9) to (3.11) we obtain another solution if j is replaced by $-j-1$ in (3.10) and (3.12). This is in agreement with the fact[16] that $y^{1-\gamma}{}_1F_1(1+\alpha-\gamma, 2-\gamma; y)$ is also a solution of (3.11). In the language of group theory, it is known that the transformation $j \to -j-1$ carries us to an equivalent representation.[14] Notice that for the discrete series D_j^+ we have $m = -j + n$ where n is an non-negative integer. It follows that

$$\alpha = -n, \tag{3.13}$$

which is exactly the condition that the confluent hypergeometric series will terminate at some order and becomes a polynomial.[16]

IV. THE SECOND AND THIRD CLASSES OF SO(2,1) REALIZATIONS

A. Pöschl–Teller potentials

If we choose the realization (2.17), then Eq. (2.8) becomes the one-dimensional Schrödinger equation with the Pöschl–Teller potential[17]

$$\left[-\frac{d^2}{dx^2} - \frac{m^2 - \frac{1}{4}}{\cosh^2 x}\right]\psi_{jm}(x) = E\psi_{jm}(x). \tag{4.1}$$

Both the discrete and continuous series appear in this realization as in Sec. II B.

If we choose the realization (2.18), we obtain

$$\left[-\frac{d^2}{dx^2} + \frac{m^2 - \frac{1}{4}}{\sinh^2 x}\right]\psi_{jm}(x) = E\psi_{jm}(x), \tag{4.2}$$

which is a repulsive modified Pöschl–Teller potential and only continuous series representation are allowed.

Generalizations of the above realizations lead to other solvable potentials whose solutions are special cases of the hypergeometric functions. Such generalizations are discussed in the next section. It is not possible to obtain within the SO(2,1) realizations the general hypergeometric function since it contains three independent parameters and SO(2,1) provides only two (j and m).

B. Ginocchio potentials

The discrete spectrum appearing in the realization (2.2) is quadratic. It is, however, possible to find more complicated examples which are in the same above classes of SO(2,1) realizations but where the spectrum has a more complicated structure. The technique[18] to do that is by transforming the realization to one in which the resulting potential is not only a function of m but also of j. For a given potential with fixed parameters the energy is then given by a more complicated form than (2.14). Now we illustrate the technique to potentials discussed by Ginocchio.[11]

Starting from the second class realization (2.17), we apply a similarity transformation S to it and obtain

$$K_\pm = SJ_\pm S^{-1}, \qquad K_3 = SJ_3S^{-1} \equiv J_3, \tag{4.3}$$

where

$$S^{-1} = \cosh^{1/2} x/(\lambda^2 + \sinh^2 x)^{1/4}. \tag{4.4}$$

Then we perform the following coordinate transformations:

$$z = \tanh x,$$
$$y = z/(\lambda^2 + (1-\lambda^2)z^2)^{1/2}.$$
$$r = (1/\lambda^2)[\operatorname{arctanh} y + (\lambda^2 - 1)^{1/2} \times \arctan((\lambda^2-1)^{1/2}y)]. \tag{4.5}$$

In the algebraic equation for the Casimir operator

$$-(C_2 + \tfrac{1}{4})\Psi_{jm} = -(j+\tfrac{1}{2})^2\Psi_{jm}, \tag{4.6}$$

we multiply it by a factor $f(z)$, where

$$f(z) = \lambda^4/[\lambda^2 + (1-\lambda^2)z^2]. \tag{4.7}$$

In order to restore its eigenequation form we subtract $\lambda^4(j+1/2)^2$ from $f(z)$ to obtain

$$H_G\Psi_{jm} = E'\Psi_{jm}, \tag{4.8}$$

where $E' = -\lambda^4(j+\frac{1}{2})^2$. In Eq. (4.8), we have

$$H_G = [f(z) - \lambda^4](j+\tfrac{1}{2})^2 - f(z)(C_2 + \tfrac{1}{4}), \tag{4.9}$$

where

$$C_2 = K_3^2 - \tfrac{1}{2}(K_+K_- + K_-K_+).$$

Equation (4.8) has a solution of the form

$$\Psi_{jm} = \psi_{jm}(r)\, e^{im\phi}, \tag{4.10}$$

where $\psi_{jm}(r)$ satisfies a Schrödinger equation

$$\left[-\frac{d^2}{dr^2} + V(r)\right]\psi_{jm}(r) = E'\psi_{jm}(r), \tag{4.11}$$

with a potential

$$V(r) = -\lambda^2\nu(\nu+1)(1-y^2) + [(1-\lambda^2)/4][5(1-\lambda^2)y^4 - (7-\lambda^2)y^2 + 2](1-y^2). \tag{4.12}$$

In (4.12), the parameter ν depends on both m and j through

$$(\nu+\tfrac{1}{2})^2 = m^2 + (j+\tfrac{1}{2})^2(\lambda^2-1). \tag{4.13}$$

The potential (4.12) was recently discussed in Ref. 11. For a given such potential with fixed λ and ν we can find the discrete spectrum by using (4.13) and the condition $m = -j+n$, $n = 0,1,2,\ldots$ to solve for $j+1/2$ and the allowed values of n, i.e.,

$$E' = -\lambda^4(j+\tfrac{1}{2})^2 = -\tfrac{1}{4}\{[(1-\lambda^2)(2n+1)^2 + \lambda^2(2\nu+1)^2]^{1/2} - (2n+1)\}^2, \tag{4.14}$$

where $n = 0,1,2,\ldots,\{\nu\}$.

V. SO(2,2) AS A POTENTIAL GROUP

A detailed discussion of SO(2,2) and solvable potentials associated with it can be found elsewhere.[9] Here we just summarize the main results which are needed for the discussion of the hypergeometric equation in Sec. VI.

A. The algebra SO(2,2) and its symmetric representation

Consider the differential realization of the SO(2,2) algebra on the (2,2) hyperboloid H^3:

$$x_1^2 + x_2^2 - x_3^2 - x_4^2 = \rho^2 > 0. \tag{5.1}$$

This hyperboloid can be parametrized with three parameters (χ,ϕ,θ) as follows:

$$\begin{aligned} x_1 &= \rho\cosh\chi\cos\phi, \quad x_2 = \rho\cosh\chi\sin\phi, \\ x_3 &= \rho\sinh\chi\cos\theta, \quad x_4 = \rho\sinh\chi\sin\theta, \end{aligned} \tag{5.2}$$

where ϕ and θ are rotation angles in 1-2 and 3-4 planes, respectively. It can be shown that the six operators J_i, K_i $(i = 1,2,3)$ defined below form an SO(2,2) algebra:

$$\begin{aligned} J_1 &= N_{23}, \quad J_2 = -N_{13}, \quad J_3 = M_{12} = -i\frac{\partial}{\partial\phi}, \\ K_1 &= N_{14}, \quad K_2 = N_{24}, \quad K_3 = M_{34} = -i\frac{\partial}{\partial\theta}, \end{aligned} \tag{5.3}$$

where

$$\begin{aligned} M_{ab} &= x_aP_b - x_bP_a, \\ N_{ab} &= x_aP_b + x_bP_a - i\delta_{a,b}I, \end{aligned} \tag{5.4}$$

for $a,b = 1,2,3,4$. In (5.4)

$$P_a = -i\frac{\partial}{\partial x_a}, \tag{5.5}$$

and I is the unit operator. The Casimir operator of the SO(2,2) algebra

$$C_2 = J_3^2 + K_3^2 - J_1^2 - J_2^2 - K_1^2 - K_2^2,$$

is found to be

$$C_2 = \frac{\partial^2}{\partial\chi^2} + (\tanh\chi + \coth\chi)\frac{\partial}{\partial\chi} + \frac{1}{\cosh^2\chi}\left(-\frac{\partial^2}{\partial\phi^2}\right) - \frac{1}{\sinh^2\chi}\left(-\frac{\partial^2}{\partial\theta^2}\right). \tag{5.6}$$

This is a symmetric representation of the SO(2,2) group. A basis $|\omega,m_1,m_2\rangle$ for this representation is characterized by

$$\begin{aligned} C_2|\omega,m_1,m_2\rangle &= \omega(\omega+2)\,|\omega,m_1,m_2\rangle, \\ J_3|\omega,m_1,m_2\rangle &= m_1|\omega,m_1,m_2\rangle, \\ K_3|\omega,m_1,m_2\rangle &= m_2|\omega,m_1,m_2\rangle. \end{aligned} \tag{5.7}$$

This basis can be written explicitly as

$$\Psi_{\omega,m_1,m_2} = e^{i(m_1\phi + m_2\theta)}\psi_{\omega m_1 m_2}(\chi). \tag{5.8}$$

B. The modified Pöschl–Teller potential

Here we summarize results from Ref. 9.

Performing a similarity transformation with $S = [\sinh(4\chi)]^{1/2}$ on the realization given in Sec. V A we find that the algebraic Eqs. (5.7) have a solution of the form (5.8) where $\psi_{\omega m_1 m_2}(\chi)$ satisfies

$$\left[-\frac{1}{4}\frac{d^2}{d\chi^2} - \frac{m_1^2 - 1/4}{\cosh^2 2\chi} + \frac{m_2^2 - 1/4}{\sinh^2 2\chi}\right]\psi_{\omega m_1 m_2}(\chi) = -(\omega+1)^2\psi_{\omega m_1 m_2}(\chi). \tag{5.9}$$

This is a Schrödinger equation with the modified Pöschl–Teller potential which depends on two parameters m_1 and m_2.

Two types of symmetric representations appear in (5.9).

(i) The discrete series with

$$\begin{aligned} \omega &= -1, -2, -3, \ldots, \\ m_2 &= 0, \pm 1, \pm 2, \ldots, \\ m_1 &= -\omega + |m_2|,\ -\omega + |m_2| + 2,\ -\omega + |m_2| + 4, \ldots. \end{aligned} \tag{5.10}$$

The corresponding states are the bound states of (5.9).

(ii) The continuous series where

$$\omega = -1 + ik \quad (0<k<\infty), \qquad m_1, m_2 = 0, \pm 1, \pm 2, \ldots . \tag{5.11}$$

These are the scattering states.

VI. THE HYPERGEOMETRIC EQUATION AND SO(2,2) GROUP

In order to obtain from the SO(2,2) realization in Sec. V A the hypergeometric equation, we use a coordinate transformation

$$z = \tanh^2 \chi, \tag{6.1}$$

and then write the Casimir operator (5.6) as

$$C_2 = 4\Big[z(1-z)^2 \frac{\partial^2}{\partial z^2} + (1-z)^2 \frac{\partial}{\partial z} + \frac{1-z}{4}\Big(-\frac{\partial^2}{\partial \phi^2}\Big) - \frac{1-z}{4z}\Big(-\frac{\partial^2}{\partial \theta^2}\Big)\Big]. \tag{6.2}$$

Next, we define a null operator

$$Q = \frac{1}{4(1-z)}[C_2 - \omega(\omega+2)] = z(1-z)\frac{\partial^2}{\partial z^2} + (1-z)\frac{\partial}{\partial z} + \frac{1}{4}\Big(-\frac{\partial^2}{\partial \phi^2}\Big) - \frac{1}{4z}\Big(-\frac{\partial^2}{\partial \theta^2}\Big) - \frac{\omega(\omega+2)}{4(1-z)}. \tag{6.3}$$

The null action of Q on the basis states $\Psi_{\omega m_1 m_2}$ leads to the equation for $\psi_{\omega m_2 m_2}(z)$ as

$$Q_c \psi_{\omega m_1 m_2}(z) \equiv \Big[z(1-z)\frac{d^2}{dz^2} + (1-z)\frac{d}{dz} + \frac{1}{4}m_1^2 - \frac{1}{4z}m_2^2 - \frac{\omega(\omega+2)}{4(1-z)}\Big]\psi_{\omega m_1 m_2}(z) = 0. \tag{6.4}$$

A similarity transformation

$$S^{-1} = z^a(1-z)^b \tag{6.5}$$

will transform Q_c to

$$H_{\rm HG} = SQ_cS^{-1} = z(1-z)\frac{d^2}{dz^2} + (2a+1-(2a+2b+1)z)\frac{d}{dz} + \Big(a^2 - \frac{m_2^2}{4}\Big)\frac{1}{z} + \Big(b^2 - b - \frac{\omega(\omega+2)}{4}\Big)\frac{1}{1-z} - \Big(b^2 - \frac{m_1^2}{4} + 2ab + a^2\Big). \tag{6.6}$$

By choosing in (6.5)

$$a = \tfrac{1}{2}m_2, \qquad b = \tfrac{1}{2}(\omega+2), \tag{6.7}$$

we obtain from (6.6) the hypergeometric operator

$$H_{\rm HG} = z(1-z)\frac{d^2}{dz^2} + (\gamma - (\alpha+\beta+1)z)\frac{d}{dz} - \alpha\beta. \tag{6.8}$$

The parameters α,β,γ of the hypergeometric function ${}_2F_1(\alpha,\beta,\gamma;z)$ which satisfies $H_{\rm HG}\;{}_2F_1 = 0$, are given in terms the group quantum numbers ω, m_1, m_2 by

$$(\omega+1)^2 = (\alpha+\beta-\gamma)^2, \qquad m_1^2 = (\alpha-\beta)^2, \qquad m_2^2 = (\gamma-1)^2. \tag{6.9}$$

A possible solution to (6.9) is

$$\gamma = m_2 + 1, \qquad \beta = \tfrac{1}{2}(\omega + m_1 + m_2 + 2), \qquad \alpha = \tfrac{1}{2}(\omega - m_1 + m_2 + 2). \tag{6.10}$$

Another solution to (6.9) is obtained by the substitutions $\omega \to -\omega - 2$ and $m_{1,2} \to -m_{1,2}$ in Eqs. (6.7) and (6.10). This agrees with the fact[16] that

$$z^{1-\gamma}(1-z)^{\gamma-\alpha-\beta}\;{}_2F_1(1-\alpha, 1-\beta, 2-\gamma; z)$$

is also a solution of (6.8).

An alternative realization of the hypergeometric equation in terms of $y = 1 - z$ can be achieved through the replacement

$$\gamma \to \alpha + \beta - \gamma + 1, \tag{6.11}$$

in the above parameters and the prescription of the algebraic quantum numbers.

A solvable class of potentials associated with the hypergeometric equation is the modified Pöshl–Teller potential described in Sec. V B. A more general class is that of the Natanzon potentials[10] whose SO(2,2) description is provided in Ref. 9.

The hypergeometric equation can also be realized with the SO(m,n) group on the (m,n) hyperboloid for $m,n \geqslant 2$ with the replacements:

$$(\omega+1)^2 \to \langle C_2^{(m,n)}\rangle + ((m+n-2)/2)^2, \qquad m_1^2 \to \langle C_2^{(m)}\rangle + ((m-2)/2)^2, \qquad m_2^2 \to \langle C_2^{(n)}\rangle + ((n-2)/2)^2, \tag{6.12}$$

where $C_2^{(m,n)}, C_2^{(m)}$, and $C_2^{(n)}$ are the Casimir operators of SO(m,n), SO(m), and SO(n), respectively. The SO(2,2) realization is the simplest case of the nontrivial SO(m,n) realizations.

VII. CONCLUSIONS

In this paper we have proposed generalized realizations of potential groups to describe the confluent hypergeometric equation and the hypergeometric equation with the SO(2,1) and SO(2,2) potential groups, respectively. Our work implies that all solvable potential problems whose solutions are related to the hypergeometric and confluent hypergeometic functions can be cast into such group structures. In Ref. 9 we discuss in detail classes of solvable potentials associated with SO(2,2) and the general class of the Natanzon potentials, Pöschl–Teller potentials, Rosen–Morse potentials, and others. We also use algebraic techniques to calculate the bound state spectra and scattering matrices of these potentials. For further solvable potentials we may have to explore other groups and special functions.

As a concluding remark, we would like to point out the relationship between the SO(2,2) realizations and SO(2,1) realizations. In Sec. II, we have shown that there can be three solutions to $k_1(x)$. Two of them are related to the hypergeometric equation, i.e., the cases where $k_1^2(x) \neq 1$. These realizations can be reached from the SO(2,2) realization by setting one of the m quantum numbers equal to a constant, since the SO(2,2) group contains several SO(2,1) subgroups with different choices of generators. The other case where $k_1^2(x) = 1$, which provides the confluent hypergeometric equation, can also be obtained from the SO(2,2) realization by properly taking the limit[7]

$$(m_1 - m_2)/2 \to \infty .$$

It is the same limit which carries the hypergeometric equation into the confluent hypergeometric equation.

ACKNOWLEDGMENTS

The authors would like to thank F. Iachello and F. Gürsey for many helpful discussions.

This work was supported in part by the Department of Energy contract numbers DE-AC-02-76 ER 03074 and DE-AS-05-83 ER 40101, DE-FG-05-87 ER 40361.

[1]W. Pauli, Z. Phys. **36**, 336 (1926).
[2]M. Gell-Mann and Y. Ne'eman, *The Eightfold Way* (Benjamin, New York, 1964).
[3]A. Arima and F. Iachello, Ann. Phys. **99**, 253 (1974); Ann. Phys. **111**, 201 (1978); Ann. Phys. **123**, 468 (1979).
[4]F. Iachello, Chem. Phys. Lett. **78**, 581 (1981); F. Iachello and R. D. Levine, J. Chem. Phys. **77**, 3046 (1982).
[5]Y. Alhassid, F. Gürsey, and F. Iachello, Phys. Rev. Lett. **50**, 873 (1983); Ann. Phys. **148**, 346 (1983); Ann. Phys. **167**, 181 (1986).
[6]A. Frank and K. B. Wolf, J. Math. Phys. **26**, 1973 (1985).
[7]J. Wu, Ph.D. thesis, Yale University, 1985 (unpublished).
[8]A. O. Barut, A. Inomata, and R. Wilson, J. Phys. A **20**, 4083 (1987).
[9]J. Wu, Y. Alhassid, and F. Gürsey, Ann. Phys. **196**, 163 (1989).
[10]G. Natanzon, Vestnik Leningrad Univ. **10**, 22 (1971); Teor. Mat. Fiz. **38**, 146 (1979) (English transl.).
[11]J. Ginnocchio, Ann. Phys. **152**, 203 (1984); Ann. Phys. **159**, 467 (1985).
[12]A. Frank and K. B. Wolf, Phys. Rev. Lett. **52**, 1737 (1984); A. Frank, Y. Alhassid, and F. Iachello, Phys. Rev. A **34**, 677 (1986).
[13]Y. Alhassid, J. Engel, and J. Wu, Phys. Rev. Lett. **53**, 17 (1984); J. Wu, F. Iachello, and Y. Alhassid, Ann. Phys. **173**, 68 (1987).
[14]See, for example, in G. Lindbland and B. Nagel, Ann. Inst. Henri Poincaré **13**, 27 (1970).
[15]P. M. Morse, Phys. Rev. **34**, 57 (1929).
[16]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).
[17]G. Pöschl and E. Teller, Z. Phys. **83**, 143 (1933).
[18]Y. Alhassid, F. Iachello, and R. D. Levine, Phys. Rev. Lett. **54**, 1746 (1985).

Parametrization of SU(n) with n—1 orthonormal vectors

Adam M. Bincer
Department of Physics, University of Wisconsin at Madison, Madison, Wisconsin 53706

(Received 1 February 1989; accepted for publication 18 October 1989)

A generalization to SU(n) of a well-known relation in SU(2) is proposed. It relies on the observation that an element of SU(n) has associated with it in a natural way $n-1$ orthonormal vectors in $\mathbf{R}^{n^2-1}$. The meaning of these $n-1$ vectors is discussed as they relate to the geometry of the adjoint representation of SU(n).

I. INTRODUCTION

Various parametrizations of SU(n) [and in particular of SU(2) and SU(3)] have been discussed in the literature. In this paper, we follow closely the procedure of Macfarlane, Sudbury, and Weisz[1] in arriving at the parametrization described in Sec. II. In discussing the geometry of the adjoint representation in Sec. IV, we followed the lead of Michel and Radicati.[2] A number of additional references are given in the above two papers, between which we should like to call attention to the paper by Rosen.[3]

If g is an element of SU(n) then, as is well known, it can be parametrized as

$$g = e^{iB}, \tag{1.1}$$

where B is a Hermitian traceless $n\times n$ matrix. If λ_i, $1\leqslant i\leqslant n^2-1$, are a basis for such matrices then we may write

$$B = \lambda_i b_i = \boldsymbol{\lambda}\cdot\mathbf{b}, \tag{1.2}$$

with $\mathbf{b}$ a vector in $\mathbf{R}^{n^2-1}$. Thus we obtain a parametrization of g in terms of the n^2-1 real numbers, the components of $\mathbf{b}$.

At the SU(2) level, one introduces θ as the magnitude of $\mathbf{b}$ so that

$$\mathbf{b} = \theta\mathbf{m}, \quad \theta = \sqrt{\mathbf{b}\cdot\mathbf{b}}, \tag{1.3}$$

with $\mathbf{m}$ a unit vector. This yields the famous relation (with $\boldsymbol{\lambda}$ given by the Pauli σ)

$$e^{i\boldsymbol{\lambda}\cdot\mathbf{b}} = 1\cos\theta + i\boldsymbol{\lambda}\cdot\mathbf{m}\sin\theta. \tag{1.4}$$

In this version, g is parametrized by an invariant and a unit vector.

The question is how should Eqs. (1.3) and (1.4) be generalized to SU(n)? It is clear that at the SU(n) level one must have

$$e^{i\boldsymbol{\lambda}\cdot\mathbf{b}} = 1M_0 + i\boldsymbol{\lambda}\cdot\mathbf{M} \tag{1.5}$$

simply because the $\boldsymbol{\lambda}$ and the unit matrix span the space of $n\times n$ matrices. Moreover, M_0 must be a function of invariants formed out of $\mathbf{b}$, and $\mathbf{M}$ must be a vector formed out of $\mathbf{b}$. Since one can form out of $\mathbf{b}$ $n-2$ additional linearly independent vectors we may write

$$\mathbf{M} = \mathbf{b}M_1 + \sum_{\alpha=2}^{n-1}\mathbf{b}_\alpha M_\alpha, \tag{1.6}$$

where $\mathbf{b}_\alpha$, $2\leqslant\alpha\leqslant n-1$, are the additional vectors, and M_α, $1\leqslant\alpha\leqslant n-1$, are functions of invariants formed out of $\mathbf{b}$.

This is the generalizaion to SU(n) proposed by Macfarlane, Sudbury, and Weisz.[1] Although perfectly satisfactory in its way, it does not really parallel the parametrization in Eq. (1.3). To this end we propose setting

$$\mathbf{b} = \theta_\alpha\mathbf{m}_\alpha, \tag{1.7}$$

where the θ_α are invariants formed out of $\mathbf{b}$, $\mathbf{m}_\alpha$ are unit vectors, and α ranges over $n-1$ values. We shall show that this parametrization arises naturally and that the $\mathbf{m}_\alpha$ have interesting properties. In particular they form an orthonormal set:

$$\mathbf{m}_\alpha\cdot\mathbf{m}_\beta = \delta_{\alpha\beta}. \tag{1.8}$$

Using these concepts we shall obtain two different generalizations to SU(n) of Eq. (1.4). In Sec. II, we summarize some well-known facts and obtain our first generalization of Eq. (1.4). In Sec. III, a generalization in the form of a product is proposed. In Sec. IV, we comment on the geometry of $\mathbf{R}^{n^2-1}$ and the meaning of the $n-1$ directions associated with the various $\mathbf{m}_\alpha$.

II. THE PARAMETRIZATION $\mathbf{b}=\theta_\alpha\mathbf{m}_\alpha$

As basis for the SU(n) algebra we take the $n\times n$ matrices λ_i, $1\leqslant i\leqslant n^2-1$, defined to satisfy

$$\lambda_i\lambda_j = (2/n)\delta_{ij}1 + (d_{ijk} + if_{ijk})\lambda_k, \tag{2.1}$$

implying

$$\operatorname{Tr}\lambda_i\lambda_j = 2\delta_{ij}, \tag{2.2}$$

$$[\lambda_i,\lambda_j] = 2if_{ijk}\lambda_k, \tag{2.3}$$

$$\{\lambda_i,\lambda_j\} = (4/n)\delta_{ij}1 + 2d_{ijk}\lambda_k. \tag{2.4}$$

For $n=2$ these are the Pauli σ, for $n=3$ these are the Gell-Mann λ.

Let the matrix B and the vector $\mathbf{b}$ be defined by Eqs. (1.1) and (1.2). If B' is obtained from B by a similarity transformation

$$B' = UBU^{-1}, \tag{2.5}$$

then

$$\operatorname{Tr}(B')^k = \operatorname{Tr}(UBU^{-1})^k = \operatorname{Tr}B^k. \tag{2.6}$$

Taking $U\in$ SU(n) it follows that traces of powers of B are SU(n) invariants. In view of the Cayley–Hamilton theorem and since

$$\operatorname{Tr}B^0 = n, \quad \operatorname{Tr}B^1 = 0, \tag{2.7}$$

there are altogether $n-1$ such independent invariants, namely,

$$I_k = \operatorname{Tr}B^k, \quad 2\leqslant k\leqslant n. \tag{2.8}$$

 0022-2488/90/030563-05$03.00

Specifically for $k=2,3,4$ we have

$$I_2 = b_i b_j \operatorname{Tr} \lambda_i \lambda_j = 2\mathbf{b}\cdot\mathbf{b}, \tag{2.9}$$

$$I_3 = b_i b_j b_k \operatorname{Tr} \lambda_i \lambda_j \lambda_k = 2 d_{ijk} b_i b_j b_k, \tag{2.10}$$

$$\begin{aligned} I_4 &= b_i b_j b_k b_l \operatorname{Tr} \lambda_i \lambda_j \lambda_k \lambda_l \\ &= ((4/n)\delta_{ij}\delta_{kl} + 2 d_{ijs} d_{kls}) b_i b_j b_k b_l, \end{aligned} \tag{2.11}$$

which are explicit examples of how the $n-1$ invariants are formed out of $\mathbf{b}$. Note that the invariance of I_2 is the statement that the length of $\mathbf{b}$ is an invariant.

As is well known, we may associate with B a different set of invariants, namely its n eigenvalues φ_k, $1 \leqslant k \leqslant n$, which are the roots of the characteristic equation of B. Because B is traceless we have

$$\sum_{k=1}^{n} \varphi_k = 0 \tag{2.12}$$

and we may form out of the φ_k $n-1$ independent combinations.

We would like to call attention to one more choice of $n-1$ independent invariants. The matrix B may be diagonalized and if S denotes the appropriate diagonalizing transformation matrix then

$$SBS^{-1} = \lambda_\alpha \theta_\alpha, \tag{2.13}$$

where we expand the diagonal matrix SBS^{-1} in terms of the basis λ_i using *only* the subset involving the *diagonal* matrices λ_α. Quite explicitly these λ_α are $n-1$ in number and, setting $\alpha = 3,8,\ldots, n^2-1$, we may write

$$\lambda_{a^2-1} = \left\{ (1-a) D_a + \sum_{k=1}^{a-1} D_k \right\} \sqrt{\frac{2}{a(a-1)}}, \quad 2 \leqslant a \leqslant n, \tag{2.14}$$

where D_k denotes the matrix with zeros everywhere except for unity at the intersection of the k th row and k th column.

It follows that φ_k and θ_α are related by

$$\varphi_k = -\theta_{k^2-1} \sqrt{\frac{2(k-1)}{k}} + \sum_{l=k+1}^{n} \theta_{l^2-1} \sqrt{\frac{2}{l(l-1)}}, \tag{2.15}$$

$$-\theta_{k^2-1} \sqrt{\frac{2(k-1)}{k}} = \varphi_k + \frac{1}{k} \sum_{l=k+1}^{n} \varphi_l, \tag{2.16}$$

for $1 \leqslant k \leqslant n$.

We now observe that the invariants I_k are also easily expressed in terms of the θ_α:

$$I_k = \operatorname{Tr} B^k = \operatorname{Tr}(SBS^{-1})^k = \operatorname{Tr}(\lambda_\alpha \theta_\alpha)^k. \tag{2.17}$$

Specifically for $k = 2,3,4$ have

$$I_2 = \operatorname{Tr} \lambda_\alpha \theta_\alpha \lambda_\beta \theta_\beta = 2\theta_\alpha \theta_\alpha, \tag{2.18}$$

$$\begin{aligned} I_3 &= \operatorname{Tr} \lambda_\alpha \theta_\alpha \lambda_\beta \theta_\beta \lambda_\gamma \theta_\gamma \\ &= 2 \sum_{a=3}^{n} \theta_{a^2-1} \sqrt{\frac{2}{a(a-1)}} \left\{ (2-a)\theta_{a^2-1}^2 \right. \\ &\quad \left. + 3 \sum_{b=2}^{a-1} \theta_{b^2-1}^2 \right\}, \end{aligned} \tag{2.19}$$

$$\begin{aligned} I_4 = 4 \Bigg\{ & \sum_{2 \leqslant a \leqslant n} \theta_{a^2-1}^4 \frac{a^2-3a+3}{a(a-1)} \\ & + \sum_{2 \leqslant a < b \leqslant n} \left[\theta_{a^2-1}^3 \theta_{b^2-1} \frac{4(2-a)}{\sqrt{a(a-1)b(b-1)}} \right. \\ & \left. + \theta_{a^2-1}^2 \theta_{b^2-1}^2 \frac{6}{b(b-1)} \right] \\ & + \sum_{2 \leqslant a < b < c \leqslant n} \theta_{a^2-1}^2 \theta_{b^2-1} \theta_{c^2-1} \frac{12}{\sqrt{b(b-1)c(c-1)}} \Bigg\}, \end{aligned} \tag{2.20}$$

where the details of the derivation are given in the Appendix.

The crucial point is that Eqs. (2.9) and (2.18) yield

$$\mathbf{b}\cdot\mathbf{b} = \theta_\alpha \theta_\alpha \tag{2.21}$$

and because the n^2-1 components b_i, on the one hand, and the $n-1$ θ_α, on the other hand, are independent it follows by differentiation of Eq. (2.21) that

$$\mathbf{b} = \theta_\alpha \mathbf{m}_\alpha, \quad \mathbf{m}_\alpha \equiv \frac{\partial \theta_\alpha}{\partial \mathbf{b}}, \tag{2.22}$$

as advertised in Eq. (1.7).

If we denote by b the one-column matrix with n^2-1 rows, by θ the one-column matrix with $n-1$ rows, and by m the rectangular matrix with matrix elements

$$m_{i\alpha} = \frac{\partial \theta_\alpha}{\partial b_i}, \quad \begin{array}{l} i = 1,2,3,\ldots,n^2-1, \\ \alpha = 1,3,8,\ldots,n^2-1, \end{array} \tag{2.23}$$

then Eqs. (2.21) and (2.22) become (where T indicates "transpose")

$$b^T b = \theta^T \theta \tag{2.24}$$

and

$$b = m\theta. \tag{2.25}$$

Inserting Eq. (2.25) into (2.24) shows that m is an orthogonal matrix

$$m^T m = 1 \tag{2.26}$$

or

$$\mathbf{m}_\alpha \cdot \mathbf{m}_\beta = \delta_{\alpha\beta}, \tag{2.27}$$

i.e., the $\mathbf{m}_\alpha$ form a system of $n-1$ orthonormal vectors.

We shall come back to the geometry of these $\mathbf{m}_\alpha$ later. At this point we return to Eq. (1.5) and observe that

$$\begin{aligned} M_0 &= (1/n) \operatorname{Tr} e^{i\lambda\cdot\mathbf{b}} \\ &= \frac{1}{n} \sum_{k=1}^{n} e^{i\varphi_k} \\ &= \frac{1}{n} \sum_{k=1}^{n} \exp i \left[-\theta_{k^2-1} \sqrt{\frac{2(k-1)}{k}} \right. \\ &\quad \left. + \sum_{l=k+1}^{n} \theta_{l^2-1} \sqrt{\frac{2}{l(l-1)}} \right] \end{aligned} \tag{2.28}$$

and

$$\mathbf{M} = -\frac{1}{2}\mathrm{Tr}\, i\lambda e^{i\lambda\cdot\mathbf{b}} = -\frac{1}{2}\mathrm{Tr}\frac{\partial}{\partial\mathbf{b}}e^{i\lambda\cdot\mathbf{b}} = -\frac{1}{2}\frac{\partial}{\partial\mathbf{b}}\mathrm{Tr}\, e^{i\lambda\cdot\mathbf{b}} = -\frac{n}{2}\frac{\partial M_0}{\partial\mathbf{b}}$$

$$= \frac{i}{2}\sum_{k=1}^{n}\left[\mathbf{m}_{k^2-1}\sqrt{\frac{2(k-1)}{k}} - \sum_{s=k+1}^{n}\mathbf{m}_{s^2-1}\sqrt{\frac{2}{s(s-1)}}\right]$$

$$\times \exp i\left[-\theta_{k^2-1}\sqrt{\frac{2(k-1)}{k}} + \sum_{l=k+1}^{n}\theta_{l^2-1}\sqrt{\frac{2}{l(l-1)}}\right]. \tag{2.29}$$

Quite explicitly for, say, $n=4$ we obtain the result

$$e^{i\lambda\cdot\mathbf{b}} = e^{i(\omega_3+\omega_8+\omega_{15})}\{\tfrac{1}{4}[1 + e^{-2i\omega_3} + e^{-i\omega_3-3i\omega_8}$$
$$+ e^{-i\omega_3-i\omega_8-4i\omega_{15}}] + \tfrac{1}{2}\lambda\cdot[\mathbf{n}_3(1-e^{-2i\omega_3})$$
$$+ \mathbf{n}_8(1+e^{-2i\omega_3} - 2e^{-i\omega_3-3i\omega_8})$$
$$+ \mathbf{n}_{15}(1+e^{-2i\omega_3}+e^{-i\omega_3-3i\omega_8}$$
$$- 3e^{-i\omega_3-i\omega_8-4i\omega_{15}})]\}, \tag{2.30}$$

where we have introduced for convenience

$$\omega_{k^2-1} \equiv \theta_{k^2-1}\sqrt{2/(k(k-1))}, \tag{2.31}$$

$$\mathbf{n}_{k^2-1} \equiv \mathbf{m}_{k^2-1}\sqrt{2/(k(k-1))}. \tag{2.32}$$

III. PRODUCT PARAMETRIZATION

In this section, we propose to generalize Eq. (1.4) to SU(n) in the form of a product of $n-1$ factors. With the summation-over-repeated-indices convention suspended we write

$$e^{i\lambda\cdot\mathbf{b}} = S^{-1}Se^{i\lambda\cdot\mathbf{b}}S^{-1}S$$
$$= S^{-1}e^{iS\lambda\cdot\mathbf{b}S^{-1}}S$$
$$= S^{-1}\left(\exp i\sum_{k=2}^{n}\lambda_{k^2-1}\theta_{k^2-1}\right)S$$
$$= \Pi_{k=2}^{n}P_{k^2-1}^{(n)}, \tag{3.1}$$

where

$$P_{k^2-1}^{(n)} = S^{-1}(\exp i\lambda_{k^2-1}\theta_{k^2-1})S = S^{-1}\left(e^{i\omega_{k^2-1}}\sum_{a=1}^{k-1}D_a + e^{i(1-k)\omega_{k^2-1}}D_k + \sum_{a=k+1}^{n}D_a\right)S$$
$$= (k-1)e^{i\omega_{k^2-1}}S^{-1}\left\{\frac{1}{n}\mathbf{1} + \sum_{p=k}^{n}\frac{\lambda_{p^2-1}}{\sqrt{2p(p-1)}}\right\}S$$
$$+ e^{i(1-k)\omega_{k^2-1}}S^{-1}\left\{\frac{1}{n}\mathbf{1} - \sqrt{\frac{k-1}{2k}}\lambda_{k^2-1} + \sum_{p=k+1}^{n}\frac{\lambda_{p^2-1}}{\sqrt{2p(p-1)}}\right\}S$$
$$+ S^{-1}\left\{\frac{n-k}{n}\mathbf{1} - k\sum_{p=k+1}^{n}\frac{\lambda_{p^2-1}}{\sqrt{p(p-1)}}\right\}S. \tag{3.2}$$

Next we note that Eqs. (2.13) and (2.22) imply

$$\sum_{k=2}^{n}\lambda\cdot\mathbf{m}_{k^2-1}\theta_{k^2-1} = \sum_{k=2}^{n}S^{-1}\lambda_{k^2-1}\theta_{k^2-1}S, \tag{3.3}$$

and therefore

$$\lambda\cdot\mathbf{m}_{k^2-1} = S^{-1}\lambda_{k^2-1}S, \tag{3.4}$$

because the θ_{k^2-1} for different k are independent. This means that we may rewrite Eq. (3.2) for $P_{k^2-1}^{(n)}$, $2\leqslant k\leqslant n$, as follows:

$$P_{k^2-1}^{(n)} = \mathbf{1} + [(k-1)/2][e^{i\omega_{k^2-1}} - e^{i(1-k)\omega_{k^2-1}}]$$
$$\times\lambda\cdot\mathbf{n}_{k^2-1} + [(k-1)e^{i\omega_{k^2-1}}$$
$$+ e^{i(1-k)\omega_{k^2-1}} - k]\left[\frac{1}{n}\mathbf{1} + \frac{1}{2}\sum_{p=k+1}^{n}\lambda\cdot\mathbf{n}_{p^2-1}\right]. \tag{3.5}$$

Quite explicitly for, say, $n=4$ we obtain the following product representation:

$$e^{i\lambda\cdot\mathbf{b}} = P_3^{(4)}P_8^{(4)}P_{15}^{(4)}, \tag{3.6}$$

where

$$P_3^{(4)} = \mathbf{1} + i\sin\omega_3\lambda\cdot\mathbf{n}_3 + (\cos\omega_3 - 1)$$
$$\times(\tfrac{1}{2}\mathbf{1} + \lambda\cdot\mathbf{n}_8 + \lambda\cdot\mathbf{n}_{15}), \tag{3.7}$$

$$P_8^{(4)} = \mathbf{1} + (e^{i\omega_8} - e^{-2i\omega_8})\lambda\cdot\mathbf{n}_8 + (2e^{i\omega_8} + e^{-2i\omega_8} - 3)$$
$$\times(\tfrac{1}{4}\mathbf{1} + \tfrac{1}{2}\lambda\cdot\mathbf{n}_{15}), \tag{3.8}$$

$$P_{15}^{(4)} = \mathbf{1}e^{i\omega_{15}} + (e^{-3i\omega_{15}} - e^{i\omega_{15}})(\tfrac{1}{4}\mathbf{1} - \tfrac{3}{2}\lambda\cdot\mathbf{n}_{15}), \tag{3.9}$$

while for $n=3$ we have

$$e^{i\lambda\cdot\mathbf{b}} = P_3^{(3)}P_8^{(3)}, \tag{3.10}$$

where

$$P_3^{(3)} = \mathbf{1} + i\sin\omega_3\lambda\cdot\mathbf{n}_3 + (\cos\omega_3 - 1)(\tfrac{2}{3}\mathbf{1} + \lambda\cdot\mathbf{n}_8), \tag{3.11}$$

$$P_8^{(3)} = \mathbf{1}e^{i\omega_8} + (e^{-2i\omega_8} - e^{i\omega_8})(\tfrac{1}{3}\mathbf{1} - \lambda\cdot\mathbf{n}_8); \tag{3.12}$$

and for $n=2$ we have

$$e^{i\lambda\cdot\mathbf{b}} = P_3^{(2)} = \cos\omega_3\mathbf{1} + i\sin\omega_3\lambda\cdot\mathbf{n}_3. \tag{3.13}$$

IV. GEOMETRY OF THE ADJOINT REPRESENTATION

Here we discuss the meaning of the $n-1$ orthonormal vectors $\mathbf{m}_\alpha$ involved in the parametrization of the adjoint representation according to Eq. (1.7): $\mathbf{b}=\theta_\alpha \mathbf{m}_\alpha$. The parameter space of the adjoint representation of SU(n), i.e., the space of n^2-1 real parameters b_i, $1\leqslant i\leqslant n^2-1$, is $\mathbf{R}^{n^2-1}$. As is well known, this space is isotropic under the group of rotations $O(n^2-1)$.

We are interested in the transformations that leave $I_k=\mathrm{Tr}(\boldsymbol{\lambda}\cdot\mathbf{b})^k$, $2\leqslant k\leqslant n$, invariant. The invariance of I_2 is the invariance of the length of $\mathbf{b}$ to which the group $O(n^2-1)$ corresponds. Since, however, we also demand invariance of I_k, $3\leqslant k\leqslant n$, we are dealing with a subgroup of $O(n^2-1)$ that leaves $I_3, I_4,..., I_n$ invariant.

It is perhaps worth noting that for $n=2$ we have just the one invariant I_2 and so as far as the adjoint representation is concerned SU(2) and $O(3)$ are indistinguishable. More precisely we have here the well-known isomorphism between $\mathrm{SU}(2)/Z(2)$ and $O(3)$. Since $\mathbf{R}^{n^2-1}$ is isotropic under $O(n^2-1)$ we have for $n=2$ that $\mathbf{R}^3$ is isotropic—there is only one kind of vector $\mathbf{m}_3$.

For $n\geqslant 3$ the corresponding $\mathbf{R}^{n^2-1}$ is not isotropic, there are $n-1$ different types of vectors $\mathbf{m}_\alpha$. These vectors may be distinguished by considering their little groups. Consider an arbitrary vector $\mathbf{b}$ defined by $B=\mathbf{b}\cdot\boldsymbol{\lambda}$, $B\in\mathrm{su}(n)$, and we ask for its little group. We can transform B into diagonal form in view of the theorem according to which a Hermitian matrix may be diagonalized by means of a unitary similarity transformation and this, of course, leaves all I_k invariant.

The resultant diagonal matrix is of the form

$$a_1\sum_{k=1}^{n_1} D_k + a_2 \sum_{k=n_1+1}^{n_2} D_k + \cdots + a_p \sum_{k=n_p+1}^{n} D_k, \tag{4.1}$$

where a_1 is the n_1-fold degenerate eigenvalue, a_2 is the (n_2-n_1)-fold degenerate eigenvalue,..., a_p is the $(n-n_p)$-fold degenerate eigenvalue. It is clear that the little group of the above matrix is

$$\mathrm{S}[\mathrm{U}(n_1)\times \mathrm{U}(n_2-n_1)\times\cdots\times \mathrm{U}(n-n_p)]. \tag{4.2}$$

In particular, this group is *minimal* if all eigenvalues are different, the corresponding $\mathbf{b}$ might be called *generic* and it is given by Eq. (1.7) with $\theta_\alpha\neq 0$ for all α. The minimal little group is

$$S[\mathrm{U}(1)^n], \tag{4.3}$$

whose dimension is $(n-1)$.

On the other hand the little group (4.2) is *maximal* if as many eigenvalues as possible are the same. All eigenvalues cannot be the same since these matrices have zero trace but we could have just two distinct eigenvalues. Thus the maximal little group is

$$\mathrm{SU}[\mathrm{U}(n-1)\times \mathrm{U}(1)], \tag{4.4}$$

whose dimension is $(n-1)^2$. The corresponding $\mathbf{b}$ might be called *special* and it is given by Eq. (1.7) with only $\theta_{n^2-1}\neq 0$:

$$\mathbf{b}_{\mathrm{special}} = \theta_{n^2-1}\mathbf{m}_{n^2-1}. \tag{4.5}$$

We can now state the meaning of the different unit vectors $\mathbf{m}_{a^2-1}$, $2\leqslant a\leqslant n$: their little groups are given by

$$\mathrm{S}[\mathrm{U}(a-1)\times \mathrm{U}(1)\times \mathrm{U}(n-a)],\quad 2\leqslant a\leqslant n. \tag{4.6}$$

We conclude with the remark that for SU(2) the little group for the generic and special case are the same—every direction in the SU(2) case is special. In SU(3) the little group for the special case is $\mathrm{S}[\mathrm{U}(2)\times\mathrm{U}(1)]$ whereas for the generic case it is $\mathrm{S}[\mathrm{U}(1)^3]$ and so they are different. This has been remarked upon by Michel and Radicati who called the special vectors q-vectors, and vectors orthogonal to them r-vectors. In our language this means that $\mathbf{m}_8$ is a q-vector and $\mathbf{m}_3$ an r-vector.

V. ADDITIONAL REMARKS

We thank the referee for calling to our attention some additional references. Of these, the work by Macfarlane, Sudbury, and Weisz[4] and Macfarlane[5] continues in the spirit of Ref. 1 to develop further parametrizations of unitary matrices in a manner that does not require the knowledge of the eigenvalues. Conceptually closer to our approach is the work of Barnes and Delbourgo[6] and Barnes, Dondi, and Sarkar.[7] In fact, these authors introduce and make extensive use of a set of $n-1$ orthonormal vectors $\mathbf{p}_A$ (A ranging over $n-1$ values). It follows from a comparison of their work with ours that our orthonormal vectors $\mathbf{m}_{k^2-1}$ are related to the $\mathbf{p}_A$ by

$$\mathbf{m}_{k^2-1} = \left\{(1-k)\mathbf{p}_k + \sum_{a=1}^{k=1}\mathbf{p}_a\right\}[k(k-1)]^{-1/2},$$
$$2\leqslant k\leqslant n.$$

In addition, Barnes and Delbourgo mention, but make no use of, another set denoted by $\mathbf{p}'_A$, which they find by applying the Schmidt orthogonalization procedure. Remarkably enough we find by comparison of their work and ours that

$$\mathbf{m}_{k^2-1} = \mathbf{p}'_{k-1},\quad 2\leqslant k\leqslant n.$$

The Schmidt orthogonalization procedure is by its very nature somewhat arbitrary. Our results would seem to indicate that this set $\mathbf{m}_{k^2-1}$ of $n-1$ orthonormal vectors arises natually and so may have a canonical significance. It must, of course, be noted that this set is natural relative to our choice of the λ_{k^2-1} matrices in the standard way given by Eq. (2.14).

APPENDIX

Here we derive the expression for I_4 given by Eq. (2.20). The derivation of Eq. (2.19) is similar and simpler:

$$\begin{aligned} I_4 &= \mathrm{Tr}(\lambda_\alpha\theta_\alpha)^4 \\ &= \sum_\alpha \theta_\alpha^4\,\mathrm{Tr}\,\lambda_\alpha^4 + 4!\sum_{\alpha<\beta}\left(\frac{\theta_\alpha^3\theta_\beta}{3!}\mathrm{Tr}\,\lambda_\alpha^3\lambda_\beta + \frac{\theta_\alpha^2\theta_\beta^2}{2!2!}\mathrm{Tr}\,\lambda_\alpha^2\lambda_\beta^2 + \frac{\theta_\alpha\theta_\beta^3}{3!}\mathrm{Tr}\,\lambda_\alpha\lambda_\beta^3\right) + 4!\sum_{\alpha<\beta<\gamma}\left(\frac{\theta_\alpha^2\theta_\beta\theta_\gamma}{2!}\mathrm{Tr}\,\lambda_\alpha^2\lambda_\beta\lambda_\gamma\right. \\ &\quad \left. + \frac{\theta_\alpha\theta_\beta^2\theta_\gamma}{2!}\mathrm{Tr}\,\lambda_\alpha\lambda_\beta^2\lambda_\gamma + \frac{\theta_\alpha\theta_\beta\theta_\gamma^2}{2!}\mathrm{Tr}\,\lambda_\alpha\lambda_\beta\lambda_\gamma^2\right) + 4!\sum_{\alpha<\beta<\gamma<\delta}\theta_\alpha\theta_\beta\theta_\gamma\theta_\delta\,\mathrm{Tr}\,\lambda_\alpha\lambda_\beta\lambda_\gamma\lambda_\delta. \end{aligned} \tag{A1}$$

Using Eq. (2.14) (for $a<b<c<d$):

$$\mathrm{Tr}\,\lambda^4_{a^2-1} = [2/(a(a-1))]^2[a-1+(a-1)^4]$$
$$= 4(a^2-3a+3)/(a(a-1)), \quad (A2)$$

$$\mathrm{Tr}\,\lambda^3_{a^2-1}\lambda_{b^2-1} = 4(2-a)/\sqrt{ab(a-1)(b-1)}, \quad (A3)$$

$$\mathrm{Tr}\,\lambda_{a^2-1}\lambda^3_{b^2-1} = 0, \quad (A4)$$

$$\mathrm{Tr}\,\lambda^2_{a^2-1}\lambda^2_{b^2-1} = 4/(b(b-1)), \quad (A5)$$

$$\mathrm{Tr}\,\lambda^2_{a^2-1}\lambda_{b^2-1}\lambda_{c^2-1} = 4/\sqrt{bc(b-1)(c-1)}, \quad (A6)$$

$$\mathrm{Tr}\,\lambda_{a^2-1}\lambda^2_{b^2-1}\lambda_{c^2-1} = 0, \quad (A7)$$

$$\mathrm{Tr}\,\lambda_{a^2-1}\lambda_{b^2-1}\lambda^2_{c^2-1} = 0, \quad (A8)$$

$$\mathrm{Tr}\,\lambda_{a^2-1}\lambda_{b^2-1}\lambda_{c^2-1}\lambda_{d^2-1} = 0. \quad (A9)$$

Introducing Eqs. (A2)–(A9) into (A1) yields Eq. (2.20).

[1]A. J. Macfarlane, A. Sudbury, and P. H. Weisz, Commun. Math. Phys. **11**, 77 (1968).
[2]L. Michel and L. A. Radicati, Ann. Inst. Henri Poincaré **18**, 185 (1973).
[3]S. P. Rosen, J. Math. Phys. **12**, 673 (1971).
[4]A. J. Macfarlane, A. Sudbury, and P. Weisz, Proc. R. Soc. London Ser. A **314**, 217 (1970).
[5]A. J. Macfarlane, J. Math. Phys. **21**, 2579 (1980).
[6]K. J. Barnes and R. Delbourgo, J. Phys. A **5**, 1043 (1972).
[7]K. J. Barnes, P. H. Dondi, and S. C. Sarkar, Proc. R. Soc. London Ser. A **330**, 389 (1972).

Local realizations of kinematical groups with a constant electromagnetic field. I. The relativistic case

Javier Negro and Mariano A. del Olmo
Departamento de Física Teórica, Facultad de Ciencias, Universidad de Valladolid, 47011 Valladolid, Spain

(Received 26 April 1989; accepted for publication 18 October 1989)

This paper is devoted to the study of the description of elementary physical systems interacting with an external constant electromagnetic field and the construction of their differential wave equations from a group-theoretical point of view. In this context certain local realizations of the Poincaré group are studied. The linearization of this problem is carried out by building the associated representation group that turns out to be the well-known Maxwell group. In this way the usual method (concerning local realizations) that has been employed in studying free systems to the interacting case is extended.

I. INTRODUCTION

It is well known that a physical system in quantum mechanics is described by a (semiunitary) projective representation of the group of space-time transformations that leaves invariant the quantum system.[1] If the representation is irreducible the physical system is said to be elementary. The physical relevant representations in the space-time description are characterized by locality, and they are called locally operating realizations or local realizations (l.rl.'s).[2] These realizations have been exhaustively studied in several papers.[2–4] It is possible to obtain the differential equations satisfied by the wave functions via the realizations associated to the system. This theory has been working when applied to free systems, but unfortunately the situation is quite different when interactions appear. There is not a general characterization of the interacting systems and the differential equations are then usually obtained modifying the free equations with the help of conditions such as invariance, minimal electromagnetic coupling, etc.

There are in the literature[5–7] some attempts to give a group-theoretical formulation of the most simple interaction case: an elementary particle with an external constant electromagnetic field (e.m.f.). This study, in spite of its limitations, is relevant if it allows us to answer some basic questions such as how the charge appears, what is the parameter of the interaction, how to obtain the local equations and the minimal coupling or the kind of local realizations linked with these equations. Thus Schrader,[6] in his paper of 1972, computes the irreducible representations of the covariance group of the Klein–Gordon and Dirac equations with minimal electromagnetic coupling (the Maxwell group) and compares them with those supported by the equations. He finds that each equation contains a family of irreducible representations. Hoogland[7] studies the local realizations of the invariance group of a constant e.m.f. subgroup of the Poincaré group, and finds the minimal coupling equations related with these local realizations. In this work we continue Hoogland's program not only considering a subgroup of the Poincaré group but the whole Poincaré group, so we will connect the method of Hoogland with the results of Schrader. Furthermore, we will pay attention to some questions posed in Hoogland's work, as we will see later. The work of Bacry *et al.*[5] on this subject is also worthy to mention. They study the invariance subgroup of the e.m.f. and its irreducible realizations. The origin of the electrical charge is explained in a similar way as the mass is in the Galilei group, i.e., it is connected with the second cohomology group. However, they assumed since the beginning the existence of the minimal coupling equations. In some sense the paper of Hoogland[7] completes that work by adding the locality concept. Let us finally quote that Becker and Hussin[8] have extended the Schrader's method to find a nonrelativistic Maxwell group.

Our paper is deeply concerned with the questions that have already been dealt with in the previous mentioned works. Our results are not completely new but they are understood from a different point of view, which in our opinion will contribute to the clarification of the whole subject. This work participates in Lévy–Leblond's[9] spirit about the reformulation of the quantum theory making use of symmetry principles.

This paper is structured as follows. Section II is devoted to a study of the local realizations of the Poincaré group when the physical system lives in a constant e.m.f. We assume the existence of a phase called factor system that depends not only on the group elements but also on the e.m.f. These new factor systems, now not trivial, of the Poincaré group define an extension of this group, the Maxwell group, as it is possible to see in Sec. III. In Sec. IV, we construct the local representations (l.rp.'s) of that group which give rise to the local realizations of the Poincaré group of Sec. II. That is, we prove that the Maxwell group is a representation group for the Poincaré group. Section V is devoted to the irreducible representations of the Maxwell group already computed by Schrader.[6] We will mainly pay attention to some details about the little groups and their invariants. In Sec. VI we formulate some invariant equations like Klein–Gordon or Dirac, starting from the local realizations, and we point out a method to obtain the corresponding minimal coupling equations when the e.m.f. is not constant. Some remarks and conclusions end the paper.

 0022-2488/90/030568-11$03.00

II. REALIZATIONS OF THE POINCARÉ GROUP WITH A CONSTANT ELECTROMAGNETIC FIELD

A. The general outlook

Let us consider an elementary physical system interacting with a constant e.m.f. in such a way that the field is not modified by such elementary quantum systems, i.e., the field is external. Thus, the transformation properties of that field are well known. In some sense the field can be described under a kinematical point of view, in a similar way as the space-time is. We will also suppose that the interacting physical system will be described by a realization of the Poincaré group such that if the field disappears then we will get a free system described by the corresponding irreducible local realization of the Poincaré group.

Making use of the above assumption about the kinematical character in the description of the external e.m.f., the Poincaré group P will act on the manifold $X \times F$, where X is the space-time manifold and F the set of constant e.m.f.'s, in the following way:

$$g\colon (x,f) \to (\Lambda x + a, \Lambda f) \equiv (gx, gf) \equiv g(x,f), \tag{2.1}$$

with $g \equiv (a,\Lambda) \in P$, $(x,f) \in X \times F$, and $(\Lambda f)^{\mu\nu} = \Lambda^{\mu}_{\alpha} \Lambda^{\nu}_{\beta} f^{\alpha\beta}$, f being the e.m. skewsymmetric tensor. Although the action of P in the space-time is transitive, its action on F is not; for this reason we must be limited to an orbit. Let $\theta_{(x_0,f_0)}$, be the orbit of $X \times F$ under P corresponding to the generic point (x_0,f_0). The isotopy group of the point (x_0,f_0) will be

$$\Gamma_{(x_0,f_0)} = \{g \in P \mid g(x_0,f_0) = (x_0,f_0)\}. \tag{2.2}$$

Different points of the same orbit have conjugated isotopy groups. The group $\Gamma_{(x_0,f_0)}$ is a subgroup of the Lorentz group L. Let θ_{f_0} be an orbit of F under P, with f_0 a generic element of F. We call Γ_{f_0} the isotopy group of f_0. The relationship between both orbits is clear:

$$\theta_{(x_0,f_0)} = X \times \theta_{f_0}, \tag{2.3}$$

and between the corresponding isotopy groups is

$$\Gamma_{f_0} = T_4 \wedge \Gamma_{(x_0,f_0)}.$$

The orbit $\theta_{(x_0,f_0)}$ is diffeomorphic to the homogeneous space $P/\Gamma_{(x_0,f_0)}$. If we take a normalized Borel cross section $s\colon P/\Gamma_{(x_0,f_0)} \to P$ defined by

$$s(x,f) = (x,\Lambda_f) \in P, \tag{2.4}$$

where $\Lambda_f \in L$ verifying $\Lambda_f f_0 = f$, we can write the elements of P as

$$g \equiv (a,\Lambda) = s(g(x_0,f_0))\gamma(g) \equiv s(g)\gamma(g) \tag{2.5}$$

with $\gamma(g) \in \Gamma_{(x_0,f_0)}$. If $x_0 = 0$ then

$$g \equiv (a,\Lambda) = (a,\Lambda_{gf_0})(0,\gamma(g)), \tag{2.6}$$

since $(a,1) \in \Gamma_{f_0}$ $\forall a \in T_4$ then if $g = (a,\Lambda)$, $\gamma(g) = \gamma(\Lambda)$ $\forall g \in P$, so we will usually write $\gamma(\Lambda)$ instead of $\gamma(g)$. On the other hand, the orbit θ_{f_0} is diffeomorphic to P/Γ_{f_0}. Having chosen a normalized Borel cross section $r\colon \theta_{f_0} \to P$ by $r(f) = (0,\Lambda_f)$, we can decompose the elements g of P in a unique way as

$$g \equiv (a,\Lambda) = (0,\Lambda_{gf_0})(\Lambda_{gf_0}^{-1}(a),\gamma(\Lambda)), \tag{2.7}$$

where $(\Lambda_{gf_0}^{-1}(a),\gamma(\Lambda)) \in \Gamma_{f_0}$ and Λ_{gf_0}, $\gamma(g)$ are defined as in the previous case.

In the sequel we will ever consider the universal covering group P^* of $P(P^* = T_4 \wedge \mathrm{SL}(2,C)$, while $P = T_4 \wedge L)$. Evidently the action of P^* on $X \times F$ is given by the canonical homomorphism existing between a group and its universal covering group. For the sake of simplicity, we will call P the universal covering of the Poincaré group, and we shall not make any distinction between the Lorentz group and its covering group $\mathrm{SL}(2,C)$.

The wave function describing a system interacting with an e.m.f. f_0 will be denoted by $\psi(x,f_0)$, $x \in X$, and it will live in the Hilbert space H_{f_0}. For another observer characterized by an element g of P relative to the initial reference frame, the wave function will be $\psi'(x,f)$, $\forall x \in X$ with $f = gf_0$ and it will belong to the Hilbert space H_f. The quantum system will be described up to a local equivalence by wave functions of any of the Hilbert spaces H_f, $f \in \theta_{f_0}$. That is, we assume the existence of a local relationship between the wave functions describing the same physical state in different frames. Mathematically,

$$\begin{aligned}\psi'(g(x,f)) &\equiv (U(g)\psi)(g(x,f)) \\ &= A(g;x,f)\psi(x,f), \quad \forall (x,f) \in \theta_{(x_0,f_0)},\end{aligned} \tag{2.8}$$

where A is a Borel nonsingular matrix-valued function, called a gauge matrix: $A\colon P \times \theta_{(x_0,f_0)} \to \mathrm{GL}(n,C)$. The set of transformations $\{U(g)\}_{g \in P}$ is a generalization of the well-known locally operating realizations.[2,3] Here the transformation also depends on the e.m.f. as well as on the group element and the space-time point, as usual. We call U a local realization in a constant field, or an F-local realization. (However, we shall speak of U as a l.rl. simply, since from the context there is no risk of confusion with the usual ordinary l.rl.'s.) Note that this kind of realizations with a local phase, i.e., depending on the group element and the point of the space-time, was also used by Jackiw[10] in his study of the two-dimensional conformal group.

The composition of the transformations $U(g')$ and $U(g)$ must have the same physical effect as $U(g'g)$ when they act on the state given by ψ. This leads to the following relationship between the corresponding gauge matrices:

$$A(g';g(x,f))A(g;x,f) = w(g',g;f)A(g'g;x,f), \tag{2.9}$$

where the phase w is a factor system of P which is a Borel function: $w\colon G \times G \times \theta_f \to U(1)$. Owing to the associative nature of the transformations, the factor systems are characterized by the following property:

$$\begin{aligned}w(g_3,g_2g_1;f)w(g_2,g_1;f) &= w(g_3,g_2;g_1f) \\ &\times w(g_3g_2,g_1;f).\end{aligned} \tag{2.10}$$

The relationship (2.10) shows that w is a two-cocycle.

We will say that two local realizations U and U' are locally pseudoequivalent if there are: (i) a local operator S defined in every H_f, $\forall f \in \theta_{f_0}$, i.e., $(S\psi)(x,f) = S(x,f)\psi(x,f)$, with $S\colon \theta_{(x_0,f_0)} \to \mathrm{GL}(n,C)$ a Borel matrix-valued function and (ii) a Borel function $\lambda\colon P \times \theta_{f_0} \to U(1)$ verifying

$$\begin{aligned}&(U'(g)\psi)(g(x,f)) \\ &= \lambda(g,f)[(SU(g)S^{-1})\psi](g(x,f)).\end{aligned} \tag{2.11}$$

Then the gauge matrices associated to local pseudoequivalent realizations are related by

$$A'(g;x,f) = \lambda(g,f)S(g(x,f))A(g;x,f)S^{-1}(x,f), \qquad (2.12)$$

and the corresponding factor systems by

$$w'(g',g;f) = w(g',g;f)\lambda(g',gf)\lambda(g,f)\lambda^{-1}(g'g,f). \qquad (2.13)$$

The relationship (2.13) defines a relation of equivalence among the factor systems of P. The set of equivalence classes of factor systems has a group structure, $H^2_F(P,\theta_{f_0};U(1))$.

B. The characterization of the factor systems

Now, our task is to characterize the factor systems appearing in the local realizations. First of all, we obtain a matrix realization of the group $\Gamma_{(x_0,f_0)}$, according to the expression (2.9) when we take the restriction of the gauge-matrix A to $\Gamma_{(x_0,f_0)}\times\{(x_0,f_0)\}$. Then the restriction of the factor system to $\Gamma_{(x_0,f_0)}\times\Gamma_{(x_0,f_0)}\times\{f_0\}$ is a factor system of $\Gamma_{(x_0,f_0)}$ in the usual sense, i.e.,

$$w(\gamma_3,\gamma_2\gamma_1;f_0)w(\gamma_2,\gamma_1;f_0) = w(\gamma_3,\gamma_2;f_0)w(\gamma_3\gamma_2,\gamma_1;f_0), \qquad (2.14)$$

that is, independent of f_0.

Secondly, the restriction of w to $\Gamma_{f_0}\times\Gamma_{f_0}\times\{f_0\}$ is also a factor system now associated to a unitary local realization of Γ_{f_0}. For this reason we are going to study the different isotopy groups Γ_{f_0} and its factor systems.

The manifold F splits into orbits of the type θ_{f_0} under the action of P. Different orbits can have isomorphic conjugated isotopy groups and then we say that they belong to the same stratum. The different strata are classified as follows.

There are two invariants[5] characterizing the elements f which belong to an orbit θ_{f_0} of f under P: $f\cdot f\,(=f_{\mu\nu}f^{\rho\sigma})$ and $^*f\cdot f\,(=\frac{1}{2}\epsilon_{\mu\nu\rho\sigma}f^{\rho\sigma}f^{\mu\nu})$, so we have five different kinds of orbits:

(i) $^*f\cdot f\neq 0$ ($^*f\cdot f=-4\mathbf{E}\cdot\mathbf{B}$). Where the electric and magnetic fields are defined by $E^i=f^{0i}$ and $B^i=\frac{1}{2}\epsilon^{ijk}f_{jk}$, $i,j,k=1,2,3$, respectively, as usual. In this case there is one element in every orbit, $\hat{f}(\hat{\mathbf{E}},\hat{\mathbf{B}})$, such that $\hat{\mathbf{E}}=(0,0,\hat{E}_3)$, $\hat{\mathbf{B}}=(0,0,\hat{B}_3)$, with $\hat{E}_3\in\mathbf{R}^*$, $\hat{B}_3\in\mathbf{R}^+$. Then the orbit is said to be of parallel type.

(ii) $^*f\cdot f=0$, $f\cdot f>0$ ($f\cdot f=2(\mathbf{B}^2-\mathbf{E}^2)$). We can consider here the following point in each orbit, $\hat{f}(\hat{\mathbf{E}},\hat{\mathbf{B}})$, $\hat{\mathbf{E}}=(0,0,0)$, $\hat{\mathbf{B}}=(0,0,\hat{B}_3)$, $\hat{B}_3\in\mathbf{R}^+$. This field is called of magnetic type.

(iii) $^*f\cdot f=0$, $f\cdot f<0$. Now, we can choose a field of electric type, $\hat{\mathbf{E}}=(0,0,\hat{E}_3)$, $\hat{\mathbf{B}}=(0,0,0)$, $\hat{E}_3\in\mathbf{R}^+$.

(iv) $^*f\cdot f=0$, $f\cdot f=0$, $f\neq 0$. There exists only one orbit. It is possible to consider as representative field the following one: $\hat{\mathbf{E}}=(1,0,0)$, $\hat{\mathbf{B}}=(0,1,0)$.

(v) $f=0$. The orbit is only one point.

The three first orbits belong to the same stratum. The isotopy group Γ_{f_0}, when we take $f_0=\hat{f}$, is made up by the space-time translations and the rotations and boosts along the z axis. Thus, Γ_{f_0} is a six parameter Lie group. The factor systems of Γ_{f_0} are well known[2] and each element of $H^2(\Gamma_{f_0},U(1))$ is labeled by two real parameters $[\epsilon,\beta]$. A two-cocycle lifting of $[\epsilon,\beta]$ is, for example,

$$w_{\epsilon,\beta}(k',k;f_0) = \exp\{i\tfrac{1}{2}(a'\wedge a^{\gamma'})_{\mu\nu}\Phi^{\mu\nu}\}, \qquad (2.15)$$

where k', $k\in\Gamma_{f_0}$, being $k=(a,\gamma)$, with $a\in T_4$ and $\gamma\in\Gamma_{(x_0,f_0)}$,

$$a^{\gamma}=(0,\gamma)(a,1)(0,\gamma)^{-1}, \quad (a'\wedge a)_{\mu\nu}=\tfrac{1}{2}(a'_\mu a_\nu - a'_\nu a_\mu),$$

and

$$\Phi^{\mu\nu}\equiv\begin{cases}\epsilon=\Phi^{03}=\Phi^{30}\\ \beta=\Phi^{12}=-\Phi^{21}\\ 0, \quad \text{otherwise.}\end{cases} \qquad (2.16)$$

Later we shall discuss about the dependence between the parameters ϵ,β and the field f_0, i.e., $\Phi^{\mu\nu}=\Phi^{\mu\nu}(f_0)$.

The isotopy group of the fourth orbit Γ_{f_0} is made up by the space-time translations and the parabolic transformations generated by J_2-K_1 and J_1+K_2 (being J_i and K_i, $i=1,2$, the generators of rotations and boosts along the i axis). In this case $H^2(\Gamma_{f_0},U(1))=\mathbf{R}^5$. We can choose the following lifting as a two-cocycle representative of each class:

$$w(k',k;f_0)=\prod_{i=1}^{5}w_i(k',k)$$

being

$$\begin{aligned}
w_1(k',k)&=\exp\{i\alpha_1\tfrac{1}{2}(a'\wedge a^{\gamma'})_{\mu\nu}f_0^{\mu\nu}\},\\
w_2(k',k)&=\exp\{i\alpha_2\tfrac{1}{2}(a'\wedge a^{\gamma'})_{\mu\nu}{}^*f_0^{\mu\nu}\},\\
w_3(k',k)&=\exp\{i\alpha_3[\rho'a_2+\sigma'a_1+\tfrac{1}{2}(\rho'^2-\sigma'^2)\\
&\qquad\times(a_0-a_3)]\}, \qquad (2.17)\\
w_4(k',k)&=\exp\{i\alpha_4[\rho'a_1+\sigma'a_2-\sigma'\rho'(a_0-a_3)]\},\\
w_5(k',k)&=\exp\{i\alpha_5\rho'\sigma\},
\end{aligned}$$

with $\alpha_i\in\mathbf{R}$, $i=1,...,5$; $k,k'\in\Gamma_{f_0}$, where $k=(a,\gamma(\rho,\sigma))$ with $a\in T_4$ and

$$\gamma(\rho,\sigma)=\exp\{\rho(J_2-K_1)+\sigma(J_1+K_2)\}\in\Gamma_{(x_0,f_0)}. \qquad (2.18)$$

However, not all the five-factor systems are physically interesting. Thus, if we also consider the discrete symmetries α_2 and α_4 become zero,[2,11] while the factor system α_5 cannot be related with a local realization.[3]

Finally, the last orbit is trivial because it corresponds to the zero field.

The unknown factor systems have some properties that we are going to enumerate. The interested reader can prove them making use of equivalences (2.13).

(i) $w(e,e;f)=w(e,g;f)=w(g,e;f)=1$, $\forall f\in\theta_{f_0}$, $\forall g\in P$, where e is the identity element of P.

(ii) $w(k,r(f);f_0)=w(r(f),k;f_0)=1$, $\forall k\in\Gamma_{f_0}$, $\forall f\in\theta_{f_0}$, r being the cross section from θ_{f_0} to P defined above.

(iii) $w(r(f'),r(f);f_0)=1$, $\forall f'$, $f\in\theta_{f_0}$.

(iv) $w(r(f)k',k;f_0)=w(k',k;f_0)$, $\forall k',k\in\Gamma_{f_0}$, $f\in\theta_{f_0}$,

$$w(k',r(f)k;f_0)=w(k'r(f),k;f_0).$$

(v) $w(r(f')k',r(f);f_0)=w(r(f'),r(f)k;f_0)=1$,

$\forall k,k'\in\Gamma_{f_0}$, $\forall f'$, $f\in\theta_{f_0}$.

(vi) $w(r(f')k',r(f)k;f_0)=w(r(f')k'r(f),k;f_0)$

$$\equiv w(\delta(r(f')k'r(f)),k;f_0),$$

$\forall k'$, $k\in\Gamma_{f_0}$, $\forall f',f\in\theta_{f_0}$, $\delta\in\Gamma_{(f_0)}$.

(vii) $w(g',g;f)=w(g',gr(f);f_0)$, $\forall g'$, $g\in P$, $\forall f\in\theta(f_0)$. (2.19)

In particular, if $g'=(a',1)$, $g=(a,1)$, and $r(f)=(0,\Lambda_f)$, then

$$w((a',1),(a,1);f) = w((a',1),(a,1)(0,\Lambda_f);f_0)$$
$$= w(\Lambda_f^{-1}(a'),\Lambda_f^{-1}(a);f_0), \qquad (2.19\text{vii}')$$

with $(\Lambda_f^{-1}(a),1) = (0,\Lambda_f^{-1})(a,1)(0,\Lambda_f)$.

We wish to make a comment on the meaning of these properties. If two groups are isomorphic, it can immediately be shown that this isomorphism can be extended in a canonical way to their extensions (in particular, extensions by $\mathbf{R}$). So if we have two subgroups H, H' of the same group G, related by an inner isomorphism given by $g\in G$ [i.e., g: $h\to g(h) = ghg^{-1}$, $h\in H$, $g(h)\in H'$] then, given a factor system w of H, it will determine the factor system w^g of H' defined by

$$w^g(g(h),g(h')) \equiv w(h,h'),$$
$$h,h'\in H, \quad g(h),g(h')\in H'. \qquad (2.20)$$

The property (2.19vii) tells us that if the factor system w is restricted to Γ_{f_0}, then the section r: $\theta_{f_0}\to P$ determines in that way factor systems defined in every Γ_f, $\forall f\in\theta_{f_0}$. Thus, let $k,k'\in\Gamma_{f_0}$, and $\Lambda_f(k),\Lambda_f(k')\in\Gamma_f$, then if we apply properties (2.19i–vii), we find that $w(\Lambda_f(k'),\Lambda_f(k);f) = w(k',k;f_0)$, in agreement with (2.20).

We shall apply the above properties to get the explicit form of the factor systems. First, we consider the factor systems associated with the strata of the three former kinds of orbits. If we restrict ourselves to Γ_{f_0}, we can give $w(k',k;f_0)$ by (2.15). It is straightforward to check that

$$w(\gamma(a'),\gamma(a);f_0) = w(a',a;f_0), \quad \forall\gamma\in\Gamma_{(x_0,f_0)} \qquad (2.21)$$

and

$$w(a',a;f) = w(\Lambda_f^{-1}(a'),\Lambda_f^{-1}(a);f_0), \qquad (2.22)$$

with the notation of (2.19). These properties imply that w is a scalar function under Lorentz transformations of the arguments $a,a'\in\mathbf{R}^4$ and $f\in\theta_{f_0}$. The values of $w(g',g;f)$ are independent of the choice of the section r of θ_{f_0} on P. The only independent scalars that can be made out of a, a', and f are $(a'\wedge a)\cdot f$ and $(a'\wedge a)\cdot{}^*f$, thus w must be an arbitrary function of such scalars. In fact, comparing with (2.15) we see that it must be a linear combination of both terms. The portionality constants can be interpreted as the electric and magnetic charges, respectively, of the interaction system. However, the term containing ${}^*f$ disappears, as well as the magnetic charge, when discrete symmetries are included.[2,11] Finally we obtain

$$w(a',a;f) = \exp\{\tfrac{1}{2}ie(a'\wedge a)_{\mu\nu}f^{\mu\nu}\}. \qquad (2.23)$$

Taking into account (2.17) we have $\epsilon = e\hat{E}_3$ and $\beta = e\hat{B}_3$, where $\hat{f} = (\hat{\mathbf{E}},\hat{\mathbf{B}})$ and $\hat{\mathbf{E}} = (0,0,\hat{E}_3)$, $\hat{\mathbf{B}} = (\hat{0},0,\hat{B}_3)$. Making use of the other properties we get

$$w(g',g;f) = \exp\{\tfrac{1}{2}ie(a'\wedge\Lambda'a)_{\mu\nu}(\Lambda'\Lambda f)^{\mu\nu}\}, \qquad (2.24)$$

where g', $g\in P$.

Now we shall consider the fourth orbit type. As noted above, when w is restricted to $\Gamma_{f_0}\times\Gamma_{f_0}\times\{f_0\}$ there are two kinds of inequivalent classes of factor systems given by $(2.17)_1$ and $(2.17)_3$. For the factor system $(2.17)_1$ we arrive at the same results as those we have obtained for the stratum just seen. However the factor systems obtained from $(2.17)_3$ have much different features than the ones already studied. In this case $w(k',k;f_0)\neq w(\gamma(k'),\gamma(k);f_0)$, with k', $k\in\Gamma_{f_0}$, and $\gamma',\gamma\in\Gamma_{(x_0,f_0)}$, so it is not a scalar function of its arguments. It means that w depends on the choice of the section r. Another section r' gives rise to a distinct system w', although equivalent to w. Due to this reason the explicit expression of w is not easy and will not be written down here since we shall not make any use of it in this paper.

While the factor system originated from $(2.17)_1$ has the same functional form as the ones existing in the other kinds of orbits (i), (ii), and (iii), the factor system that comes from $(2.17)_3$ is exclusive of orbit (iv). We also observe that the local realization associated to this strange factor system has few signs of "electromagnetic character." In fact, the construction of such realizations is done by making use of the subgroup Γ_{f_0} and the homogeneous space P/Γ_{f_0}, but it is not possible to obtain from them any object transforming like an e.m.f. does, such as it happens to be with the realizations associated to "covariant" factors. Because of such reasons we shall not take them into account for the rest of our work, although its study, when we do not consider its relation to the e.m.f., has its own interest.[2]

C. The gauge matrices

To complete this section it remains to write down the explicit expression of the gauge matrices, and this labor will be carried out in the sequel. We will make use of local equivalences whenever possible in order to compute a representative within each class of realizations.

Being given a normalized gauge matrix $A(g;x,f)$ of P (i.e., such that $A(e;x,f) = 1$, $\forall(x,f)\in\theta_{(x_0,f_0)}$) there is another one equivalent and normalized, A', such that $A'(s(g);x_0,f_0) = 1$, $\forall g\in G$. According to (2.12), it suffices to take $S(x,f) = A(s(x,f);x_0,f_0)$. Thus, the expression of a gauge matrix $\forall g\in P$ and at the point (x_0,f_0) is

$$A(g;x_0,f_0) = A(\gamma(g);x_0,f_0). \qquad (2.25)$$

And the general expression at any point (x,f) is

$$A(g;x,f) = w(gs(x,f);f_0)A(\gamma(gs(x,f));x_0,f_0), \qquad (2.26)$$

where w is the factor system of P associated to A according to (2.9). Making use of the explicit expression of w (2.24) we get

$$A(g;x,f) = \exp\{ie\tfrac{1}{2}(a\wedge\Lambda x)\cdot\Lambda f\}A(\gamma(gs(x,f));x_0,f_0),$$
$$\forall g\equiv(a,\Lambda)\in P. \qquad (2.27)$$

Note that a system of gauge matrices of P is determined by (i) a factor system of P and (ii) a finite dimensional matrix realization of the subgroup $\Gamma_{(x_0,f_0)}$.

Taking into account (2.27), the local realizations of the Poincaré group have the form

$$(U(g)\psi)(g(x,f)) = \exp\{ie\tfrac{1}{2}(a\wedge\Lambda x)\cdot\Lambda f\}$$
$$\times A(\gamma(gs(x,f));x_0,f_0)\psi(x,f). \qquad (2.28)$$

III. A REPRESENTATION GROUP FOR P: THE MAXWELL GROUP

Our plan is to find a new group for P, such that the local realizations of the Poincaré group that we are studying can be lifted to linear representations of this group.[3] Evidently such a representative group cannot be a central extension of P by $\mathbf{R}$, because as it is well known such extensions are trivial.

Let θ_{f_0} be the orbit under P of the e.m.f. f_0. We consider the real vector space $V(\theta_{f_0})$ of the real functions defined on θ_{f_0}. Let $V(\theta_{f_0})\times P$ be the set with the following composition law:

$$(\phi',g')(\phi,g) = (\phi' + g'\phi + \phi_{(g',g)},g'g), \tag{3.1}$$

with $(\phi,g)\in V(\theta_{f_0})\times P$. The action of P on $V(\theta_{f_0})$ is $(g\phi)(f) = \phi(g^{-1}f)$ and $\phi_{(g',g)}(f) = \mu(g',g;f)$, μ being the exponent of a factor system of P, i.e., $w(g',g;f) = \exp\{i\mu(g',g;f)\}$. It is easy to prove that this law endows that set with a group structure. However, there is a difficulty due to the infinite dimension of that extended group. Since the aim of the representation group concept is the building of a "minimal" group that allows us such a lifting, we may restrict ourselves to a finite dimensional subgroup in the following way. We take the minimal subspace of $V(\theta_{f_0})$ invariant under P including the functions $\phi_{(g',g)}(f) = \mu(g',g;f)$. Such a subspace is easy to compute. It is made up of functions $\alpha\in A(4,\mathbf{R})$, where $A(4,\mathbf{R})$ is the real vector space of 4×4 real skewsymmetric matrices, hereafter named A simply, which are defined by: $\alpha(f) = \alpha_{\mu\nu}f^{\mu\nu}, f\in\theta_{f_0}$. Thus we get a six-dimensional extension of P by the Abelian group A. The action of P on A is given by $(g\alpha)(f) = \alpha(g^{-1}f) \to (g\alpha)^{\mu\nu} = \Lambda^\mu_\rho\Lambda^\nu_\sigma\alpha^{\rho\sigma}$, $g = (a,\Lambda)\in P$, and $\phi^e{}_{(g,g')} = \frac{1}{2}e(\Lambda\Lambda')^{-1} \times (a\wedge\Lambda a')$, $g, g'\in P$ is a system of two-(exponential) cocycles of P with values on A with the mentioned action.

The composition law of the extended group is

$$(\alpha',g')(\alpha,g) = (\alpha' + \Lambda'\alpha + \tfrac{1}{2}e(\Lambda'\Lambda)^{-1} \times (a'\wedge\Lambda'a),g'g), \tag{3.2}$$

and the inverse of a generic element (α,g) is $(\alpha,g)^{-1} = (-\Lambda^{-1}\alpha, -\Lambda^{-1}a,\Lambda^{-1})$.

Now this group is useful when one is interested in the realizations associated to the exponent μ^e, corresponding to the factor system (2.23) labeled with the real number e, and any of the orbits θ_{f_0}, since in the course of the construction of the extended group, we have lost any track of orbits. In fact, as it will be checked later (Sec. IV), it is enough to consider only one value of e (e.g., $e = 1$) to obtain a representation group to lift any local realization (independently of the orbit or the value of e). Such a group ($e = 1$) will be called the Maxwell group,[6] M.

The canonical epimorphism p: $M\to P$ is given by $p(\alpha,g) = g$. Thus, we can extend the action of P on $X\times F$ (and on F) to M by $\bar g(x,f) = p(\bar g)(x,f) = g(x,f)$, $\forall\bar g\equiv(\alpha,g)\in M$. Then $\bar\Gamma_{f_0} = p^{-1}(\Gamma_{f_0})$ ($\bar\Gamma_{(x_0,f_0)} = p^{-1}(\Gamma_{(x_0,f_0)})$). Fixing a cross section $\bar s$: $M/\bar\Gamma_{f_0}\to M$ by

$$\bar s(x,f) = (0,x,\Lambda_f), \tag{3.3}$$

each element $\bar g$ of M factorizes as

$$\bar g\equiv(\alpha,a,\Lambda) = (0,a,\Lambda_{\bar g f_0})(\Lambda^{-1}{}_{\bar g f_0}\alpha,0,\gamma(g)), \tag{3.4}$$

where $(\Lambda_{\bar g f_0}^{-1}\alpha,0,\gamma(g))\equiv\gamma(\bar g)\in\bar\Gamma_{f_0}$. With this cross section we get

$$\bar s^{-1}(\bar g(x,f))\overline{gs}(x,f)\equiv\gamma(\overline{gs}(x,f)) = (\Lambda^{-1}{}_{\Lambda f}\alpha + \tfrac{1}{2}\Lambda^{-1} \times(a\wedge\Lambda x),0,\Lambda^{-1}{}_{\Lambda f}\Lambda\Lambda_f). \tag{3.5}$$

In a similar way the cross section $\bar r$: $M/\bar\Gamma_{(x_0,f_0)}\to M$ is defined by

$$\bar r(f) = (0,0,\Lambda_f). \tag{3.6}$$

The Maxwell group has 16 infinitesimal generators, ten of them, $M_{\mu\nu}$ and Π_μ, corresponding to the Poincaré group and the other six, $F_{\mu\nu}$, to the Abelian group A. The Maxwell Lie algebra is defined by the following nonzero commutators[6]

$$\begin{aligned}
[M_{\mu\nu},M_{\rho\sigma}] &= g_{\nu\rho}M_{\mu\sigma} + g_{\mu\sigma}M_{\nu\rho} - g_{\mu\rho}M_{\nu\sigma} - g_{\nu\sigma}M_{\mu\rho},\\
[M_{\mu\nu},F_{\rho\sigma}] &= g_{\nu\rho}F_{\mu\sigma} + g_{\mu\sigma}M_{\nu\rho} - g_{\mu\rho}M_{\nu\sigma} - g_{\nu\sigma}M_{\mu\rho},\\
[M_{\mu\nu},\Pi_\rho] &= g_{\nu\rho}\Pi_\mu - g_{\mu\rho}\Pi_\nu,\\
[\Pi_\mu,\Pi_\nu] &= F_{\mu\nu}.
\end{aligned} \tag{3.7}$$

IV. THE LOCAL REPRESENTATIONS OF THE MAXWELL GROUP

First of all, we check here that the local realizations of P with an external constant e.m.f. (2.8) are related with some local representations of the Maxwell group that we call "physical" representations.

Fixed a cross section ρ: $P\to M$ by $\rho(g) = (0,g)$ then we prove that the application R of M defined by

$$(R(\alpha,0)\psi)(x,f) = e^{ie\alpha\cdot f}\psi(x,f), \tag{4.1}$$

with $e\in R$ and $(\alpha,0)\in A$,

$$(R(0,g)\psi)((0,g)(x,f))\equiv(R(\rho(g))\psi)(\rho(g)(x,f)) = (U(g)\psi)(g(x,f)), \tag{4.2}$$

where U is the local realization of the Poincaré group given by (2.28) and $R(\alpha,g)\equiv R(\alpha,0)R(0,g)$ constitutes a local representation of M.

Indeed, the explicit form of R is

$$(R(\alpha,g)\psi)((\alpha,g)(x,f)) = e^{ie\alpha\cdot\Lambda f}A(g;x,f)\psi(x,f), \tag{4.3}$$

where the matrices $A(g;x,f)$ obey (2.9). If we define the new matrices $\bar A(\bar g;x,f)$ by $\bar A(\bar g;x,f) = e^{ie\alpha\cdot\Lambda f}A(g;x,f)$, it is straightforward to show that $\bar A$ is a system of gauge matrices with associated factor system $\bar w = 1$, i.e., they verify (2.9) with $\bar w = 1$, since $(0,g')(0,g) = (\mu(g',g),g'g)$ and $w(g',g;f) = \exp\{ie\mu(g',g;f)\}$, $\forall\bar g'$, $\bar g\in M$. Note that now we are working with factor systems of M. This means that the gauge matrices $\bar A(\bar g;x,f)$ originate a local linear representation of M.

A. The local representations as induced representations from $\bar\Gamma_{(x_0,f_0)}$

We will show now a property similar to that one satisfied by the ordinary (physical) local representations,[3] i.e., physical local representations are induced from the isotopy group of a fixed point, subgroup of the representation group. In our case we take the point (x_0,f_0) of the orbit $\theta_{(x_0,f_0)}$ and its isotopy group $\overline{\Gamma}_{(x_0,f_0)}$, a subgroup of the Maxwell group.

Being given a finite-dimensional linear representation σ of $\overline{\Gamma}_{(x_0,f_0)}$ and a normalized Borel cross section $\overline{s}$: $\theta_{(x_0,f_0)} \to M$, defined by (3.3) $\overline{s}(x,f) = (0,s(x,f)) = (0,x,\Lambda_f)$, where s was defined in (2.4), then the linear representation $R^{\overline{s}}_\sigma$ of M induced by σ of $\overline{\Gamma}_{(x_0,f_0)}$ has the following expression:

$$(R^{\overline{s}}_\sigma(\overline{g})\psi)(\overline{g}(x,f)) = \sigma(\overline{s}^{-1}(\overline{g}(x,f))\overline{g}\overline{s}(x,f))\psi(x,f). \qquad (4.4)$$

This induced representation is a local representation of M as it is easy to see if we realize that $\sigma(\overline{s}^{-1}(\overline{g}(x,f))\overline{g}\overline{s}(x,f))$ is a system of gauge matrices with an associated trivial factor system. It is possible to prove that if we take another cross section $\overline{s}'$, the induced representation $R^{\overline{s}'}_\sigma$ is locally equivalent to $R^{\overline{s}}_\sigma$, for that reason from now on we will remove the superindex $\overline{s}$. On the other hand, if we take two equivalent representations σ and σ' of $\overline{\Gamma}_{(x_0,f_0)}$, the induced representations R_σ and $R_{\sigma'}$ are locally equivalent as can be shown as well (for more details about induced and local representations see, for example, Refs. 3, 12). Conversely, if we choose the following special representation of $\overline{\Gamma}_{(x_0,f_0)}$,

$$\sigma(\alpha,0,\gamma) = \exp(ie\alpha\cdot f_0)\Delta(\gamma), \qquad (4.5)$$

with Δ a finite-dimensional representation of $\Gamma_{(x_0,f_0)}$, then taking into account (3.5) and replacing in (4.4) we get

$$\begin{aligned}(R_\Delta(\overline{g})\psi)(\overline{g}(x,f)) &= \exp\{ie[\Lambda^{-1}{}_{\Lambda f}\alpha\cdot f_0 + \tfrac{1}{2}\Lambda^{-1}{}_{\Lambda f}(a\wedge\Lambda x)\cdot f_0]\}\\ &\quad\times\Delta(\Lambda^{-1}{}_{\Lambda f}\Lambda\Lambda_f)\psi(x,f),\end{aligned} \qquad (4.6)$$

and when we restrict R_Δ to A, $R_\Delta|A$, we recover (4.1). Since

$$\begin{aligned}\Lambda^{-1}{}_{\Lambda f}\alpha\cdot f_0 + \tfrac{1}{2}\Lambda^{-1}{}_{\Lambda f}(a\wedge\Lambda x)\cdot f_0\\ = \alpha\cdot\Lambda f + \tfrac{1}{2}(a\wedge\Lambda f)\cdot\Lambda f\end{aligned} \qquad (4.7)$$

and $\gamma(gs(x,f)) \equiv s^{-1}(g(x,f))gs(x,f) = \Lambda^{-1}{}_{\Lambda f}\Lambda\Lambda_f \in \Gamma_{(x_0,f_0)}$, then R_Δ is a representation like that one defined by (4.3), with the gauge matrices given by (2.28); so any local physical representation of M is locally equivalent to a $\overline{\Gamma}_{(x_0,f_0)}$-induced representation.

Finally, if we give a physical local linear representation R of M and choose a cross section ρ of P on M, $\rho(g) = (0,g)$, then we have that $R(\rho(g)) = R(0,g) \equiv U(g)$ defines a local realization of P. So from R_Δ we get, up to a local equivalence a local realization of P like (2.28), as it was to be expected from the last results.

We recall that we are looking for a kind of local realizations of P such that they could describe the interaction between an elementary quantum system and a constant e.m.f. in such a way that if we eliminate that field we should obtain a realization of P describing a free physical system. These free realizations U of P are induced from finite-dimensional representations D of the Lorentz group L, and they have the expression

$$(U(g)\Psi)(gx) = D(\Lambda)\Psi(x). \qquad (4.8)$$

Now, suppose that the finite dimensional representation Δ of $\Gamma_{(x_0,f_0)}$ has the property that it can be extended to the whole Lorentz group, i.e., there exists a finite-dimensional representation D of L verifying $D|\Gamma_{(x_0,f_0)} = \Delta$. With this assumption, let U_D be the local realization of P induced by it according to (4.5) and (4.6), then if we make an equivalence by means of the operator T defined by

$$(T\psi)(x,f) = D(\Lambda_f)\psi(x,f) \equiv \Psi(x,f), \qquad (4.9)$$

the new local realization U'_D locally equivalent to U_D is $U'_D(g) = TU'_D(g)T^{-1}$, and its explicit expression is

$$\begin{aligned}(U'_D(g)\Psi)(g(x,f)) &= \exp\{\tfrac{1}{2}ie(a\wedge\Lambda x)\cdot\Lambda f\}\\ &\quad\times D(\Lambda)\Psi(x,f),\end{aligned} \qquad (4.10)$$

and of course, when $f \Rightarrow 0$ we recover the free realization (4.8). Such local realizations U' (or the corresponding rp.'s R) are called local covariant realizations (local covariant rp.'s).

B. Pseudoequivalence of local realizations

In preceding papers,[2,3,7] the local pseudoequivalence (l.pe.) of ordinary l.rp.'s of the extended invariance subgroup $\overline{\Gamma}_{f_0}$ of the field f_0 has been studied. Here we study the l.pe. of the l.rp's of P by means of its representation group M, and try to relate it with the well known one of $\overline{\Gamma}_{f_0}$. We shall reach the following main result:

Theorem 1: The local pseudoequivalence classes of local representations of M are in a one to one correspondence with the ones of $\overline{\Gamma}_{f_0}$.

Proof: First, we write down some basic facts on l.rp.'s of M. As already mentioned, the system of gauge matrices $\overline{A}$ of a l.rp., R, of M verify a relationship like (2.10) with $\overline{w} = 1$,

$$\overline{A}(\overline{g}';\overline{g}(x,f))\overline{A}(\overline{g};x,f) = \overline{A}(\overline{g}'\overline{g};x,f). \qquad (4.11)$$

Thus $\overline{A}|\overline{\Gamma}_{(x_0,f_0)}\times\{(x_0,f_0)\}$ is a matrix representation of $\overline{\Gamma}_{(x_0,f_0)}$, and $\overline{A}|\overline{\Gamma}_{f_0}$ gives rise to a l.rp. of the subgroup $\overline{\Gamma}_{f_0}$ in the Hilbert space H_{f_0}.

If we have two pseudoequivalent l.rp.'s, R and R' of M, i.e., there exists a local operator T and a one-coboundary map λ: $M\times\theta_{f_0}\to U(1)$ homomorphic in the first argument $\lambda(\overline{g}'\overline{g};f) = \lambda(\overline{g}';\overline{g}f)\lambda(\overline{g};f)$, such that $R'(\overline{g}) = \lambda(\overline{g};f)\mathrm{T}R(\overline{g})T^{-1}$ in the Hilbert space H_f, $f\in\theta_{f_0}$ [recall that they are strictly equivalent only when $\lambda(\overline{g};f) = 1$, $\forall f\in\theta_{f_0}$]. It is immediate to show that $R'|\overline{\Gamma}_{f_0}$ and $R|\overline{\Gamma}_{f_0}$ must be also pseudoequivalent and the corresponding gauge matrices restricted to $\overline{\Gamma}_{(x_0,f_0)}$ are pseudoequivalent matrix representations of $\overline{\Gamma}_{(x_0,f_0)}$.

Conversely, let R and R' be two l.rp.'s of M such that $R|\overline{\Gamma}_{f_0} \equiv R_{f_0}$ and $R'|\overline{\Gamma}_{f_0} \equiv R'_{f_0}$ are pseudoequivalent; then R and R' are pseudoequivalent too. The proof goes on as follows. The pseudoequivalence of R_{f_0} and R'_{f_0} implies the existence of a local operator S and a homomorphism λ: $\overline{\Gamma}_{f_0}\to U(1)$, such that $R'_{f_0}(\overline{k}) = \lambda(\overline{k})SR_{f_0}(\overline{k})S^{-1}$, $\forall\overline{k}\in\overline{\Gamma}_{f_0}$, then the associated gauge matrices must verify

$$\overline{A}'(\overline{k};x,f_0) = \lambda(\overline{k})S(\overline{k}x)\overline{A}(\overline{k};x,f_0)S^{-1}(x). \qquad (4.12)$$

Making use of an equivalence we arrive at the following expression for $\overline{A}$ (and $\overline{A}'$):

$$\overline{A}(\overline{g};x,f) = \overline{A}(\overline{\Lambda}_{\overline{g}f}^{-1}\overline{g}\overline{\Lambda}_f;\Lambda_f^{-1}x,f_0), \qquad (4.13)$$

where by (3.6) $\overline{r}(f) = (0,0,\Lambda_f) \equiv \overline{\Lambda}_f \in M$. Then if we apply (4.12) and call $\overline{\Lambda}_{\overline{g}f}^{-1}\overline{g}\overline{\Lambda}_f = \overline{k}(\overline{g},f)\in\overline{\Gamma}_{f_0}$, we obtain

$$\begin{aligned}\overline{A}'(\overline{k}(\overline{g},f);\overline{\Lambda}_{\overline{g}f}^{-1}x,f_0) &= \lambda(\overline{k}(\overline{g},f))S(\overline{k}(\overline{g},f)\Lambda_f^{-1}x)\overline{A}(\overline{k}(\overline{g},f);\Lambda_f^{-1}x,f_0)\\ &\quad\times S^{-1}(\Lambda_f^{-1}x).\end{aligned} \qquad (4.14)$$

With the new notation: $\lambda(\bar{k}(\bar{g},f))\equiv\lambda(\bar{g},f)$, $S(\Lambda_f^{-1}x)\equiv S(x,f)$, we can rewrite (4.14) as

$$\bar{A}'(\bar{g};x,f)=\lambda(\bar{g},f)S(\bar{g}(x,f))\bar{A}(\bar{g};x,f)S^{-1}(x,f). \tag{4.15}$$

Thus the gauge matrix systems $\bar{A}$ and $\bar{A}'$ are pseudoequivalent and so are the associated l.rp.'s R and R', respectively, of M.

In fact, since the pseudoequivalence classes of ordinary l.rl.'s of $\bar{\Gamma}_{f_0}$ are given in terms of those of the inducing subgroup $\bar{\Gamma}_{(x_0,f_0)}$,[3] the result given above can be stated in the following form.

Theorem 2: Two l.rp.'s R and R' of M, and, therefore, the associated l.rl.'s U and U' of P, are locally pseudoequivalent if and only if they are induced from pseudoequivalent matrix realizations of $\bar{\Gamma}_{(x_0,f_0)}$ and the homomorphism λ: $\bar{\Gamma}_{(x_0,f_0)}\rightarrow U(1)$ involved can be extended to a homomorphism of $\bar{\Gamma}_{f_0}$, or what is the same, it can be extended to a onecoboundary of M, as was defined at the beginning of this section.

C. Representations of *M* in the phase space

If we replace the manifold $X\times F$ for $X\times A$ the action of M, via P, on $X\times A$ is transitive. Such a manifold is diffeomorphic to the homogeneous space M/L. The action of a generic element $\bar{g}\equiv(\alpha,a,\Lambda)$ at a point $(x,\phi)\in X\times A$ is

$$(\alpha,a,\Lambda)(x,\phi)=(\Lambda x+a,\Lambda\phi+\alpha+\tfrac{1}{2}\Lambda^{-1}(a\wedge\Lambda x)), \tag{4.16}$$

with $\Lambda\phi\equiv\Lambda^{\mu}_{\rho}\Lambda^{\nu}_{\sigma}\phi^{\rho\sigma}$. Here, we have chosen the cross section $\tilde{s}$: $M/L\rightarrow M$ by $\tilde{s}(x,\phi)=(\phi,x,1)$.

If we fix the point $(x_0,\phi_0)\equiv(0,0)$ of $X\times A$, its isotopy subgroup is L, so we compute the induced representations of M from L. Being given a finite-dimensional representation D of L, the induced representation $\tilde{R}_D$ of M acts on the support space of functions defined on $X\times A$ in the following way:

$$(\tilde{R}_D(\bar{g})\tilde{\Psi})(\bar{g}(x,\phi)=D(\tilde{s}^{-1}(\bar{g}(x,\phi))\bar{g}\tilde{s}(x,\phi))\tilde{\Psi}(x,\phi). \tag{4.17}$$

After a brief computation we get

$$(\tilde{R}_D(\bar{g})\tilde{\Psi})(\bar{g}(x,\phi))=D(\Lambda)\tilde{\Psi}(x,\phi). \tag{4.18}$$

Note that in this representation the phase factor does not appear explicitly, but it is hidden at the point transformation $\bar{g}(x,\phi)$.

The relationship between F and A goes via duality, which means that each point of F labels a unitary representation of A. Thus the relationship between both kinds of representations of M is given by a Fourier transformation such as

$$\Psi(x,f)=\int\tilde{\Psi}(x,\phi)\exp(ief\cdot\phi)d\phi. \tag{4.19}$$

The representation R_D of M defined in the space of functions Ψ obtained from $\tilde{R}_D$ is

$$(R_D(\bar{g})\Psi)(x,f)=\int(\tilde{R}_D(\bar{g})\tilde{\Psi})(x,\phi)\exp(ief\cdot\phi)d\phi. \tag{4.20}$$

Substituting (4.16) and (4.18) we get

$$(R_D(\bar{g})\Psi)(x,f)=\exp(ie\tfrac{1}{2}(a\wedge\Lambda x)\cdot f)D(\Lambda)\times\Psi(\bar{g}^{-1}(x,f)). \tag{4.21}$$

This can be written in our usual notation as

$$(R_D(\bar{g})\Psi)(\bar{g}(x,f))=\exp(ie\tfrac{1}{2}(a\wedge\Lambda x)\cdot\Lambda f)\times D(\Lambda)\Psi(x,f), \tag{4.22}$$

which corrresponds to the local realization of the Poincaré group as expressed in (4.10).

That kind of representations $\tilde{R}$ of M is not of physical interest because it contains implicitly a family of representations corresponding to different values of the orbits in F under M and to different values of e as well, both being of physical meaning. The trouble is that each of these physical l.rp.'s labeled by the field's orbit and the charge, cannot be described in a local way out of the $\tilde{R}$ representations.

V. IRREDUCIBLE REPRESENTATIONS OF THE MAXWELL GROUP

As it is well known, local representations, in general, are not irreducible representations. In this section we will characterize the unitary irreducible representations (u.i.rp.'s) of M following closely Schrader's program,[6] so we will not give a very detailed explanation. Only some attention will be paid to the computation of the generators of the little groups. This study will clarify the origin of some invariants which will be employed later in the next section.

A. The Mackey–Kirillov method

Let $M=H\wedge L$ be the Maxwell group as a semidirect product, where H is the subgroup A of the 4×4 real skew-symmetric matrices and the space-time translations T_4. The composition law in H, according to (3.1), is

$$(\alpha',a')(\alpha,a)=(\alpha'+\alpha+\tfrac{1}{2}(a'\wedge a),a'+a). \tag{5.1}$$

Since H is a maximal nilpotent subgroup of M, the u.i.rp.'s of M (M satisfies the regularity conditions required by the inducing theory of Mackey[13]) can be computed inducing from the ones of H founded by means of Kirillov's theory.[14] This program is applied as follows.

Let $\hat{H}$ be the set of equivalence classes of u.i.rp.'s of H. The group M acts on $\hat{H}$ in the following way. Being given a u.i.rp. U which belongs to the class $\hat{U}\in\hat{H}$, then $\bar{g}$: $U\rightarrow U^{\bar{g}}$, with

$$U^{\bar{g}}(h)=U(\bar{g}^{-1}h\bar{g}),\quad\forall h\in H,\quad\forall\bar{g}\in M. \tag{5.2}$$

This action splits $\hat{H}$ in orbits $\Theta(\hat{U})$. If M_0 is the isotopy group of $\hat{U}$, then we must find the u.i.rp.'s of M_0 which when restricted to H are multiple of $\hat{U}$. Finally, we obtain the u.i.rp.'s of M induced from such representations of M_0. Note that $M_0=H\wedge L_0$, where $L_0\subset L$, and that the equivalence classes of u.i.rp.'s of M_0 are given in terms of: (i) an orbit $\Theta(\hat{U})$ of $\hat{H}$ under the action of M, and (ii) a u.i.rp. up to equivalence of L_0 if $H^2(L_0,U(1))=0$. The invariants of the representation specify these two conditions and so are referred to as orbital or little group invariants. In this subsection we develop, in several steps, the part of that program to find the orbital invariants. The next one will be devoted to the little group invariants.

1. The computation of $\hat{H}$

The infinitesimal generators of H are denoted by $F^{\mu\nu}$ and Π^{μ} (Sec. III). A generic element of the Lie algebra $\mathbf{H}$ of H is denoted (β,b), i.e.,

$$(\beta,b) \equiv F^{\mu\nu}\beta_{\mu\nu} + \Pi^{\mu}b_{\mu},$$

where $\beta \in A$ and $b \in \mathbf{R}^4$. The action of M on $\mathbf{H}$ is

$$\bar{g}\colon (\beta,b) \to \bar{g}(\beta,b) = (\Lambda\beta + a \wedge \Lambda b, \Lambda b), \tag{5.3}$$

where $\bar{g} = (\alpha,a,\Lambda) \in M$. Let $\mathbf{H}^*$ be the dual of the Lie algebra $\mathbf{H}$, if $(f,p) \in \mathbf{H}^*$ then

$$(f,p)\cdot(\beta,b) = -f^{\mu\nu}\beta_{\mu\nu} + p^{\mu}b_{\mu}, \tag{5.4}$$

with $f \in A$ and $p \in \mathbf{R}^4$. Now the action of M on $\mathbf{H}^*$ is given by

$$\bar{g}(f,p)\cdot(\beta,b) = (f,p)\cdot \bar{g}^{-1}(\beta,b). \tag{5.5}$$

Explicitly

$$\bar{g}(f,p) = (\Lambda f, \Lambda(p + af)), \tag{5.6}$$

where $af = a_{\mu}f^{\mu\nu}$. In particular the subgroup H acts on $\mathbf{H}^*$ by

$$h \equiv (\alpha,a,1)\colon (f,p) \to (f,p + af). \tag{5.7}$$

The construction of the u.i.rp.'s of H starts by fixing an element $(f,p) \in \mathbf{H}^*$ and computing a maximal subalgebra $\mathbf{N} \subset \mathbf{H}$ such that $[\mathbf{N},\mathbf{N}] \subset \mathrm{Ker}\ (f,p)$. Let N be the subgroup of H corresponding to $\mathbf{N}$. A unitary representation of N is given by

$$U_{(f,p)}\colon (\alpha,a) \to \exp\{i(pa - f\alpha)\}, \quad (\alpha,a) \in N. \tag{5.8}$$

This representation induces the representation $U_{(f,p)N}$ of H. Two induced representations $U_{(f,p)N}$ and $U_{(f',p')N'}$ are equivalent if and only if (f,p) and (f',p') lie in the same orbit under H. Thus, the equivalence classes of u.i.rp.'s of H are in a one to one correspondence with the orbits of $\mathbf{H}^*$ under H. According to (5.7) these orbits can be classified in the following way:

(i) Orbits of type I. Here $f = 0$, and the orbits have one point $\Theta(0,p) = \{(0,p)\}$, $p \in \mathbf{R}^4$.

(ii) Orbits of type II. They are characterized by $f\cdot{}^*f = C_2 \neq 0 (\Leftrightarrow \det(f) \neq 0)$. These orbits are four-dimensional and are given by $\Theta(f) = \{(f,p) \mid p \in \mathbf{R}^4\}$.

(iii) Orbits of type III. In this case $\det(f) = 0$, with $f \neq 0$. Here the orbits $\Theta(f,q) = \{(f,p) \mid {}^*fp = q,\ p \in \mathbf{R}^4\}$, $q \in \mathrm{Im}{}^*f$, are two-dimensional.

2. The action of M on $\hat{H}$

As we have mentioned above the group M acts on the representations $U_{(f,p)N}$ giving rise to orbits. Bearing in mind that the class of $U_{(f,p)N}$ is labeled by an orbit of $\mathbf{H}^*$ under H, then the points of any orbit under M are orbits under H, so we will call them superorbits, and are determined by (5.6). An element Λ of the Lorentz group acts on any orbit Θ by

(I) $\Lambda: \Theta(0,p) \to \Theta(0,\Lambda p)$. The superorbit originated here is denoted by $\Theta[p]$.

(II) $\Lambda: \Theta(f) \to \Theta(\Lambda f)$. Notation: $\Theta[f]$.

(III) $\Lambda: \Theta(f,q) \to \Theta(\Lambda f, \Lambda q)$. Notation: $\Theta[f,q]$.

The superorbits $\Theta[p]$ correspond to the well known u.i.rp.'s of the Poincaré group because of the null field.

The superorbits of type II, $\Theta[f]$, are made up of points (f,p), where $p \in \mathbf{R}^4$ and f belongs to a four-dimensional manifold given by the values of two invariants: $f\cdot f = C_1$ and $f\cdot{}^*f = C_2$, so it has dimension eight. The stability group of each point is a two-dimensional Abelian group. In each superorbit $\Theta[f]$ it is possible to find a characteristic element $\hat{f} \equiv (\hat{\mathbf{E}},\hat{\mathbf{B}})$ such that $\hat{\mathbf{E}} = (0,0,\hat{E}_3)$ and $\hat{\mathbf{B}} = (0,0,\hat{B}_3)$, with $\hat{B}_3 \in \mathbf{R}^*$ and, for example, $\hat{E}_3 \in \mathbf{R}^+$. So this orbit is called of parallel type. The stability group of $\hat{f}$ is generated by rotations and boosts along the z axis; i.e., J_3 and K_3.

The superorbits of type III, $\Theta[f,q]$, are classified depending on the values of f and q in the following way.

(1) Magnetic-like (M). There exists an element $[\hat{f},\hat{q}]$ where $\hat{f}(\hat{\mathbf{E}},\hat{\mathbf{B}})$ has the form $\hat{\mathbf{E}} = (0,0,0)$, $\hat{\mathbf{B}} = (0,0,\hat{B}_3)$, $\hat{B}_3 \in \mathbf{R}^+$. The characteristic values of q give the subclassification:

$$\begin{aligned}
&M_0,\ \hat{q} = (0,0,0,0),\\
&M_t,\ \hat{q} = \hat{B}_3\ (0,0,0,p^0),\\
&M_s,\ \hat{q} = \hat{B}_3\ (p^3,0,0,0),\\
&M_l,\ \hat{q} = \hat{B}_3(\pm 1,0,0,\pm 1) \quad \text{(four cases)}, \quad p_0,p^3 \in \mathbf{R}.
\end{aligned} \tag{5.9}$$

The invariants characterizing these orbits are:

(a) $f\cdot f = \hat{f}\cdot\hat{f} = C_1 > 0$,

(b) $f\cdot{}^*f = \hat{f}\cdot{}^*\hat{f} = 0 = C_2$,

(c) $({}^*f\cdot p)^2 = q^2 = C_3$, ($<0$ in M_s, $\geqslant 0$ otherwise). (5.10)

A sign operator[6] should also be added to distinguish the sign of p^0 in M_t, p^3 in M_s, etc. The isotopy groups of $\Theta(\hat{f},\hat{q})$ are generated by $\{J_3\}$ in superorbits M_t, M_s, M_l, and $\{J_3,K_3\}$ in M_0.

(2) Electric-like (E). In this case $\hat{f}(\hat{\mathbf{E}},\hat{\mathbf{B}})$, $\hat{\mathbf{E}} = (0,0,\hat{E}_3)$, $\hat{\mathbf{B}} = (0,0,0)$, $\hat{E}_3 \in \mathbf{R}^+$, and $\hat{q}$ gives rise to

$$\begin{aligned}
&E_0,\ \hat{q} = (0,0,0,0).\\
&E_q,\ \hat{q} = \hat{E}_3(0,-p^2,0,0), \quad p^2 \in \mathbf{R}^*.
\end{aligned} \tag{5.11}$$

The invariants here are the same as in the magnetic type superorbits but $C_1 < 0$ and $C_3 \leqslant 0$. The little groups of $\Theta(\hat{f},\hat{q})$ are given in terms of their generators by $\{J_3,K_3\}$ (E_0 superorbit) and $\{K_3\}$ (E_q superorbit).

(3) Radiation-like (R). The characteristic elements $[\hat{f},\hat{q}]$ are $\hat{f}(\hat{\mathbf{E}},\hat{\mathbf{B}})$, $\hat{\mathbf{E}} = (0,0,1)$, $\hat{\mathbf{B}} = (0,1,0)$, and

$$\begin{aligned}
&R_0,\ \hat{q} = (0,0,0,0).\\
&R_-,\ \hat{q} = (p^2,-p^2,0,0,), \quad p^2 \in \mathbf{R}^*.\\
&R_+,\ \hat{q} = (0,0,p^0 + p^1,0), \quad p^0 + p^1 \in \mathbf{R}^*.
\end{aligned} \tag{5.12}$$

The manifold of these superorbits is specified by the same invariants as in the magnetic and electric type superorbits, more a sign in R_-, R_+, however $C_1 = 0$, $C_3 \leqslant 0$. The isotopy groups of $\Theta(\hat{f},\hat{q})$ are generated by $\{J_2 - K_3, J_3 + K_2\}$ in R_0, R_- superorbits, and $\{J_3 + K_2\}$ in R_+.

Let us summarize the general facts on type III superorbits. (i) Concerning the invariants whose values label the superorbits: there are two of them which have the same functional form, $f\cdot f = C_1$ and $f\cdot{}^*f = C_2$, as in type II superorbits. The third one, $({}^*f\cdot p)^2 = C_3$, is specific of type III superorbits. (ii) Concerning the isotopy groups: they are one-dimensional and J_3, K_3, $J_3 + K_2$ are the respective generators when the point is the characteristic element of M, E and R superorbit. Only for some special cases, they are two-dimensional Abelian subgroups of L. These cases are M_0, E_0, R_0, and R_-.

B. The computation of little groups

As we have just stated, each superorbit contains a set of orbits each of which consists of a set of points (f,p). We recall that the action of M on this set was given by (5.6). The task to be done here is the computation of the generators of the isotopy subgroup of a point (f,p) belonging to any of the type II or III superorbits. Of course we already know its dimension: two for points of type II, and one for those of type III, except for the special cases already marked that will not be discussed along this paper. Since the isotopy subgroups are Abelian, we can show that a u.i.rp. of M can be fixed by the real eigenvalues of their hermitian generators, which correspond to their unidimensional u.rp.'s.

1. Type II superorbits

Let $(f,p)\in\Theta[f]$. We begin with the search of the subgroup of L which leaves invariant f. According to (5.6), the infinitesimal action of $\omega\in L$ on f is

$$\omega\colon f^{\mu\nu}\to f^{\mu\nu}+\omega^{\mu}_{\lambda}f^{\lambda\nu}+\omega^{\nu}_{\lambda}f^{\mu\lambda}. \tag{5.13}$$

If ω leaves f invariant then $\omega^{\mu}_{\lambda}f^{\lambda\nu}+\omega^{\nu}_{\lambda}f^{\mu\lambda}=0$. Making use of a vectorial notation $\omega(\mathbf{v},\boldsymbol{\theta}), f(\mathbf{E},\mathbf{B})$, we can rewrite it as

$$\begin{aligned}&\boldsymbol{\theta}\wedge\mathbf{E}+\mathbf{v}\wedge\mathbf{B}=\mathbf{0},\\&\boldsymbol{\theta}\wedge\mathbf{B}-\mathbf{v}\wedge\mathbf{E}=\mathbf{0}.\end{aligned} \tag{5.14}$$

There are two independent solutions,[5]

a) $(\mathbf{v},\boldsymbol{\theta})\propto(\mathbf{E},\mathbf{B})\to\omega^{\mu\nu}\propto f^{\mu\nu}$,

b) $(\mathbf{v},\boldsymbol{\theta})\propto(-\mathbf{B},\mathbf{E})\to\omega^{\mu\nu}\propto {}^*f^{\mu\nu}$.

The second step is to display how the elements of M having the form $(0,a,\Lambda)$, where Λ belongs to the Abelian subgroup generated by $\{f\cdot M, {}^*f\cdot M\}$, act on the other component p of the point (f,p). By an infinitesimal translation, a, we get, a: $p\to p+a\cdot f$, while the Lorentz transformation ω gives ω: $p\to p+\omega\cdot p$. The notation $(a\cdot f)^{\mu}=a_{\nu}f^{\nu\mu}$, $(f\cdot a)^{\mu}=f^{\mu\nu}a_{\nu}$, has been used. The element $(0,a,\omega)$ of the little group must satisfy

$$f\cdot a+\omega\cdot p=0 \tag{5.15}$$

whose solutions are as follows.

(a) If $\omega=f$, $f\cdot a-f\cdot p=0$. This implies that $a-p$ is in the Ker of f, but as we are in a type II superorbit, this is $\{0\}$, and $a=p$. The corresponding transformations are $\exp\{a_{\mu}\Pi^{\mu}+\frac{1}{2}\omega_{\mu\nu}M^{\mu\nu}\}$. If we use for the generators the notation of Sec. III, and if we insert the above solutions we obtain: $\exp\{p_{\mu}\Pi^{\mu}+\frac{1}{2}f_{\mu\nu}M^{\mu\nu}\}$. If we keep in mind that $F^{\mu\nu}$ is represented by $-if^{\mu\nu}$, Π^{μ} by ip^{μ}, and that $p_{\mu}\Pi^{\mu}$ acts on (f,p) as $\frac{1}{2}\Pi\cdot\Pi$, we have the following first hermitian generator of the little group

$$\tfrac{1}{2}\Pi_{\mu}\Pi^{\mu}-\tfrac{1}{2}F_{\mu\nu}M^{\mu\nu}\equiv\tfrac{1}{2}(\Pi^2-F\cdot M). \tag{5.16}$$

(b) If $\omega={}^*f$, we have $f^{\mu\nu}a_{\nu}-{}^*f^{\mu\nu}p_{\nu}=0$. Multiplying by ${}^*f$, we get $\frac{1}{4}{}^*f\cdot fa_{\nu}={}^*f_{\nu\mu}{}^*f^{\mu\lambda}p_{\lambda}$. Then we obtain the second generator substituting that expression for a_{μ}:

$$4(\Pi\cdot{}^*F)^2-({}^*F\cdot F)(M\cdot{}^*F). \tag{5.17}$$

However, the eigenvalues of these two generators cannot be arbitrary because the isotopy subgroup has a compact part and it provides discrete values. For example, the field $\hat{f}$ has $\{J_3,K_3\}$ as generators of its isotopy group, and it must be granted that J_3 should be represented by an operator with half-integer eigenvalues. Thus, the u.i.rp.'s are labeled by one real and one half-integer numbers. Nevertheless, such a restriction on the values of the Casimir operators is not considered at this moment because it has not a covariant expression.

2. Type III superorbits

The Lorentz subgroup which leaves invariant f is generated, as in type II, by $f\cdot M$ and ${}^*f\cdot M$. Then we look for solutions of the equation $f\cdot a-\omega\cdot p=0$.

(a) When we try with $\omega=f$, one solution, not unique this time, is $a=p$, and one gets the same generator as in type II superorbits (5.16):

$$\Pi^2-M\cdot F. \tag{5.18}$$

(b) However, when $\omega={}^*f$, there is no solution of that equation, as can be easily proved, except for the special cases quoted above, which are not included here. Observe that when the superorbit is of magnetic type, the isotopy group is compact and so its u.i.rp.'s are given by a half-integer number, but in the other cases of type II there is not such a constraint.

These computations end our program of characterizing the u.i.rp.'s of the Maxwell group. We shall make use of them in the next section devoted to the wave equations.

VI. INVARIANT EQUATIONS

To begin with, we consider the formulation of the equations describing a spinless elementary system interacting with an external constant e.m.f. The local realization of the Poincaré group and the equation when the system is free are

$$\begin{aligned}&(U_0(g)\psi)(gx)=\psi(x),\quad x\in X,\ g\in P,\\&p^2\psi(x)=m^2\psi(x),\quad p^{\mu}\equiv i\partial^{\mu},\quad m^2\in\mathbf{R}^+.\end{aligned} \tag{6.1}$$

The local representation of M which describes this system, now interacting, must be, c.f. (4.3) and (4.10),

$$\begin{aligned}&(R_0(\alpha,g)\Psi)((\alpha,g)(x,f))\\&\quad=\exp\{ie(\alpha\cdot\Lambda f+\tfrac{1}{2}(a\wedge\Lambda x)\cdot\Lambda f)\}\Psi(x,f),\end{aligned} \tag{6.2}$$

with $(\alpha,g)\in M$, $g=(a,\Lambda)$, $(x,f)\in X\times F$. This l.rp. of M is reducible and it contains a family of irreducible representations labeled by the values of the Casimir operators (Sec. V):

$$\begin{aligned}&\text{(a)}\quad F\cdot F,\\&\text{(b)}\quad F\cdot{}^*F,\\&\text{(c)}\quad \Pi^{\mu}\Pi_{\mu}-M^{\mu\nu}F_{\mu\nu},\\&\text{(d)}\quad 4({}^*F\cdot\Pi)^2-(M\cdot{}^*F)(F\cdot{}^*F).\end{aligned} \tag{6.3}$$

To this list, we should add the operators associated to some signs of some orbits, as we remarked before, but they are not taken into account here because they have not a local translation as differential equations. We recall that this is not a new problem, for example, we already know that the Klein–

Gordon equation describes systems with both positive and negative energies.

The explicit expressions of the Maxwell group generators Π, M, F in the l.rp. (6.2) are:

$$\Pi_0 = -\partial_0 - \tfrac{1}{2}ie\mathbf{x}\cdot\mathbf{E},$$
$$\mathbf{\Pi} = \boldsymbol{\partial} - \tfrac{1}{2}ie(\mathbf{x}\wedge\mathbf{B} + t\,\mathbf{E}),$$
$$M_{\mu\nu} = (x_\mu\partial_\nu - x_\nu\partial_\mu) + M_{\mu\nu}(f),$$
$$F_{\mu\nu} = ief_{\mu\nu}, \tag{6.4}$$

where $f^{\mu\nu}\equiv(\mathbf{E},\mathbf{B})$ and $M_{\mu\nu}(f)$ acts on the f-part of wave functions $\Psi(x,f)$ as $M_{\mu\nu}(f)\equiv(f^{\lambda}_{\mu}\partial_{\nu\lambda} + f^{\lambda}_{\nu}\partial_{\mu\lambda}) - (f^{\lambda}_{\mu}\partial_{\lambda\nu} - f^{\lambda}_{\nu}\partial_{\lambda\mu})$ with $\partial_{\mu\nu} = \partial/\partial f^{\mu\nu}$. Substituting in (6.3) the generators (6.4) and considering that $M(f)\cdot f$ has eigenvalue 0 when it acts on $\Psi(x,f)$, we get the following wave equations:

(a) $f^2\Psi(x,f) = C_1\Psi(x,f)$,

(b) $f\cdot{}^*f\Psi(x,f) = C_2\Psi(x,f)$,

(c) $(i\partial_\mu - eA_\mu(x,f))^2\Psi(x,f) = m^2\Psi(x,f)$,

where $A_\mu = (-\tfrac{1}{2}\mathbf{x}\cdot\mathbf{E}, \tfrac{1}{2}(\mathbf{x}\wedge\mathbf{B} + t\,\mathbf{E}))$. (6.5)

(d) We write this equation in a particular frame of reference and for a special orbit. The general expression is too complicated at least to understand its physical meaning. When the e.m.f. is of magnetic type, for example (the same conclusions can be drawn when it is of parallel type), our frame is such that we have the characteristic fields $\hat{\mathbf{E}} = (0,0,0)$, $\hat{\mathbf{B}} = (0,0,\hat{B})$. Then we obtain

$$[(i\partial_3)^2 - (i\partial_0)^2]\Psi(x,\hat{\mathbf{B}}) = C_3\Psi(x,\hat{\mathbf{B}}). \tag{6.5d}$$

If we mix Eqs. (6.5c) and (6.5d), we arrive at the equation

$$[(i\partial_1 - eA_1(x,\hat{\mathbf{B}}))^2 + (i\partial_2 - eA_2(x,\hat{\mathbf{B}}))^2]\Psi(x,\hat{\mathbf{B}}) = C'\Psi(x,\hat{\mathbf{B}}). \tag{6.5d'}$$

This is the known equation of the Landau levels of a charged particle interacting with a constant magnetic field. Only discrete values $C' = e\hat{B}(2n+1)$ are allowed for the existence of square integrable solutions,[3,7] and as it was mentioned before this is related with the compactness of the little group, i.e., when we use these values we get assured that $\exp\{2\pi J_3\}$ is represented by the identity. When the superorbits are of electric or light type, the corresponding equations (6.5d') allow continuous values of C'. For example, if $\hat{\mathbf{B}} = 0$, and $\hat{\mathbf{E}} = (0,0,\hat{E})$, Eq. (6.5d') becomes

$$[(i\partial_0 - eA_0(x,\hat{\mathbf{E}}))^2 - (i\partial_3 - eA_3(x,\hat{\mathbf{E}}))^2]\Psi(x,\hat{\mathbf{E}}) = C'\Psi(x,\hat{\mathbf{E}}). \tag{6.5d''}$$

Note that if $f\Rightarrow 0$, the only surviving invariant is (6.3c) and we find the Klein–Gordon equation as one can expect. So when (6.3c) is the only specified invariant, together with those selecting f, it describes a whole family of u.i.rp.'s of M, each labeled by the values of C' which are discrete or continuous, depending on the superorbit of f.

In the case of a $\frac{1}{2}$-spin system interacting with a constant e.m.f., first we write the associated local realization of the Poincaré group and the Dirac equation when the system is free, which are, respectively,

$$(U_{1/2}(g)\psi)(gx) = D_{1/2,1/2}(\Lambda)\psi(x),$$

$$\gamma^\mu p_\mu\psi(x) = m\psi(x), \tag{6.6}$$

with

$$\gamma^k = \begin{bmatrix}\sigma^k & 0\\ 0 & \sigma^k\end{bmatrix},\quad \gamma^0 = \begin{bmatrix}0 & I\\ I & 0\end{bmatrix},$$

$k = 1,2,3$, σ^k are the Pauli matrices, $\frac{1}{2}(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu) = i\sigma^{\mu\nu}$, $D_{1/2,1/2}(\Lambda) = \exp\{i\,\tfrac{1}{4}\omega_{\mu\nu}\sigma^{\mu\nu}\}$, $p_\mu = i\partial_\mu$, $g = (a,\Lambda)\in P$.

The local representation that we associate to the interacting system is, cf. (4.3) and (4.10),

$$(R_{1/2}(\alpha,g)\Psi)((\alpha,g)(x,f)) = \exp\{ie(\alpha\cdot f + \tfrac{1}{2}(a\wedge\Lambda x)\cdot\Lambda f)\}D_{1/2,1/2}(\Lambda)\Psi(x,f), \tag{6.7}$$

with $(\alpha,g)\equiv(\alpha,g,\Lambda)\in M$. This choice has been made because: (a) when $f\Rightarrow 0$, the representation (6.7) becomes that one of the free system (6.6), and (b) when in a reference system where $f = \hat{f}(\hat{\mathbf{E}},\hat{\mathbf{B}})$, with $\hat{\mathbf{E}} = (0,0,\hat{E})$, $\hat{\mathbf{B}} = (0,0,\hat{B})$, we consider rotations around the z axis this realization acts in the same way as it does a free $\frac{1}{2}$-spin system. This is part of the track of spin in an interacting system. In this representation the infinitesimal generators Π^μ, $M^{\mu\nu}$, $F^{\mu\nu}$ of M are given by the same expressions as they were in the last representation (6.4), except that $M^{\mu\nu}$ must be enlarged by adding the spin term $\frac{1}{4}(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu) = \frac{1}{2}i\sigma^{\mu\nu}$. Replacing such an expression for the generators in (6.3) we find the wave equations:

(a) $f\cdot f\Psi(x,f) = C_1\Psi(x,f)$,

(b) $f\cdot{}^*f\Psi(x,f) = C_2\Psi(x,f)$, (6.8)

(c) $[(i\partial_\mu - eA_\mu(x,f))^2 - \tfrac{1}{2}e\sigma^{\mu\nu}f_{\mu\nu}]\Psi(x,f) = m^2\Psi(x,f)$,

(d') when we have $\hat{f}(\hat{\mathbf{B}},\hat{\mathbf{E}})$, $\hat{\mathbf{E}} = (0,0,\hat{E})$ and $\hat{\mathbf{B}} = (0,0,\hat{B})$ we have, making use of (c)

$$[(i\partial_1 - eA_1(x,\hat{f}))^2 + (i\partial_2 - eA_2(x,\hat{f}))^2]\Psi(x,\hat{f}) = (C' \pm eB)\Psi(x,\hat{f}),$$

the signs + or − correspond to $+\frac{1}{2}$ or $-\frac{1}{2}$ spin, respectively. Equation (6.8c) is equivalent to the Dirac equation written as a second degree equation. It is connected with the two-component Feynmann–GellMann equations.[15] As clear from its origin, this equation gives the characterization of the little group representation, i.e., the spin characterization. In Casimir (6.3d) the spin appears if the e.m.f. verifies $F\cdot{}^*F\neq 0$. However, equations (6.8c) and (6.8d) jointly give the equation (6.8d'), which corresponds to Landau's levels with energy eigenvalues depending on the spin whenever the field is not of electric or radiation type. Again the origin of these discrete levels comes from the fact that $\exp\{2\pi J_3\}$ is now to be represented by −1.

Now we wish to make some comments on the above results. The l.rp. R_0 contains a number of local subrepresentations (l.srp.'s), each one characterized by the orbit of the e.m.f. (invariants C_1 and C_2), the mass (m^2) and, for example, the Landau levels (C'). It can be shown that all these l.srp.'s are locally inequivalent, and as a consequence they must describe different physical situations. The same can be said on the l.rp. $R_{1/2}$ used to describe interacting $\frac{1}{2}$-spin particles.

Note that we did not make use of the minimal electromagnetic coupling $(p - eA)$ from the beginning, but it appeared when the realizations of the generators were substituted in the Casimir expressions and after a counterbalance term. So, the Casimir $\Pi^2 - M\cdot F$ is not generalizable to the case in which the e.m.f. is not constant if we want to obtain the correct minimal electromagnetic coupling equations. There are two main questions to be answered when one is dealing with a general not constant e.m.f. The first one is to compute the factor systems $w(g',g;f)$ associated to a general field $f = f(x)$. The second one is to find the equations related with such realizations. In the computation of the factor systems the difficulties are derived from the fact that the group is infinite dimensional and that makes it impossible to generalize our preceding method. However, we will show another way that carries directly to the minimal coupling (6.5c) which may be applied to the not constant e.m.f. situation, once the general factor systems have been founded out.

Making use only of the generators of the Maxwell group, we are looking for an equation such that if $f \Rightarrow 0$ it becomes the free wave equation. At the point $x_0 = 0$, the generators of the isotropy group $\overline{\Gamma}_{f_0}$, Π^μ and $\widehat{M}^{\mu\nu}$ act in the same way as the P^μ and $\widehat{L}^{\mu\nu}$ (generators of the subgroup Γ_{f_0} of the Poincaré group.) Thus, if the free equation is

$$E(p^\mu)\Psi(x) = 0, \tag{6.9}$$

we can consider the support function space of the covariant representation of M (4.10) with the condition that at the point (x_0, f_0) verifies

$$E(\Pi^\mu)\Psi(x_0, f_0) = 0. \tag{6.10}$$

Independently of the correct equation satisfied by Ψ at any point (x,f) of the orbit $\theta_{(x_0,f_0)}$, it is clear that just at the point (x_0, f_0) the equation (6.10) is fulfilled. Now making use of Lorentz transformations, (6.10) becomes

$$E(\Pi^\mu)\Psi(x_0, f) = 0, \tag{6.11}$$

with $f \in \theta_{f_0}$ (remark that $x_0 = 0$). Finally, if we make a translation from x_0 to another point $x \in X$, we get

$$E(p^\mu - eA^\mu)\Psi(x,f) = 0, \tag{6.12}$$

which is the free equation (6.9) with the right minimal coupling. This procedure to construct wave equations is general enough and it gives more importance to the characterization of local realizations than it does to the description of irreducible representations. Observe that in this way, the Dirac equation with e.m. coupling appears directly and we do not have to make use of a second-degree equation.

VII. CONCLUSIONS

In the study of the problem of elementary systems interacting with an external constant e.m.f., we have made use of a new kind of local realizations of the Poincaré group, with an equivalence relation physically reasonable. It has been determined the classes of relevant factor systems, as well as a representation group known as the Maxwell group. The physically interesting l.rp.'s of this representation group were characterized and we fixed the equivalence criteria to classify them. Finally, we have compared such l.rp.'s with the irreducible ones by means of invariant equations. Some of these equations can be obtained from the ones for the corresponding free system by means of a minimal coupling.

This point of view is interesting by itself because it shows the importance that l.rl.'s have in describing interacting systems. If we study the l.rl.'s of the whole Poincaré group without having to restrict to an invariance subgroup, we obtain the maximum of information and furthermore, it allows a covariant formulation. For example, in this context we can speak about the spin of the interacting system. Also, this formulation leads to an easy comparation with free systems, which is of most importance. On the other hand, this is a natural extension of ordinary l.rl.'s so that the Maxwell group can be thought of as a particular case of the representation group concept, and then its origin should not be considered as an isolated special case.

This theory can be developed for any kinematical group. Indeed we have made use of a very general notation and the results may still be valid, with some slight changes if necessary, when the group is not Poincaré. However, we have to define a suitable constant e.m.f. consistent with such a kinematical group.[16] Therefore, we shall give some results in a forthcoming paper on the Galilean case.

ACKNOWLEDGMENTS

We wish to thank the Comisión Interministerial de Ciencia y Tecnología (CICYT) and the Caja de Ahorros de Salamanca for partial financial support.

[1]E. P. Wigner, Ann. Math. **40**, 149 (1939).
[2]H. Hoogland, "On local and global equivalence of wave functions in Quantum Mechanics," Ph.D. thesis, Nigmegen University, 1977.
[3]J. F. Cariñena, M. A. del Olmo, and M. Santander, J. Math. Phys. **26**, 2096 (1985).
[4]O. Garcia-Prada, M. A. del Olmo, and M. Santander, J. Math. Phys. **29**, 1083 (1988).
[5]H. Bacry, Ph. Combe, and J. L. Richard, Nuovo Cimento A **67**, 267 (1970).
[6]R. Schrader, Fortschr. Phys. **20**, 701 (1972).
[7]H. Hoogland, J. Phys. **20**, 797 (1978).
[8]J. Beckers and V. Hussin, J. Math. Phys. **24**, 1295 (1983).
[9]J. M. Lévy-Leblond, Riv. Nuovo Cimento **4**, 99 (1974).
[10]R. Jackiw, "Representations of the two-dimensional conformal group," Preprint CTP 1375, MIT (1986).
[11]Ph. Combe and J. L. Richard, Physica **71**, 66 (1974).
[12]D. J. Simms, *Lie Groups and Quantum Mechanics*, Lecture Notes in Mathematics, Vol. 52 (Springer, Berlin, 1969).
[13]G. W. Mackey, Acta Math. **99**, 267 (1958).
[14]A. A. Kirillov, Usp. Math. Nauk. **10**, 57 (1962).
[15]S. S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper and Row, New York, 1964).
[16]M. le Bellac and J. M. Lévy-Leblond, Nuovo Cimento B **14**, 217 (1973).

Highest weight irreducible unitary representations of Lie algebras of infinite matrices. I. The algebra gl(∞)

Tchavdar D. Palev[a)]

Arnold Sommerfeld Institut für Mathematische Physik, Technische Universität Clausthal, D-3392 Clausthal-Zellerfeld, West Germany

(Received 18 April 1989; accepted for publication 27 September 1989)

Two classes of irreducible highest weight modules of the general linear Lie algebra gl(∞), corresponding to two different Borel subalgebras, are constructed. Both classes contain all unitary representations. Within each module a basis is introduced. Expressions for the transformation of the basis under the action of the algebra are written down.

I. INTRODUCTION

During the last decade the infinite-dimensional Lie algebras became a field of increasing interest in several branches of mathematics and physics (see Refs. 1 and 2 and the references therein). We mention here as an example, related to the topic of the present paper, the results of the Kyoto school[3–6] on the applications of certain highest weight representations of gl(∞), which led to solutions of a large class of integrable equations. The algebra gl(∞), and more precisely its completion and central extension a_∞ (see Refs. 2 and 7), has several other applications (string theory, two-dimensional statistical models, etc.). This is due to the circumstance that the highest weight irreducible modules of a_∞ are carrier spaces for several representations of the infinite-dimensional Heisenberg algebra, the Virasoro algebra, and other Kac–Moody algebras (see Ref. 8 for applications in quantum physics).

The present paper is devoted to a study of two classes of highest weight representations of the infinite-dimensional general linear Lie algebra gl(∞), corresponding to two different matrix realizations of gl(∞), namely (see the notation at the end of the Introduction),

$$\mathrm{gl}_0(\infty) = \{(a_{ij})\,|\,i,j\in\mathbb{N}, \text{ all but a finite number of } a_{ij}\in\mathbb{C} \text{ are zero}\}, \tag{1.1}$$

$$\mathrm{gl}_\infty = \{(A_{ij})\,|\,i,j\in\mathbb{Z}, \text{ all but a finite number of } A_{ij}\in\mathbb{C} \text{ are zero}\}. \tag{1.2}$$

In both realizations the Lie bracket is the ordinary matrix commutator. Let $e_{ij}\in\mathrm{gl}_0(\infty)$, $i,j\in\mathbb{N}$ (resp. $E_{ij}\in\mathrm{gl}_\infty$, $i,j\in\mathbb{Z}$) be a Weyl matrix, i.e., a matrix with 1 on the ith row and the jth column and zero elsewhere. All e_{ij}, $i,j\in\mathbb{N}$ (resp. E_{ij}, $i,j\in\mathbb{Z}$) constitute a basis in gl(∞), usually called a Weyl basis. The commutation relations on gl(∞) are a linear extension of the relations

$$[e_{ij},e_{kl}] = \delta_{jk}e_{il} - \delta_{li}e_{kj}\,, \quad i,j,k,l\in\mathbb{N}\,, \tag{1.3}$$

$$[E_{ij},E_{kl}] = \delta_{jk}E_{il} - \delta_{li}E_{kj}\,, \quad i,j,k,l\in\mathbb{Z}. \tag{1.4}$$

Let

$$N_+^0 = \text{lin. env.}\{e_{ij}\,|\,i<j\in\mathbb{N}\} \subset \mathrm{gl}_0(\infty) \tag{1.5}$$

and

$$N_+ = \text{lin. env.}\{E_{ij}\,|\,i<j\in\mathbb{Z}\} \subset \mathrm{gl}_\infty \tag{1.6}$$

be subalgebras of gl(∞), consisting of all strictly upper triangular matrices in $\mathrm{gl}_0(\infty)$ and gl_∞, respectively. The irreducible gl(∞) module V is said to be a gl_∞ [resp. $\mathrm{gl}_0(\infty)$] highest weight module[2] [corresponding to the "Borel" subalgebra N_+ (resp. to N_+^0)], if there exists a vector $x_\Lambda\in V$, called a highest weight vector, such that

$$N_+x_\Lambda = 0\,, \quad E_{ii}x_\Lambda = M_ix_\Lambda\,, \quad \forall\, i\in\mathbb{Z} \tag{1.7}$$

$$(\text{resp.} \quad N_+^0x_\Lambda = 0, \quad e_{ii}x_\Lambda = m_ix_\Lambda, \quad \forall i\in\mathbb{N}). \tag{1.8}$$

The sequence of complex numbers

$$[M] \equiv \{M_i\,|\,i\in\mathbb{Z}\} \tag{1.9}$$

$$(\text{resp.} \quad [m] \equiv \{m_i\,|\,i\in\mathbb{N}\}) \tag{1.10}$$

is said to be the highest weight (that is, the signature) of the gl_∞ [resp. $\mathrm{gl}_0(\infty)$] module V.

The highest weight irreducible gl(∞) representations are infinite-dimensional analogs of the finite-dimensional irreducible highest weight representations of gl(n). In the case of gl(n), the finite-dimensional irreducible highest weight representations give all finite-dimensional irreducible representations, irrespective on the choice of the Borel subalgebra. This is no more the case for the infinite-dimensional algebras and, in particular, for gl(∞). Here we study the highest weight representations, corresponding to N_+^0 and N_+, which we called highest weight representations of $\mathrm{gl}_0(\infty)$ and gl_∞, respectively. The highest weight representations of $\mathrm{gl}_0(\infty)$ are a natural generalization of the representations of gl(n) in the Gel'fand–Zetlin basis.[9,10] This basis is, however, inappropriate for a direct generalization to gl_∞. Therefore, we first modify it, introducing a new labeling for the basis vectors. We write down the transformations of this basis, which we call a central basis (C basis), under the action of the generators. It is difficult to say how rich are the two classes of representations we consider. It turns out, however, that they contain all gl_∞ and $\mathrm{gl}_0(\infty)$ highest weight modules, for which

$$E_{ij}^+ = E_{ji}, \quad \forall\, i,j\in\mathbb{Z} \quad (\text{resp. } e_{ij}^+ = e_{ji}, \quad \forall\, i,j\in\mathbb{N}), \tag{1.11}$$

where a^+ is the Hermitian adjoint of the operator a with respect to a certain scalar product in the representation space. Following the terminology[2] accepted in the literature we call these representations unitary.

The exposition is organized as follows. In Sec. II we recall the way one introduces the Gel'fand–Zetlin basis in the gl(n) fidirmods (Sec. II A). In a similar way we define

---

[a)] Permanent address: Institute for Nuclear Research and Nuclear Energy, blvd. Lenin 72, 1184 Sofia, Bulgaria.

 0022-2488/90/030579-08$03.00

the C basis and write down the transformation of the basis under the action of the generators (Sec. II B). Section III is devoted to the study of the highest weight irreducible $gl(\infty)$ modules. First (Sec. III A), we investigate the $gl_0(\infty)$ modules. In Sec. III B we extend the concept of the C basis to the infinite-dimensional case and apply it to gl_∞.

Throughout the paper we use the following abbreviations and notation:

LA, LA's—Lie algebra, Lie algebras;

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

GZ basis—Gel'fand–Zetlin basis;

lin. env.$\{X\}$—the linear envelope of X;

$\mathbb{C}$—the complex numbers;

$\mathbb{R}$—the real numbers;

$\mathbb{Z}$—all integers;

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

$\mathbb{N}$—all positive integers;

$$[M] = \{M_i | i\in\mathbb{Z},\ M_i\in\mathbb{C}\}; \tag{1.12}$$

$$[m] = \{m_i | i\in\mathbb{N},\ m_i\in\mathbb{C}\}; \tag{1.13}$$

$$[m]_k = [m_{1k}, m_{2k},\dots, m_{kk}], \quad \text{where } m_{ik}\in\mathbb{C}; \tag{1.14}$$

$$[M]_{2k+\theta} = [M_{-k,2k+\theta}, M_{-k+1,2k+\theta},\dots, M_{k+\theta-1,2k+\theta}], \quad \theta = 0,1, \quad k\in\mathbb{N}; \tag{1.15}$$

$$l_{ij} = m_{ij} - i, \quad L_{ij} = M_{ij} - i; \tag{1.16}$$

$$(x,y)_0 = x,y, \quad (x,y)_1 = y,x; \tag{1.17}$$

$$\text{if } p\leqslant q\in\mathbb{Z}, \text{ then } [p,q] = \{k\,|p\leqslant k\leqslant q,\ k\in\mathbb{Z}\}; \tag{1.18}$$

$$\theta(x) = 1, \quad \text{for } x\geqslant 0, \quad \theta(x) = 0, \quad \text{for } x<0. \tag{1.19}$$

II. FINITE-DIMENSIONAL REPRESENTATIONS OF gl(2N+1)

Consider the LA $gl(2N+1)$, $N\in\mathbb{Z}_+$, and let V be a $gl(2N+1)$ fidirmod. In this section we introduce a basis in V, labeled in two different ways. The first one is the known Gel'fand–Zetlin basis (GZ basis),[9] which can be easily extended to a basis in the highest weight $gl_0(\infty)$ modules. The second basis, which we call a central basis (C basis), is more appropriate for generalization in the gl_∞ modules. In both cases the definition of the basis makes use of the following proposition.

Proposition 1: Consider the $gl(n)$ fidirmod $V(n)$ as a $gl(n-1)$ module. Then

$$V(n) = \sum_{i=1}^{k} \oplus\, V_i(n-1), \tag{2.1}$$

where each $V_i(n-1)$ is a $gl(n-1)$ fidirmod and all $V_1(n-1),\dots,V_i(n-1),\dots,V_k(n-1)$ carry inequivalent representations of $gl(n-1)$.

As in the case of $gl(\infty)$ it is convenient to use two different matrix realizations for the same LA $gl(2N+1)$, namely,

$$gl_0(2N+1) = \{(a_{ij}) | a_{ij}\in\mathbb{C},\quad i,j = 1,2,\dots,2N+1\}, \tag{2.2}$$

$$gl_{2N+1} = \{(A_{ij}) | A_{ij}\in\mathbb{C},\quad i,j = -N,-N+1,-N+2,\dots,N\}. \tag{2.3}$$

A. GZ basis[9,10]

Let e_{ij}, $i,j = 1,\dots,n$, be the Weyl generators of $gl_0(2N+1)$. Consider the chain of subalgebras

$$gl_0(n) \supset gl_0(n-1) \supset \cdots \supset gl_0(k) \supset \cdots \supset gl_0(2) \supset gl_0(1), \tag{2.4}$$

where $gl_0(k)$, $k = 1,\dots,n$, is a linear span of the generators e_{ij}, $i,j = 1,\dots,k$. Let

$$V([m]_n) \supset V([m]_{n-1}) \supset \cdots \supset V([m]_k) \supset \cdots \supset V([m]_2) \supset V(m_{11}) \tag{2.5}$$

be a flag of $gl_0(k)$ fidirmods $V([m]_k)$, $k = 1,2,\dots,n$, where

$$[m]_k \equiv [m_{1k}, m_{2k},\dots, m_{kk}] \tag{2.6}$$

is the signature of $V([m]_k)$. In an ordered basis of the Cartan subalgebra

$$e_{11}, e_{22},\dots, e_{kk} \tag{2.7}$$

of $gl_0(k)$ (so that the linear envelope of all positive root vectors is a Borel subalgebra), m_{ik} is the eigenvalue of e_{ii} on the highest weight vector $x(k)\in V([m]_k)$,

$$e_{ii}x(k) = m_{ik}x(k), \quad i = 1,\dots,k. \tag{2.8}$$

Since the fidirmods of $gl_0(1)$ are one dimensional, the flag (2.5) defines a vector (m) in $V([m]_n)$, which, according to Proposition 1, is uniquely defined by the signatures $[m]_n, [m]_{n-1},\dots,[m]_k,\dots,[m]_2, m_{11}$. Therefore, one can set

$$(m) \equiv \begin{bmatrix} [m]_n \\ [m]_{n-1} \\ \vdots \\ [m]_k \\ \vdots \\ [m]_2 \\ m_{11} \end{bmatrix} \equiv \begin{bmatrix} m_{1n}, & m_{2n}, & \dots, & m_{kn}, & \dots, & m_{n-1,n}, & m_{nn} \\ m_{1,n-1}, & m_{2,n-1}, & \dots, & m_{k,n-1}, & \dots, & m_{n-1,n-1} & \\ \dots & \dots & \dots & \dots & \dots & & \\ m_{1k}, & m_{2k}, & \dots, & m_{kk} & & & \\ \dots & \dots & \dots & & & & \\ m_{12}, & m_{22} & & & & & \\ m_{11} & & & & & & \end{bmatrix}. \tag{2.9}$$

The vectors (m) [see (2.9)], corresponding to all possible flags (2.5), constitute a basis in the $gl_0(n)$ fidirmod $V([m]_n)$. This is the Gel'fand–Zetlin basis (GZ basis).[9] For later use we summarize the results of Ref. 9 in a proposition (for derivation of the results, see Ref. 10).

Proposition 2: The n-tuple $[m]_n \equiv [m_{1n}, m_{2n}, ..., m_{nn}]$ is a signature of a $\mathrm{gl}_0(n)$ fidirmod $V([m]_n)$ if and only if $m_{in} \in \mathbb{C}$ and

$$m_{in} - m_{jn} \in \mathbb{Z}_+, \quad \forall i<j = 1,...,n. \tag{2.10}$$

The GZ basis $\Gamma([m]_n)$ in $V([m]_n)$ is given with all patterns (2.9) consistent with the betweenness condition

$$m_{i,j+1} - m_{ij} \in \mathbb{Z}_+, \quad m_{ij} - m_{i+1,j+1} \in \mathbb{Z}_+ . \tag{2.11}$$

The transformation of the GZ basis under $\mathrm{gl}(n)$ is completely defined from the relations

$$e_{k-1,k}(m) = \sum_{j=1}^{k-1} \left| \frac{\Pi_{i=1}^{k}(l_{ik} - l_{j,k-1})\Pi_{i=1}^{k-2}(l_{i,k-2} - l_{j,k-1} - 1)}{\Pi_{i\neq j=1}^{k-1}(l_{i,k-1} - l_{j,k-1})(l_{i,k-1} - l_{j,k-1} - 1)} \right|^{1/2} (m)_{j,k-1} , \tag{2.12}$$

$$e_{k,k-1}(m) = \sum_{j=1}^{k-1} \left| \frac{\Pi_{i=1}^{k}(l_{ik} - l_{j,k-1} + 1)\Pi_{i=1}^{k-2}(l_{i,k-2} - l_{j,k-1})}{\Pi_{i\neq j=1}^{k-1}(l_{i,k-1} - l_{j,k-1})(l_{i,k-1} - l_{j,k-1} + 1)} \right|^{1/2} (m)_{-j,k-1} , \tag{2.13}$$

$$e_{ii}(m) = \left[\sum_{k=1}^{i} m_{ki} - \sum_{k=1}^{i-1} m_{k,i-1} \right] (m) , \tag{2.14}$$

where $l_{ij} = m_{ij} - i$ and the pattern $(m)_{\pm i,j}$ is obtained from the pattern (m) by the replacement $m_{ij} \to m_{ij} \pm 1$.

All GZ basis vectors are weight vectors. The $\mathrm{gl}_0(k)$ highest weight vector $x(k)$ in $V([m]_k)$ is given with the pattern

$$x(k) \equiv (m), \quad \text{such that } m_{ii} = m_{i,i+1} = \cdots = m_{ik}, \quad \forall i = 1,...,k. \tag{2.15}$$

In this case

$$e_{ii} x(k) = m_{ik} x(k), \quad i = 1,...,k. \tag{2.16}$$

B. Central basis (C basis)

Consider as a basis in gl_{2N+1} the Weyl matrices

$$E_{ij}, \quad i,j \in [-N,N] . \tag{2.17}$$

For any $\theta = 0,1$ and $k \in [1-\theta, N]$, define a subalgebra

$$\mathrm{gl}_{2k+\theta} = \text{lin. env.}\{E_{ij} | i,j \in [-k, k+\theta-1]\} . \tag{2.18}$$

As an ordered basis in $\mathrm{gl}_{2N+\theta}$ take

$$E_{-k,-k}, E_{-k+1,-k+1}, E_{-k+2,-k+2}, ..., E_{k+\theta-1,k+\theta-1} . \tag{2.19}$$

Then the generators E_{ij}, $i<j \in [-k, k+\theta-1]$, are the positive root vectors in $\mathrm{gl}_{2k+\theta}$. Let g: $\mathbb{Z} \to \mathbb{N}$ be a bijective mapping, defined as

$$g(z) = 2|z| + \theta(z) \in \mathbb{N}, \quad \forall z \in \mathbb{Z} . \tag{2.20}$$

Proposition 3: The mapping φ, which is a linear extension of the relations

$$\varphi(E_{ij}) = e_{g(i),g(j)}, \quad i,j \in [-N,N] , \tag{2.21}$$

is an isomorphism of gl_{2N+1} on $\mathrm{gl}_0(2N+1)$. Its restriction on $\mathrm{gl}_{2k+\theta}$ gives an isomorphism of $\mathrm{gl}_{2k+\theta}$ on $\mathrm{gl}_0(2k+\theta)$ for each $\theta = 0,1$ and $k \in [1-\theta, N]$. The chain of subalgebras

$$\mathrm{gl}_{2N+1} \supset \mathrm{gl}_{2N} \supset \cdots \supset \mathrm{gl}_{2k+\theta} \supset \cdots \supset \mathrm{gl}_2 \supset \mathrm{gl}_1 \tag{2.22}$$

is transformed by φ into the chain (2.4):

$$\mathrm{gl}_0(2N+1) \supset \mathrm{gl}_0(2N) \supset \cdots \supset \mathrm{gl}_0(2k+\theta) \supset \cdots \supset \mathrm{gl}_0(2) \supset \mathrm{gl}_0(1) . \tag{2.23}$$

The proof is straightforward. The isomorphism (2.21) allows one to turn any $\mathrm{gl}_0(2k+\theta)$ fidirmod $V([m]_{2k+\theta})$ (for each $\theta = 0,1$ and $k \in [1-\theta, N]$) into a $\mathrm{gl}_{2k+\theta}$ fidirmod by simply setting $\varphi(E_{ij}) = e_{g(i),g(j)}$. In particular, any $\mathrm{gl}_0(2N+1)$ fidirmod $V([m]_{2N+1})$ is a gl_{2N+1} fidirmod. The transformation of the GZ basis (2.9) under the action of gl_{2N+1} reads

$$\varphi(E_{ij})(m) = e_{g(i),g(j)}(m), \quad \forall (m) \in \Gamma([m]_{2N+1}) . \tag{2.24}$$

Therefore φ is a representation of gl_{2N+1} in $V([m]_{2N+1})$ and the flag

$$V([m]_{2N+1}) \supset V([m]_{2N}) \supset \cdots \supset V([m]_{2k+\theta}) \supset \cdots \supset V([m]_2) \supset V(m_{11}) , \tag{2.25}$$

can be considered as a flag of fidirmods of the chain (2.22). Hence the GZ vector (2.9) $(n = 2N+1)$, defined by (2.25), can be labeled with the signatures of $V([m]_{2N+1}), V([m]_{2N}), ..., V([m]_{2k+\theta}), ..., V([m]_2)$, $V(m_{11})$ with respect to $\mathrm{gl}_{2N+1}, \mathrm{gl}_{2N}, ..., \mathrm{gl}_{2k+\theta}, ..., gl_2, gl_1$, correspondingly. Let $[M]_{2k+\theta}$ be the $\mathrm{gl}_{2k+\theta}$ signature of $V([m]_{2k+\theta})$. By definition, $[M]_{2k+\theta}$ consists of the eigenvalue of the representatives of the Cartan generators (2.19), i.e.,

$$\begin{aligned} &\varphi(E_{-k,-k}), \varphi(E_{-k+1,-k+1}), \\ &\quad \varphi(E_{-k+2,-k+2}), ..., \varphi(E_{k+\theta-1,k+\theta-1}) \\ &\quad = e_{2k,2k}, ..., e_{44}, e_{22}, e_{11}, e_{33}, e_{55}, ..., e_{2k+\theta-1,2k+\theta-1} , \end{aligned} \tag{2.26}$$

on the $\mathrm{gl}_{2k+\theta}$ highest weight vector $y(2k+\theta) \in V([m]_{2k+\theta})$:

$$\varphi(E_{ii}) y(2k+\theta) = M_{i,2k+\theta}\, y(2k+\theta), \quad i = -k, -k+1, -k+2, ..., k+\theta-1 . \tag{2.27}$$

Therefore we set

$$[M]_{2k+\theta} \equiv [M_{-k,2k+\theta}, M_{-k+1,2k+\theta}, M_{-k+2,2k+\theta}, ..., M_{k+\theta-1,2k+\theta}] , \tag{2.28}$$

and write the vector (2.9) in terms of the signatures (2.28) as

$$
(M) = \begin{bmatrix}
M_{-N,2N+1} & ,M_{-N+1,2N+1}, \cdots & ,M_{-1,2N+1} & ,M_{0,2N+1} & ,M_{1,2N+1}, \cdots & ,M_{N-1,2N+1} & ,M_{N,2N+1} \\
M_{-N,2N} & ,M_{-N+1,2N}, \cdots & ,M_{-1,2N} & ,M_{0,2N} & ,M_{1,2N}, \cdots & ,M_{N-1,2N} & \\
\cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\
 & \cdots & \cdots & \cdots & \cdots & & \\
 & & M_{-1,3} & ,M_{03} & ,M_{13} & & \\
 & & M_{-1,2} & ,M_{02} & & & \\
 & & & M_{01} & & &
\end{bmatrix}. \tag{2.29}
$$

Clearly in the case $n = 2N+1$, (2.9) and (2.29) are two different labelings for one and the same vector $(m) \equiv (M)$. In order to determine the relations between its $gl_0(2N+1)$ coordinates m_{ik} and its gl_{2N+1} coordinates M_{pq} we observe that up to a permutation the right-hand side of (2.26) coincides with the basis (2.7) of the Cartan subalgebra of $gl_0(2k+\theta)$. Therefore $[M]_{2k+\theta}$ is a signature of $V([m]_{2k+\theta})$ with respect to the new ordering (2.26) of the basis in the Cartan subalgebra of $gl_0(2k+\theta)$. The reordering of the basis [which is an inner automorphism and can be achieved with an action of a proper element of the Weyl group of $gl_0(2k+\theta)$] is changing the highest weight vector of $V([m]_{2k+\theta})$ from $x(2k+\theta)$ to $y(2k+\theta)$. It does not change the signature, however. Therefore, $[M]_{2k+\theta} = [m]_{2k+\theta}$. This equality, written in terms of the coordinates, reads

$$
M_{j-k-1,2k+\theta} = m_{j,2k+\theta}, \quad \forall\theta = 0,1, \quad k\in[1-\theta,N], \quad j\in[1,2k+\theta]. \tag{2.30}
$$

We call the basis, written in the notation (2.29), a central basis (C basis) in $V([M]_{2N+1}) \equiv V([m]_{2N+1})$ and denote it as $\Gamma([\mathrm{M}]_{2\mathrm{N}+1})$. The problem of writing the transformation of the C basis under the action of gl_{2N+1} reduces to a change of the variables (2.30) in the right-hand sides of Eqs. (2.12)–(2.14) for $n = 2N+1$ together with a replacement of the left-hand sides according to (2.24). We formulate the final result as a proposition. Let (M) be an arbitrary C-basis vector (2.30). Denote by

$$
(M)_{\pm[j,2k+\theta]}, \quad \theta = 0,1, \quad k\in[1-\theta,N], \quad j\in[-k,k+\theta-1], \tag{2.31}
$$

the C vector obtained from (M) by the replacement

$$
M_{j,2k+\theta} \to M_{j,2k+\theta} \pm 1 . \tag{2.32}
$$

Proposition 4: The $(2N+1)$-tuple

$$
[M]_{2N+1} = [M_{-N,2N+1}, M_{-N+1,2N+1}, \ldots, M_{N,2N+1}] \tag{2.33}
$$

is a signature of a gl_{2N+1} fidirmod $V([M]_{2N+1})$ if and only if

$$
M_{k,2N+1} \in \mathbb{C}, \quad k\in[-N,N]; \quad M_{i,2N+1} - M_{j,2N+1} \in \mathbb{Z}_+, \quad \forall i<j\in[-N,N]. \tag{2.34}
$$

The C basis $\Gamma([M]_{2N+1})$ in $V([M]_{2N+1})$ consists of all patterns (2.29) for which the $(N+1)(2N+1)$ labels

$$
M_{i,2k+\theta}, \quad \theta = 0,1, \quad k\in[1-\theta,N], \quad i\in[-k,k+\theta-1], \tag{2.35}
$$

take all possible values consistent with the conditions

$$
\begin{aligned}
&M_{i-\theta,2k+1-\theta} - M_{i,2k-\theta} \in \mathbb{Z}_+, \\
&M_{i,2k-\theta} - M_{i+1-\theta,2k+1-\theta} \in \mathbb{Z}_+ ,
\end{aligned} \tag{2.36}
$$

for all $\theta = 0,1, \quad k\in[1,N], \quad i\in[\theta-k,k-1]$.

The transformation of the C basis under the action of gl_{2N+1} is defined from the relations [we write E_{ij} instead of $\varphi(E_{ij})$]

$$
\begin{aligned}
E_{(-k,k-\nu)_\mu}(M) &= \sum_{j=\nu-k}^{k-1} \left| \frac{\Pi_{i=-k}^{k-\nu}[L_{i,2k+1-\nu} - L_{j,2k-\nu} + \mu(1-2\nu)]\,\Pi_{i=1-k}^{k-1-\nu}[L_{i,2k-1-\nu} - L_{j,2k-\nu} + \mu(1-2\nu)]}{\Pi_{i\neq j=\nu-k}^{k-1}[L_{i,2k-\nu} - L_{j,2k-\nu} + (-1)^\mu \nu]\,[L_{i,2k-\nu} - L_{j,2k-\nu} + (-1)^\mu(\nu-1)]} \right|^{1/2} \\
&\quad \times (M)_{(-1)^{\mu+\nu}[j,2k-\nu]}, \quad \forall\mu,\nu = 0,1, \quad \forall k\in[1,N] \quad [\text{see (1.16) and (1.17)}],
\end{aligned} \tag{2.37}
$$

$$
\begin{aligned}
E_{kk}(M) = \Bigg[&\sum_{i=-|k|}^{|k|+\theta(k)-1} M_{i,2|k|+\theta(k)} \\
&- \sum_{i=-|k|+1-\theta(k)}^{|k|-1} M_{i,2|k|+\theta(k)-1} \Bigg] (M) ,
\end{aligned} \tag{2.38}
$$

where $k\in[-N,N]$ and $M_{00} = M_{-1,0} = 0$.

The vector $y(2k+\theta)\in\Gamma([M]_{2N+1})$, for which

$$
M_{i,2j+\varphi} = M_{i,2k+1}, \quad \forall\varphi = 0,1, \quad i\in[-j,j+\varphi-1], \tag{2.39}
$$

is annihilated by all E_{ij}, $i<j\in[-k,k+\theta-1]$, i.e., it is a $gl_{2k+\theta}$ highest weight vector in $V([M]_{2k+\theta})$. In particular, the gl_{2N+1} highest weight vector $y_\Lambda \in V([M]_{2N+1})$ is the (only) one from (2.29), for which

$$M_{i,2k+\theta}=M_{i,2N+1},\quad \forall\theta=0,1,\quad k\in[1-\theta,N],$$
$$i\in[-k,k+\theta-1]. \tag{2.40}$$

III. IRREDUCIBLE REPRESENTATIONS OF gl(∞)

A. Highest weight modules of $\mathrm{gl}_0(\infty)$

Proposition 5: To each sequence of complex numbers

$$[m]\equiv[m_1,m_2,\ldots,m_k,\ldots]\equiv\{m_i|m_i\in\mathbb{C},\quad i\in\mathbb{N}\}, \tag{3.1}$$

such that

$$m_i-m_j\in\mathbb{Z}_+,\quad \forall i<j\in\mathbb{N}, \tag{3.2}$$

there corresponds an irreducible highest weight $\mathrm{gl}_0(\infty)$ module $V([m])$ with a signature (3.1). The basis $\Gamma([m])$ in $V([m])$, which we call a GZ basis, can be chosen to consist of all patterns

$$(m)\equiv\begin{bmatrix} m_1 & ,m_2 & ,\ldots & ,m_j & ,\ldots \\ \cdots & \cdots & \cdots & \cdots & \cdots \\ m_{1j} & ,m_{2j} & ,\ldots & ,m_{jj} & \\ \cdots & \cdots & & & \\ m_{12} & ,m_{22} & & & \\ m_{11} & & & & \end{bmatrix}\equiv\begin{bmatrix}[m]\\ \vdots \\ [m]_j \\ \vdots \\ [\mathrm{m}]_2 \\ \mathrm{m}_{11}\end{bmatrix}, \tag{3.3}$$

characterized by an infinite number of coordinates

$$m_{ij},\quad \forall j\in\mathbb{N},\quad i\in[1,j], \tag{3.4}$$

which are consistent with the conditions (1) for each pattern (m) there exists a positive integer [depending on (m)] $N[(m)]\in\mathbb{N}$, such that

$$m_{ij}=m_i,\quad \forall j>N[(m)],\quad i=1,\ldots,j; \tag{3.5}$$

$$(2)\ m_{i,j+1}-m_{ij}\in\mathbb{Z}_+,\quad m_{ij}-m_{i+1,j+1}\in\mathbb{Z}_+,\quad \forall i\leqslant j\in\mathbb{N}. \tag{3.6}$$

The transformation of the basis (3.3) is determined from the action of the generators [see (2.12)–(2.14)]

$$e_{k-1,k}(m)=\sum_{j=1}^{k-1}\left|\frac{\Pi_{i=1}^{k}(l_{ik}-l_{j,k-1})\Pi_{i=1}^{k-2}(l_{i,k-2}-l_{j,k-1}-1)}{\Pi_{i\neq j=1}^{k-1}(l_{i,k-1}-l_{j,k-1})(l_{i,k-1}-l_{j,k-1}-1)}\right|^{1/2}(m)_{j,k-1},\quad k=2,3,\ldots, \tag{3.7}$$

$$e_{k,k-1}(m)=\sum_{j=1}^{k-1}\left|\frac{\Pi_{i=1}^{k}(l_{ik}-l_{j,k-1}+1)\Pi_{i=1}^{k-2}(l_{i,k-2}-l_{j,k-1})}{\Pi_{i\neq j=1}^{k-1}(l_{i,k-1}-l_{j,k-1})(l_{i,k-1}-l_{j,k-1}+1)}\right|^{1/2}(m)_{-j,k-1},\quad k=2,3,\ldots, \tag{3.8}$$

$$e_{ii}(m)=\left[\sum_{k=1}^{i}m_{ki}-\sum_{k=1}^{i-1}m_{k,i-1}\right](m),\quad i\in\mathbb{N}, \tag{3.9}$$

and the identities

$$e_{ij}=[[[\cdots[[e_{i,i+1},e_{i+1,i+2}],\ e_{i+2,i+3}],\ldots],e_{j-2,j-1}],e_{j-1,j}], \tag{3.10}$$

$$e_{ij}=[[[\cdots[[e_{i,i-1},e_{i-1,i-2}],\ e_{i-2,i-3}],\ldots],e_{j+2,j+1}],e_{j+1,j}], \tag{3.11}$$

for $i<j$ and $i>j$, respectively.

The highest weight vector $(m)_0$ is the one from (3.3) for which

$$m_{ij}=m_i,\quad \forall j\in\mathbb{N},\quad i\in[1,\ldots,j]. \tag{3.12}$$

The proof will be a consequence of a few separate steps. First we introduce some terminology and notation. Let

$$(m)\equiv\begin{bmatrix}[m]\\ \vdots \\ [m]_n \\ \vdots \\ [m]_2 \\ m_{11}\end{bmatrix}\in\Gamma([m]). \tag{3.13}$$

Then

$$\text{(a)}\ [m]_n\equiv[m_{1n},m_{1n},\ldots,m_{nn}],\quad n=1,2,\ldots, \tag{3.14}$$

is said to be an n signature of (m);

$$\text{(b)}\ (m)^{\mathrm{up}(n)}\equiv\begin{bmatrix}[m]\\ \vdots \\ [m]_j \\ \vdots \\ [m]_{n+2} \\ [m]_{n+1}\end{bmatrix}\quad\text{and}\quad (m)^{\mathrm{low}(n)}\equiv\begin{bmatrix}[m]_n\\ \vdots \\ [m]_i \\ \vdots \\ [m]_2 \\ m_{11}\end{bmatrix} \tag{3.15}$$

are said to be an n-upper part and an n-lower part of (m), respectively.

Consider the subalgebra

$$\mathrm{gl}(n)=\{e_{ij}|i,j=1,\ldots,n\}\subset\mathrm{gl}_0(\infty). \tag{3.16}$$

Observation 1: Let

$$e\in\{e_{ii},e_{k,k-1},e_{k-1,k}|i=1,\ldots,n,\quad k=2,\ldots,n\}. \tag{3.17}$$

Then, for any $(m)\in\Gamma([m])$, $e(m)$ is a linear combination of vectors from $\Gamma([m])$ with one and the same n-upper part $(m)^{\mathrm{up}(n)}$. More generally, let $U(n)$ be the set of all polynomials of the operators (3.17). Then, for every $a\in U(n)$, $a(m)$ is a linear combination of vectors (3.13) with one and the same n-upper part $(m)^{\mathrm{up}(n)}$. In particular, this property holds if a is a $\mathrm{gl}(n)$ generator or any polynomial of $\mathrm{gl}(n)$ generators.

Denote by

$$\Gamma([m]_i|i\geqslant n) \subset \Gamma([m]) \tag{3.18}$$

the set of all vectors (3.13), that have one and the same $[m]_i$ signatures, for all $i\geqslant n$. Let

$$V([m]_i|i\geqslant n) \subset V([m]) \tag{3.19}$$

be the linear span of $\Gamma([m]_i|i\geqslant n)$. From (3.7)–(3.9) one concludes that $V([m]_i|i\geqslant n)$ is invariant with respect to the action of $\mathrm{gl}(n)$. To each $(m)\in\Gamma([m]_i|i\geqslant n)$ put in correspondence its n-lower part:

$$f(m) = (m)^{\mathrm{low}(n)}, \quad \forall (m)\in\Gamma([m]_i|i\geqslant n) . \tag{3.20}$$

Let

$$\Gamma([m]_n) = \{f(m)|(m)\in\Gamma([m]_i|i\geqslant n)\} . \tag{3.21}$$

Then f maps bijectively $\Gamma([m]_i|i\geqslant n)$ on $\Gamma([m]_n)$. Observe that $\Gamma([m]_n)$ consists of all GZ patterns of a $\mathrm{gl}(n)$ fidirmod with a signature $[m]_n$. Define an action of $\mathrm{gl}(n)$ on each $(m)\in\Gamma([m]_n)$ with the relations (2.12)–(2.14). Then the linear envelope $V([m]_n)$ of $\Gamma([m]_n)$ is a $\mathrm{gl}(n)$ fidirmod with a signature $[m]_n$. From a comparison of the relations (3.7)–(3.9) with (2.12)–(2.14) and observation 1 one easily concludes that for each e [see (3.17)] the diagram

$$\begin{tikzcd} V([m]_i|i\geqslant n) \arrow[r, "e"] \arrow[d, "f"'] & V([m]_i|i\geqslant n) \\ V([m]_n) \arrow[r, "e"'] & V([m]_n) \arrow[u, "f^{-1}"'] \end{tikzcd} \tag{3.22}$$

is commutative.

Observation 2: The subspace $V([m]_i|i\geqslant n) \subset V([m])$ is an irreducible finite-dimensional $\mathrm{gl}(n)$ module with a signature $[m]_n$ and a GZ basis $\Gamma([m]_i|i\geqslant n)$.

Let e_{ij}, e_{kl} be any two generators from $\mathrm{gl}_0(\infty)$ and (m) be an arbitrary vector from $\Gamma([m])$. Consider e_{ij}, e_{kl} as elements from $\mathrm{gl}(n) \subset \mathrm{gl}_0(\infty)$, where $n\geqslant\max(i,j,k,l)$. Then (m) is a vector from the $\mathrm{gl}(n)$ fidirmod $V([m]_i|i\geqslant n) \subset V([m])$ [see (3.18) and (3.19)] and therefore [observation 2]

$$(e_{ij}e_{kl} - e_{kl}e_{ij})(m) = (\delta_{jk}e_{il} - \delta_{li}e_{kj})(m) . \tag{3.23}$$

This proves a part of Proposition 9, which we formulate as a separate statement.

Conclusion 1: The operators e_{ij}, defined with the relations (3.7)–(3.11), turn the linear space $V([m])$ into a $\mathrm{gl}_0(\infty)$ module.

Consider any two vectors $x, y\in V([m])$,

$$x = \sum_{i=1}^{p} \alpha_i(m^i) , \quad y = \sum_{i=p+1}^{q} \alpha_i(m^i) ,$$

$$(m^i)\in\Gamma([m]) , \quad \alpha_i\in\mathbf{C} , \quad i = 1,\ldots,q . \tag{3.24}$$

Let

$$N = \max\{N[(m^i)]|i = 1,\ldots,q\} . \tag{3.25}$$

Then according to (3.5) all vectors (m^i), $i = 1,\ldots,q$, have one and the same k signatures, for every $k\geqslant N$:

$$[m^i]_k \equiv [m^i_{1k},m^i_{2k},\ldots,m^i_{kk}] = [m_1,m_2,\ldots,m_k] , \tag{3.26}$$

for every $i = 1,\ldots,q$ and for any $k\geqslant N$. Therefore $(m^i)\in V([m]_k|k\geqslant N) \subset V([m])$. Hence $x, y\in V([m]_k|k\geqslant N)$. The space $V([m]_k|k\geqslant N)$ is a $\mathrm{gl}(N)$ fidirmod (observation 2) and, therefore, there exists a polynomial P of the $\mathrm{gl}(N)$ generators such that $y = Px$.

Conclusion 2: The $\mathrm{gl}_0(\infty)$ module $V([m])$ is an irreducible $\mathrm{gl}_0(\infty)$ module.

Consider the vector $(m)_0\in\Gamma([m])$ [see (3.12)]. By a straightforward computation one obtains, from Eqs. (3.9) and (3.7), that

$$e_{ii}(m)_0 = m_i(m)_0, \quad \forall i\in\mathbf{N} , \tag{3.27}$$

and

$$e_{k,k+1}(m)_0 = 0, \quad \forall k\in\mathbf{N} . \tag{3.28}$$

The last equality together with the identity (3.10) gives that $e_{ij}(m)_0 = 0 ,\forall i<j\in\mathbf{N}$. Hence [see (1.7)]

$$N^0_+ \,(m)_0 = 0. \tag{3.29}$$

Conclusion 3: The irreducible $\mathrm{gl}_0(\infty)$ module $V([m])$ is a highest weight module with a signature

$$[m] \equiv [m_1,m_2,\ldots,m_j,\ldots] \tag{3.30}$$

and a highest weight vector $(m)_0$.

This completes the proof of Proposition 8.

Denote by $\mathfrak{F}$ the class of all highest weight irreducible $\mathrm{gl}_0(\infty)$ modules, which we have obtained so far, i.e., those with signatures (3.1) and (3.2). There is no claim that $\mathfrak{F}$ contains all highest weight irreducible $\mathrm{gl}_0(\infty)$ modules. We now proceed to show that the class $\mathfrak{F}$ is rich enough and it contains all unitary irreducible highest weight $\mathrm{gl}_0(\infty)$ modules.

Define an antilinear anti-involutive mapping ω in $\mathrm{gl}_0(\infty)$ in a standard way

$$\omega(e_{ij}) = e_{ji}, \quad \forall i,j\in\mathbf{N} , \tag{3.31}$$

$$\omega(a + b) = \omega(a) + \omega(b), \quad \forall a,b\in\mathrm{gl}_0(\infty) , \tag{3.32}$$

$$\omega(\lambda a) = \lambda^*\omega(a), \quad \forall a\in\mathrm{gl}_0(\infty), \quad \lambda\in\mathbf{C} , \tag{3.33}$$

where λ^* is the complex conjugate of λ. Introduce a scalar product (,) in every $V([m])$ by declaring that the GZ basis be orthonormed. From (3.7)–(3.9) one derives that

$$(ax, y) = (x,\omega(a)y), \quad \forall a\in\mathrm{gl}_0(\infty), \quad x, y\in V([m]) , \tag{3.34}$$

(i.e., that the scalar product is contravariant with respect to ω) if and only if the signature $[m] = \{m_i|i\in\mathbf{N}\}$ consists of real numbers, $m_i\in\mathbf{R}$, $\forall i\in\mathbf{N}$. Clearly, (3.34) holds if and only if (1.11) is fulfilled. By definition (see, for instance, Ref. 2) such representations are called unitary. Thus a module $V([m]) \subset \mathfrak{F}$ carries a unitary representation of $\mathrm{gl}_0(\infty)$ only if its signature is real, i.e., if its coordinates are real numbers. We denote all such modules by $\mathfrak{F}_R$, $\mathfrak{F}_R \subset \mathfrak{F}$.

Let $V([m])$ be an arbitrary unitary irreducible highest weight $\mathrm{gl}_0(\infty)$ module with a signature $[m] \equiv [m_1,m_2,\ldots,m_i,\ldots,m_j,\ldots]$. Denote by $(m)_0$ its highest weight vector. Then

$$e_{ii}(m)_0 = m_i(m)_0, \quad \forall i\in\mathbf{N} . \tag{3.35}$$

Take any two positive integers $i<j\in\mathbb{N}$. Choose $j\leqslant n\in\mathbb{N}$ and let $gl(n)$ be the subalgebra (3.16) of $gl_0(\infty)$. The representation of $gl(n)$ in $V([m])$ is unitary and, in general, reducible. Let $V(n)$ be the irreducible unitary $gl(n)$ submodule in $V([m])$, which contains $(m)_0$. All such modules are finite dimensional[9] and with real signatures. The vector $(m)_0$ is the highest weight vector in $V(n)$ and, therefore, the signature of $V(n)$ is $[m_1,...,m_i,...,m_j,...,m_n]$. Hence $m_i\in\mathbb{R}$, $i=1,...,n$, and, according to Proposition 2,

$$m_i - m_j \in \mathbb{Z}_+ .$$

Since the latter holds for any $i<j\in\mathbb{N}$, the signature $[m]$ of $V(n)$ is a real signature of the type (3.1) and (3.2) and, therefore, $V(n)\subset \mathfrak{F}_R$. We have obtained the following result.

Proposition 6: The irreducible highest weight $gl_0(\infty)$ module $V([m])$ carries an unitary irreducible highest weight representation of $gl_0(\infty)$ if and only if $V([m])\in\mathfrak{F}_R$.

If one requires an addition that the real form of $gl_0(\infty)$, which is a linear envelope of all generators $i(e_{pq}+e_{qp})$, $e_{pq}-e_{qp}$, $p,q\in\mathbb{N}$, is integrable to a unitary representation of the group $U(\infty)$, then the coordinates of the signatures $[m]$ should be integers, $m_i\in\mathbb{Z}$ [as it is for the unitary representations of $U(n)$].

The unitary representations of the group U_0, which is a closure of $U(\infty)$ in the group U of all unitary operators in an infinite-dimensional separable Hilbert space V, were studied by Kirillov.[11] He mentioned the possibility of introducing a GZ basis in V without specifying explicitly the conditions for selecting the basis and, in particular, condition (1) in Proposition 5.

Observe that (contrary to the finite-dimensional case) the algebra $gl_0(\infty)$ does not contain the "unit matrix,"

$$I=\sum_{i\in\mathbb{N}} e_{ii} \notin gl_0(\infty) .$$

It is possible to extend $gl_0(\infty)$ to a larger LA $a_0(\infty)$, consisting of all infinite matrices (a_{ij}), $i,j\in\mathbb{N}$, which have finite number of nonzero diagonals. In this case $I\in a_0(\infty)$. If we try to apply the linear extension of Eqs. (3.7)–(3.9) in order to represent $a_0(\infty)$ in $V([m])$, then [see (3.9) and (3.12)]

$$I(m)_0=\left[\sum_{i=1}^{\infty} m_i\right](m)_0 . \tag{3.36}$$

Therefore, I is defined as an operator only in the modules of finite signatures, i.e., those $gl_0(\infty)$ modules $V([m])$, which signature $[m]$ has a finite number of nonzero elements. Clearly, $V([m])$ is a module of finite signature if and only if there exists $N\in\mathbb{N}$ such that $m_i=0$, $\forall i\geqslant N$. A more detailed analysis shows that any $gl_0(\infty)$ module of finite signature is also a highest weight irreducible unitary $a_0(\infty)$ module.

B. Highest weight modules of gl_∞

The propositions stated below are proved in a similar way as those from the previous section. Therefore, we skip all proofs.

Definition: Let

$$\begin{aligned}[M]&\equiv[...,M_p,...,M_{-1},M_0,M_1,...,M_q,...]\\ &\equiv\{M_i|M_i\in\mathbb{C},\quad i\in\mathbb{Z}\}\end{aligned} \tag{3.37}$$

be a sequence of complex numbers such that

$$M_i-M_j\in\mathbb{Z}_+,\quad \forall i<j\in\mathbb{Z} . \tag{3.38}$$

A pattern (M) consisting of all complex numbers

$$M_{i,2k+\theta-1},\quad \forall k\in\mathbb{N},\quad \theta=0,1,\quad i\in[1-\theta-k,k-1], \tag{3.39}$$

which satisfy the conditions (a) there exists $N([M])\in\mathbb{N}$ such that

$$M_{i,2k+\theta-1}=M_i,\quad \forall k>N[(M)],\quad \theta=0,1,\quad i\in[1-\theta-k,k-1] ; \tag{3.40}$$

$$\begin{aligned}&\text{(b)}\quad M_{i+\theta-1,2k+\theta}-M_{i,2k+\theta-1}\in\mathbb{Z}_+ ,\\ &M_{i,2k+\theta-1}-M_{i+\theta,2k+\theta}\in\mathbb{Z}_+,\\ &\forall k\in\mathbb{N},\quad \theta=0,1,\quad i\in[1-\theta-k,k-1] ,\end{aligned} \tag{3.41}$$

will be called a C pattern (corresponding to $[M]$).

The entries $M_{i,2k+\theta-1}$ will be referred to as coordinates of the C pattern. It is convenient to order them as indicated in the pattern below, writing as a first row the sequence $[M]$:

$$M=\begin{bmatrix} \cdots & ,M_{1-\theta-k} & ,\cdots & ,M_{-1} & ,M_0 & ,M_1 & ,\cdots & ,M_{k+\theta-1},\cdots\cdots \\ \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots \\ \cdots & ,M_{1-\theta-k,2k+\theta-1} & ,\cdots & ,M_{-1,2k+\theta-1} & ,M_{0,2k+\theta-1} & ,M_{1,2k+\theta-1} & ,\cdots & ,M_{k+\theta-1,2k+\theta-1} \\ & \cdots & \cdots & \cdots & \cdots & \cdots & \cdots & \\ & & & M_{-1,3} & ,M_{03} & ,M_{13} & & \\ & & & M_{-1,2} & ,M_{02} & & & \\ & & & & M_{01} & & & \end{bmatrix}, \tag{3.42}$$

where $k\in\mathbb{N}$, $\theta=0,1$.

Proposition 7: To each sequence (3.37) there corresponds an irreducible highest weight gl_∞ module $V([M])$ with a signature $[M]$. The basis $\Gamma([M])$ in $V([M])$ consists of all C patterns, corresponding to $[M]$. The transformation of the basis is determined from the relations [see (1.16) and (1.17)]

$$E_{(-k,k-\nu)_\mu}(M) = \sum_{j=\nu-k}^{k-1} \left| \frac{\Pi_{i=-k}^{k-\nu}[L_{i,2k+1-\nu} - L_{j,2k-\nu} + \mu(1-2\nu)]\,\Pi_{i=1-k}^{k-1-\nu}[L_{i,2k-1-\nu} - L_{j,2k-\nu} + \mu(1-2\nu)]}{\Pi_{i\neq j=\nu-k}^{k-1}[L_{i,2k-\nu} - L_{j,2k-\nu} + (-1)^\mu \nu]\,[L_{i,2k-\nu} - L_{j,2k-\nu} + (-1)^\mu(\nu-1)]} \right|^{1/2}$$

$$\times (M)_{(-1)^{\mu+\nu}[j,2k-\nu]}, \quad \forall \mu,\nu = 0,1, \quad \forall k \in \mathbb{N}, \tag{3.43}$$

$$E_{kk}(M) = \left[\sum_{i=-|k|}^{|k|+\theta(k)-1} M_{i,2|k|+\theta(k)} - \sum_{i=-|k|+1-\theta(k)}^{|k|-1} M_{i,2|k|+\theta(k)-1} \right](M), \tag{3.44}$$

where $k \in \mathbb{Z}$ and $M_{00} = M_{-10} = 0$. The gl_∞ highest weight vector $(M)_0$ is the one from $\Gamma([M])$ for which

$$M_{i,2k+\theta-1} = M_i, \quad \forall k \in \mathbb{N}, \quad \theta = 0,1,$$

$$i \in [1-\theta-k, k-1]. \tag{3.45}$$

Proposition 8: Define an antilinear anti-involutive mapping ω in gl_∞ with the Eqs. (3.32) and (3.33) for every $a,b \in \mathrm{gl}_\infty$ and

$$\omega(E_{ij}) = E_{ji}, \quad \forall i,j \in \mathbb{Z}. \tag{3.46}$$

Introduce a metric (,) in $V([M])$, postulating that the C basis $\Gamma([M])$ be orthonormed. Then $V([M])$ carries a unitary (irreducible highest weight) representation of gl_∞ if and only if its signature $[M]$ consists of real numbers. The real form of gl_∞, which is a linear span of all $i(E_{pq} + E_{qp})$, $(E_{pq} - E_{qp})$, $p,q \in \mathbb{Z}$, is integrable to a unitary representation of the corresponding group U_∞ only if the coordinates of $[M]$ are integers.

From a point of view of physical applications it is more interesting to construct the unitary irreducible highest weight representations of the completion and central extension a_∞ of gl_∞. In a forthcoming paper[12] we shall study all these representations, using essentially the results of the present investigation.

ACKNOWLEDGMENTS

I am thankful to Professor H. D. Doebner for the kind hospitality in the Arnold Sommerfeld Institute for Mathematical Physics, where most of the present results have been obtained. The paper was written during my visit to the Department of Mathematics at the University of Queensland. I am thankful to Dr. A. J. Bracken, Dr. M. D. Gould, and Dr. R. B. Zhang for fruitful discussions.

[1] V. G. Kac, "Infinite dimensional Lie algebras," in *Progress in Mathematics*, Vol. 44 (Birkhauser, Boston, 1983) [second edition (Cambridge U.P., Cambridge, 1985)].
[2] V. G. Kac and A. K. Raina, "Bombay lectures on highest weight representations of infinite-dimensional Lie algebras," in *Advanced Series in Mathematics*, Vol. 2 (World Scientific, Singapore, 1987).
[3] E. Date, M. Jimbo, M. Kashiwara, and T. Miwa, J. Phys. Soc. Jpn. **50**, 3806 (1981).
[4] E. Date, M. Kashiwara, and T. Miwa, Proc. Jpn. Acad. **57**, 387 (1981).
[5] E. Date, M. Jimbo, M. Kashiwara, and T. Miwa, Physica D **4**, 343 (1982).
[6] M. Jimbo and T. Miwa, Publ. Res. Inst. Math. Sci. **19**, 943 (1983).
[7] V. G. Kac and D. H. Peterson, Proc. Nat. Acad. Sci. USA **78**, 3308 (1981).
[8] P. Goddard and D. Olive, Intl. J. Mod. Phys. A **1**, 303 (1986).
[9] I. M. Gel'fand and M. L. Zetlin, Dokl. Akad. Nauk SSSR **71**, 825 (1950) (in Russian).
[10] G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[11] A. A. Kirillov, Dokl. Akad. Nauk SSSR **212**, 288 (1973) (in Russian).
[12] T. D. Palev, "Highest weight irreducible unitarizable representations of Lie algebras of infinite matrices. II. The algebra a_∞ to be published in J. Math. Phys.

On induced scalar products and unitarization

Bruno Gruber
Department of Physics, Southern Illinois University, Carbondale, Illinois 62901-4401 and Arnold Sommerfeld Institute for Mathematical Physics, Technical University Clausthal, D-3392 Clausthal, Germany

Romuald Lenczewski[a)]
Department of Mathematics, Southern Illinois University, Carbondale, Illinois 62901

Miguel Lorente
Departamento de Fisica, Universidad de Oviedo, 33007 Oviedo, Spain

(Received 20 March 1987; accepted for publication 18 October 1989)

Infinitesimal unitarity of representations of the simple real Lie algebras su(2), su(1,1), and so(4) is discussed with respect to scalar products induced by sesquilinear forms on their universal enveloping algebras. Sesquilinear forms are explicitly calculated. The Verma modules, their irreducible quotients, their irreducible submodules, and their infinitesimal unitarizability are discussed.

I. INTRODUCTION

This article discusses the Verma modules for su(2), su(1,1), and so(4), and the unitarizability of irreducible Verma modules, their irreducible quotient modules, and their irreducible submodules. Infinitesimal unitarity is discussed with respect to a scalar product that is induced by a sesquilinear form defined on the enveloping algebra of a given Lie algebra.[1] It is shown that the scalar product leads to the unitarized representations of su(2), su(1,1), and so(4) that are familiar from physical applications.

A Verma module d_Λ of highest weight Λ, $\Lambda \in C$, may be reducible, but not completely reducible. This happens when there exists a submodule $d_{\tilde{\Lambda}}$, invariant under the action of the Lie algebra with a complement that is not invariant. In other words, it is an indecomposable module.

An invariant submodule $d_{\tilde{\Lambda}}$ exists in d_Λ if in addition to the identity (coset) $\mathbb{1}$, there exists another extremal vector y in d_Λ. For a given weight Λ the existence of the $\tilde{\Lambda}$, as well as their explicit dependence on Λ, has been known for a long time. In Ref. 2 an algorithm is given to construct extremal vectors y that generate the submodules mentioned above for the case of simple Lie algebras.

Thus, one can consider the submodules $d_{\tilde{\Lambda}}$ or the quotient modules $d_\Lambda / d_{\tilde{\Lambda}}$. Both types of modules may turn out to be irreducible or indecomposable. An investigation of these modules for the simplest possible case of a complex simple Lie algebra sl(2) [complex su(2)] and its real forms su(2) [isomorphic to so(3)] and su(1,1) was carried out in Refs. 3 and 4.

Given a semisimple Lie algebra g, we consider first the enveloping algebra of g, $U(g)$. Then the quotient spaces of $U(g)$ modulo certain left ideals are studied. Namely, the Verma modules d_Λ of highest weight Λ, which are obtained from $U(g)$ by mapping the left ideal generated by $H_\alpha - \Lambda_\alpha$, X_α (α is any positive root) onto zero. Thus, the modules obtained give diagonal values of H_α (the operators which provide the physical quantum numbers) and they possess a state of highest weight. An analogous definition can be given for modules of lowest weight Λ.

The infinitesimal unitarity of such (irreducible) modules can be discussed once a sesquilinear form is introduced on $U(g)$. Namely, it is possible to define a sesquilinear form S that induces a scalar product on either the irreducible quotient modules for the compact real form of g or on certain irreducible Verma modules and on irreducible submodules for a noncompact real form of g.

Such a sesquilinear form was introduced by Harish-Chandra,[1] Gel'fand and Kirillov,[5] and Shapovalov.[6] With the help of this sesquilinear form Jakobsen[7,8] was able to construct infinitesimally unitary irreducible representations for the noncompact real forms corresponding to Hermitian symmetric spaces of simple Lie algebras. His method, based on the Bernstein–Gel'fand–Gel'fand theorem, gives an algorithm to calculate all the highest weights which define infinitesimally unitary representations.

In Sec. II we introduce the sesquilinear form on the universal enveloping algebra $U(g)$ of a semisimple complex Lie algebra.

In Sec. III we treat explicitly the case of sl(2) and its real form su(2). The angular momentum algebra su(2) is of fundamental importance in physics and thus it is of exemplificatory nature. We first define the sesquilinear form S_0 on the universal enveloping algebra of sl(2), then construct the Verma module d_Λ. It is shown how the sesquilinear form induced on d_Λ defines a scalar product on the finite-dimensional irreducible quotient modules, $d_l/\tilde{\pi}_l$, $l = (1/2)k$, $k \in \mathbb{N}$, thus unitarizing them for su(2).

In Sec. IV we follow with a similar discussion of su(1,1). In this case, since su(1,1) is noncompact, the sesquilinear form defined on the Verma modules induces a scalar product on the Verma modules d_l with $l < 0$, as well as on the infinite-dimensional irreducible submodules $\tilde{\pi}_l$ of the Verma modules d_l, $l = \frac{1}{2}k$, $k \in \mathbb{N}$, thus unitarizing the su(1,1) representations on these spaces. As it is for the case of su(2),

[a)] Permanent address: Department of Mathematics, Technical University of Wroclaw, 50-370 Wroclaw, Poland.

 0022-2488/90/030587-07$03.00

these unitarized representations of su(1,1) take on the form which is familiar from physical applications.

In Sec. V we calculate the sesquilinear form for Verma modules of the complexification of so(4) derived in Ref. 9. The explicit form of the sesquilinear form enables us to conclude that the standard form of irreducible Verma modules of so(4) is obtained from the modules derived in Ref. 8 by renormalizing the basis elements by their norms. The same remark applies to Verma modules of su(2) and su(1,1) derived in Ref. 3.

II. PRELIMINARIES

Let R, C denote the fields of real and complex numbers, respectively. Let g_1 be a semisimple Lie algebra over R and g its complexification. Let $B(X,Y) = \mathrm{tr}(\mathrm{ad}X\,\mathrm{ad}Y)$, $X, Y\in g$, be the Killing form.

A real form g_0 of g is called compact if $B(X,X) < 0$ for each $X\in g_0 (X \neq 0)$. There exists a compact real form g_0 and an automorphism θ of order 2 of g such that

$$\theta g_0 \subset g_0, \quad \theta g_1 \subset g_1$$

and

$$g_1 = f_1 + p_1, \quad g_0 = f_1 + ip_1,$$

where $i = \sqrt{-1}$, f_1 is the set of all $X\in g_1$ such that $\theta X = X$, and p_1 is the set of all $Y\in g_1$ such that $\theta Y = -Y$.

Let f and p be the subspaces of g spanned by f_1 and p_1, respectively, over C. It holds that

$$[f,f] \subset f, \quad [f,p] \subset p, \quad [p,p] \subset f.$$

Let h_1 be a Cartan subalgebra of g_1, with g_1 a Hermitian symmetric space, and h the complexification of h_1. Then h is a Cartan subalgebra of g and it holds that

$$[h,f] \subset f, \quad [h,p] \subset p.$$

For given g, h, let Δ be the root system of g and Δ^+ the system of positive roots. Then for each root $\alpha\in\Delta$ one can choose $X_\alpha \in g$, $H_\alpha \in h$ such that

$$[H_\alpha, X_\alpha] = (\alpha,\alpha)X_\alpha, \quad \text{if } \alpha\in\Delta^+,$$

$$[H_\alpha, X_{-\alpha}] = -(\alpha,\alpha)X_{-\alpha}, \quad \text{if } \alpha\in\Delta^+,$$

$$[X_\alpha, X_{-\alpha}] = H_\alpha, \quad \text{if } \alpha\in\Delta^+,$$

$$[X_\alpha, X_\beta] = \begin{cases} N_{\alpha+\beta}X_{\alpha+\beta}, & \text{if } \alpha+\beta\in\Delta, \\ 0, & \text{if } \alpha+\beta\notin\Delta,\ \alpha+\beta\neq 0, \end{cases}$$

where $(\alpha,\alpha) = \alpha(H_\alpha)$.

It follows that for given g_1, g, h, if α is any root of g with respect to h, then X_α is either in f or in p. We say that α is compact iff $X_\alpha \in f$ and noncompact iff $X_\alpha \in p$. The set of compact roots will be denoted by Δ_c, the set of noncompact roots by Δ_n. We have the direct decompositions

$$f = h + \sum_{\alpha\in\Delta_c} g^\alpha, \quad p = \sum_{\beta\in\Delta_n} g^\beta.$$

Let n_+ be the subalgebra of g generated by the positive root vectors X_α $(\alpha\in\Delta^+)$ and n_- the subalgebra of g generated by the negative root vectors $X_{-\alpha}$ $(\alpha\in\Delta^+)$.

Let $U(g)$ be the (universal) enveloping algebra. According to the Poincare–Birkhoff–Witt theorem, the elements

$$X((q_i),(m_i),(p_i)) = X^{q_1}_{-\alpha_1}\cdots X^{q_n}_{-\alpha_n} H_1^{m_1}\cdots H_l^{m_l} X^{p_1}_{\alpha_1}\cdots X^{p_n}_{\alpha_n},$$

where H_i's give a basis of h, form a basis of $U(g)$, with $\alpha_1, \alpha_2, \cdots, \alpha_n$ a fixed ordering of the set of positive roots.

One can consider $U(g)$ as a g-module corresponding to the adjoint representation. For each $H\in h$ we have

$$[H, X((q_i),(m_i),(p_i))] = ((p_1 - q_1)\alpha_1(H) + \cdots + (p_n - q_n)\alpha_n(H)) \times X((q_i),(m_i),(p_i)).$$

Thus, $\lambda = \Sigma(p_i - q_i)\alpha_i \in h^*$ is a weight of this representation and we can decompose $U(g)$ as a direct sum of weight subspaces,

$$U(g) = \oplus U(g)_\lambda.$$

In particular, $U(g)_0$ is the commutant of h in $U(g)$. It is a subalgebra of $U(g)$ consisting of elements of $U(g)$ with zero weight. Let π: $U(g) \to U(g)_0$ be the natural projection, i.e., the linear map

$$\pi(X((q_i),(m_i),(p_i))) = \begin{cases} X((q_i),(m_i),(p_i)), & \text{if } p_i = q_i, \quad \forall i, \\ 0, & \text{otherwise.} \end{cases}$$

For $U(g)_0$ holds $U(g)_0 = U(h) \oplus L$ where L is the two-sided ideal of $U(g)$ defined as

$$L = n_- U(g) \cap U(g)_0 = U(g) n_+ \cap U(g)_0,$$

i.e., the elements of weight zero of $U(g)$ which are not in $U(h)$. Notice that $U(h) \subset U(g)_0 \subset U(g)$. The natural projection ϕ: $U(g)_0 \to U(h)$ has L as its kernel and is called the Harish–Chandra homomorphism. In the sequel we will use the composition of the two projections introduced above, namely $\gamma = \phi\cdot\pi$.

Now, $U(h)$ is a commutative algebra and every $\Lambda\in h^*$ can be extended uniquely to a homomorphism from $U(h)$ into C. This homomorphism will be also denoted by Λ. Again, we will use the composition $\Lambda\cdot\gamma = \Lambda\cdot\phi\cdot\pi = \xi_\Lambda$.

Given a real semisimple Lie algebra g_1 and its complexification g, let σ denote the conjugation of g with respect to g_1:

$$\sigma: g \to g,$$

$$\sigma(X + iY) = X - iY \quad \text{for } X, Y\in g_1.$$

Then $-\sigma$ can be extended to an antilinear antiautomorphism η of $U(g)$ as follows:

$$\eta(1) = 1,$$

$$\eta(X) = -\sigma(X),$$

$$\eta(XY\cdots Z) = \eta(Z)\cdots\eta(Y)\eta(X).$$

We can define a sesquilinear form S on $U(g)\times U(g)$ for any $\Lambda\in h^*$ as follows:

$$S:\ U(g)\times U(g) \to C,$$

$$S(X,Y) = \xi_\Lambda(\eta(X)Y), \quad X, Y\in U(g).$$

In general, i.e., on $U(g)$ and for arbitrary Λ, S is degenerate and indefinite for a given g_1. However, for some Λ it may induce a scalar product on irreducible quotients of Verma modules of highest weight Λ and unitarize the finite-dimensional irreducible representations if g_1 is compact (see

Ref. 1). On the other hand, if g_1 is noncompact, then for some Λ it may induce a scalar product on irreducible Verma modules or on irreducible submodules of Verma modules of highest weight Λ and unitarize infinite-dimensional irreducible representations of g_1.

Let d_Λ denote a Verma module, $\tilde{\pi}_\Lambda$ an invariant submodule, and $\pi_\Lambda \sim d_\Lambda/\tilde{\pi}_\Lambda$ a quotient module. If $\rho = d_\Lambda, \tilde{\pi}_\Lambda, \pi_\Lambda$ is irreducible, then we say that ρ is infinitesimally unitary if there exists a scalar product (,) on the carrier space V of ρ such that

$$(u,\rho(X)w) = -(\rho(X)u,w),$$

for all $X\in g_1$ and $u,w\in V$. The above condition is called g_1 invariance.

If V is the irreducible quotient of a Verma module, then π_Λ is infinitesimally unitary with respect to the scalar product induced by S iff $\xi_\Lambda(\eta(z)z)$ is real and non-negative for every $z\in V$ (see Ref. 1)

If g_1 is simple, another useful result was given by Harish-Chandra in Ref. 1. If g has no totally positive roots, then sufficient (and necessary) conditions for the unitarizability of the irreducible quotient representation π_Λ are that $\Lambda(H_\alpha)$ is a non-negative integer for every compact positive root α and $\Lambda(H_\beta) = 0$ for every noncompact positive root β. On the other hand, if every noncompact positive root is totally positive, then the necessary conditions are that $\Lambda(H_\alpha)$ is a non-negative integer for every compact positive root α and $\Lambda(H_\beta)$ is real and $\leqslant 0$ for every noncompact positive root β. To obtain the sufficient conditions one has to strengthen the requirement on the noncompact positive roots, namely that $\Lambda(H_\beta) + \rho(H_\beta)$ is real and $\leqslant 0$ for every noncompact positive root β. Here, ρ denotes half of the sum of all positive roots of g (see Ref. 1).

III. su(2)

As the simplest example we consider the real simple Lie algebra su(2). The calculation of the sesquilinear form in this case is very straightforward and is basically the calculation given in Ref. 1. However, it is worth being presented here since it gives a nice illustration of the methods used.

The real algebra su(2) can be defined as the linear span (over R) of X_{12}, X_{23}, X_{13} with the following Lie brackets:

$$[X_{12},X_{23}] = X_{13}, \quad [X_{23},X_{13}] = X_{12}, \quad [X_{13},X_{12}] = X_{23}.$$

Thus, sl(2) [complex su(2)] can be obtained by introducing

$$l_+ = (1/\sqrt{2})(X_{13} - iX_{23}), \; l_- = (1/\sqrt{2})(-X_{13} - iX_{23}),$$
$$l_3 = iX_{12}$$

with brackets:

$$[l_+,l_-] = l_3, \quad [l_3,l_-] = -l_-, \quad [l_3,l_+] = l_+.$$

Notice that we opted for the basis that is employed in physics and used in Ref. 3 at the expense of mathematical uniformity, but a change can be easily introduced. Now, the following relation can be derived on $U(g)$ by induction:

$$[l_+,l_-^n] = nl_-^{n-1}l_3 - \tfrac{1}{2}n(n-1)l_-^{n-1},$$

which in turn leads to the following relation:

$$l_+^n l_-^n = l_+^{n-1}l_-^{n-1}(l_-l_+ + nl_3 - \tfrac{1}{2}n(n-1)\mathbf{1}).$$

The conjugation σ_0 of sl(2) with respect to su(2) gives the following:

$$\sigma_0(l_+) = -l_-, \quad \sigma_0(l_-) = -l_+,$$
$$\sigma_0(l_3) = -l_3, \quad \eta_0 = -\sigma_0.$$

A basis for the enveloping algebra is given by monomials $l_-^n l_3^s l_+^t$, where n,s,t are non-negative integers. For a given linear function Λ on h the Verma module d_Λ of highest weight Λ is the quotient space $U(g)/I_\Lambda$ where the left ideal I_Λ is generated by $l_3 - \Lambda(l_3)$ and l_+. In other words,

$$I_\Lambda = U(g)(l_3 - \Lambda(l_3)) + U(g)l_+.$$

Thus, a basis for d_Λ is given by cosets l_-^n, where n is a non-negative integer.

The sesquilinear form induced on such a Verma module by the sesquilinear form defined in Sec. II is given by

$$S_0(\mathbf{1},\mathbf{1}) = 1,$$

and

$$\begin{aligned}
S_0(l_-^k,l_-^n) &= \xi_\Lambda(\eta_0(l_-^k)l_-^n) = \xi_\Lambda(l_+^k l_-^n) = \delta_{kn}\Lambda\cdot\phi(l_+^n l_-^n)\\
&= \delta_{kn}\Lambda\cdot\phi(l_+^{n-1}(l_-^n l_+ + nl_-^{n-1}(l_3 - \tfrac{1}{2}(n-1)\mathbf{1})))\\
&= \delta_{kn}\Lambda\cdot\phi(nl_+^{n-1}l_-^{n-1}(l_3 - \tfrac{1}{2}(n-1)\mathbf{1}))\\
&= \delta_{kn}\Lambda\Bigl(\prod_{t=0}^{n-1}\Bigl((n-t)\Bigl(l_3 - \frac{1}{2}(n-1-t)\mathbf{1}\Bigr)\Bigr)\Bigr)\\
&= \delta_{kn}n!\prod_{t=1}^{n}\Bigl(l - \frac{1}{2}(t-1)\Bigr),
\end{aligned}$$

where $l = \Lambda(l_3)$.

We used the fact that $l_+^{n-1}l_-^n l_+ \in L$ (see Sec. I). It can be seen that S_0 is a complex valued sesquilinear form. However, when l is real then it becomes real valued. Furthermore, when l is real but not a non-negative integer or half-integer then it is indefinite. When l is a non-negative integer or half-integer then S_0 becomes degenerate and vanishes on the submodule $\tilde{\pi}_l$ generated by the extremal vector $y = l_-^{2l+1}$ and spanned by $U(n_-)y$. On the irreducible quotient module $\pi_l = d_l/\tilde{\pi}_l$ spanned by $\{l_-^k \mid 0 < k \leqslant 2l\}$ S_0 induces an su(2)-invariant scalar product,

$$S_0(l_-^k, l_-^n) = \delta_{kn} 2^{-n} n!(2l)!/(2l-n)!$$
$$= \delta_{kn} \| l_-^n \|^2.$$

Let us now look at the explicit form of the Verma module d_l (ref. 3):

$$l_3 X(n) = (l-n)X(n),$$
$$l_+ X(n) = n(l - \tfrac{1}{2}(n-1))X(n-1),$$
$$l_- X(n) = X(n+1),$$

where $X(n) = l_-^n$.

For the irreducible quotient module π_l, a new basis, orthonormal with respect to S_0, is given by

$$|l,m\rangle = \|X(l-m)\|^{-1} X(l-m),$$

where $n = l - m$, $m = -l,...,l$, and

$$\|X(l-m)\| = \left(\frac{2^{m-l}(l-m)!(2l)!}{(l+m)!}\right)^{1/2}.$$

Thus, on π_l we obtain

$$l_3 |l,m\rangle = m|l,m\rangle,$$
$$l_+ |l,m\rangle = (1/\sqrt{2})(l(l+1) - m(m+1))^{1/2} |l,m+1\rangle,$$
$$l_- |l,m\rangle = (1/\sqrt{2})(l(l+1) - m(m-1))^{1/2} |l,m-1\rangle,$$

i.e., the standard form for the matrix elements for the angular momentum algebra as used in physics.

IV. su(1,1)

A similar approach will be presented for su(1,1). The real algebra su(1,1) can be defined as the real span of $\widehat{X}_{23}, \widehat{X}_{13}, \widehat{X}_{12}$, with the following Lie brackets:

$$[\widehat{X}_{23}, \widehat{X}_{13}] = -\widehat{X}_{12}, \quad [\widehat{X}_{13}, \widehat{X}_{12}] = \widehat{X}_{23},$$
$$[\widehat{X}_{12}, \widehat{X}_{23}] = \widehat{X}_{13}.$$

There exists an automorphism ψ of $g = sl(2)$ given by

$$\psi\colon \psi(\widehat{X}_{23}) = i\widehat{X}_{23}, \quad \psi(\widehat{X}_{13}) = i\widehat{X}_{13}, \quad \psi(\widehat{X}_{12}) = \widehat{X}_{12},$$

and

$$\hat{l}_+ = il_+, \quad \hat{l}_- = il_-, \quad \hat{l}_3 = l_3,$$
$$[\hat{l}_3, \hat{l}_\pm] = \pm \hat{l}_\pm, \quad [\hat{l}_+, \hat{l}_-] = -\hat{l}_3.$$

The conjugation σ_1 of g with respect to $g_1 = \text{su}(1,1)$ gives for the elements $\hat{l}_3, \hat{l}_+, \hat{l}_-$,

$$\sigma_1(\hat{l}_+) = -\hat{l}_-, \quad \sigma_1(\hat{l}_-) = -\hat{l}_+, \quad \sigma_1(\hat{l}_3) = -\hat{l}_3.$$

Thus $\eta_1(\hat{l}_+) = \hat{l}_-$, $\eta_1(\hat{l}_-) = \hat{l}_+$, $\eta_1(\hat{l}_3) = \hat{l}_3$, and the sesquilinear form $S_1(x,y) = \xi_\Lambda(\eta_1(x),y)$ in the su(1,1) basis can be calculated as

$$S_1(\mathbb{1},\mathbb{1}) = 1, \quad n = 0,$$
$$S_1(\hat{l}_-^k, \hat{l}_-^n) = \delta_{kn}(-1)^n n! \prod_{s=1}^{n} \left(l - \frac{1}{2}(s-1)\right),$$
$$k + n > 0.$$

Again, S_1 is a complex-valued sesquilinear form that becomes real if l is real.

In the following we discuss three cases for real l.

Case A: $l > 0$, $l \neq k/2$, $k \in \mathbb{N}^+$. For these values of l the S_1 is nondegenerate, but indefinite.

Case B: $l < 0$: S_1 is a scalar product. It holds

$$S_1(\mathbb{1},\mathbb{1}) = 1, \quad n = 0,$$
$$S_1(\hat{l}_-^k, \hat{l}_-^n) = \delta_{kn} n! \prod_{s=1}^{n} \left(-l + \frac{1}{2}(s-1)\right)$$
$$= \delta_{kn} \|\hat{l}_-^n\|^2, \quad k + n > 0.$$

The su(1,1) invariance property follows from the definition of η_1. The basis for the infinitesimally unitary Verma module d_l is given by

$$|l,l\rangle = \mathbb{1}, \quad m = l,$$
$$|l,m\rangle = \|\hat{l}_-^{l-m}\|^{-1} \hat{l}_-^{l-m}$$
$$= \left\{(l-m)! \prod_{s=1}^{l-m} \left(-l + \frac{1}{2}(s-1)\right)\right\}^{-1/2} \hat{l}_-^{l-m},$$
$$l < 0; \quad m = l-1, l-2, l-3, \ldots .$$

Substitution of the new basis states $|l,m\rangle$ into the su(1,1) relations (Ref. 3)

$$\hat{l}_3 \widehat{X}(n) = (l-n)\widehat{X}(n),$$
$$\hat{l}_+ \widehat{X}(n) = -n(l - \tfrac{1}{2}(n-1))\widehat{X}(n-1),$$
$$\hat{l}_- \widehat{X}(n) = \widehat{X}(n+1),$$
$$\widehat{X}(n) = \hat{l}_-^n, \quad n = l - m,$$

one obtains the familiar form for the su(1,1) representations,

$$\hat{l}_3 |l,m\rangle = m|l,m\rangle,$$
$$\hat{l}_+ |l,m\rangle = (1/\sqrt{2})\sqrt{-l(l+1) + m(m+1)}\,|l,m+1\rangle,$$
$$\hat{l}_- |l,m\rangle = (1/\sqrt{2})\sqrt{-l(l+1) + m(m-1)}\,|l,m-1\rangle,$$
$$l < 0, \quad m = l, l-1, l-2, \ldots .$$

Case C: $l > 0$, $l = \frac{1}{2}k$, $k \in \mathbb{N}$: For these values of l the sesquilinear form S_1 is degenerate and vanishes on the irreducible submodule $\tilde{\pi}_l$ of d_l generated by the extremal vector $y = \hat{l}_-^{2l+1}$ and spanned by $U(n_-)y$. However, S_1 induces a scalar product on the submodule $\tilde{\pi}_l$, as will be shown below. Consider a Verma module $d_{l+i\epsilon}$, l a non-negative integer or halfinteger, and $\epsilon > 0$. The sesquilinear form S_1 can be factored,

$$S_1(\hat{l}_-^k y, \hat{l}_-^n y) = S_1(y,y) S_1^*(\hat{l}_-^k y, \hat{l}_-^n y),$$

with

$$S_1(y,y) = (-1)^{2l+1}(2l+1)!$$
$$\times \prod_{s=1}^{2l+1} (l + i\epsilon - \tfrac{1}{2}(s-1)) \neq 0.$$

It follows that

$$S_1^*(y,y) = 1,$$

$$S_1^*(\hat{l}_-^k y, \hat{l}_-^n y) = \delta_{kn}(-1)^n \frac{(n+2l+1)!}{(2l+1)!} \times \prod_{s=2l+2}^{n+2l+1} \left(l + i\epsilon - \frac{1}{2}(s-1) \right) = \delta_{kn} \frac{(n+2l+1)!}{2^n (2l+1)!} \prod_{s=1}^{n} (s - 2i\epsilon),$$

$$k + n > 0.$$

Thus, for $\epsilon \to 0$, the sesquilinear form S_1^* becomes a scalar product on $\tilde{\pi}_l$,

$$S_1^*(y,y) = 1,$$

$$S_1^*(\hat{l}_-^k y, \hat{l}_-^n y) = \delta_{kn} \frac{(n+2l+1)!}{2^n(2l+1)!} n! = \|\hat{l}_-^n y\|^2, \quad k+n > 0.$$

If new basis elements are defined on $\tilde{\pi}_l$,

$$|-l-1,-l-1\rangle = y,$$

$$|-l-1,m\rangle = \left\{ \frac{2^{-l-1-m}(2l+1)!}{(l-m)!(-l-1-m)!} \right\}^{1/2} \hat{l}_-^{-l-1-m} y,$$

$$n = -l-1-m \geqslant 0,$$

$$m = -l-1, -l-2, -l-3, \ldots,$$

then the infinitesimally unitary representation of su(1,1) on $\tilde{\pi}_l$ takes on the form

$$\hat{l}_3 |-l-1,m\rangle = m |-l-1,m\rangle,$$

$$\hat{l}_+ |-l-1,m\rangle = (1/\sqrt{2})\sqrt{-l(l+1)+m(m+1)} \times |-l-1,m+1\rangle,$$

$$\hat{l}_- |-l-1,m\rangle = (1/\sqrt{2})\sqrt{-l(l+1)+m(m-1)} \times |-l-1,m-1\rangle,$$

$$l = k/2, \quad k \in \mathbb{N}, \quad m = -l-1, -l-2, -l-3, \ldots .$$

These equations are, however, identical to the equations for the case $l < 0$ as discussed above. Thus, from an algebraic point of view, these representations on the invariant subspace $\tilde{\pi}_l$ of the Verma module d_l, $l = k/2$, $k \in \mathbb{N}$, are isomorphic to certain representations on the Verma modules d_l, $l < 0$. It is also seen that these infinitesimally unitary representations of su(1,1) on the invariant subspace $\tilde{\pi}_l$ of the Verma module d_l, $l = \frac{1}{2}k$, $k \in \mathbb{N}$, are in a one-to-one correspondence with the infinitesimally unitary representations of su(2) on the quotient space $d_l / \tilde{\pi}_l$, $l = \frac{1}{2}k$, $k \in \mathbb{N}$. The scalar product on $\tilde{\pi}_l$ is given by S_1, while on $d_l/\tilde{\pi}_l$, it is given by S_0. Moreover,

$$S_1(\hat{l}_-^k, \hat{l}_-^n) = (-1)^n S_0(l_-^k, l_-^n).$$

V. so(4)

In order to conform to the notation of Ref. 9 we will use in this section the complexified Lorentz algebra basis. Let $X_{\mu\nu}$ denote the generators of (real) so(3,1) with the Lie products

$$[X_{\mu\nu}, X_{\alpha\beta}] = g_{\mu\alpha} X_{\nu\beta} - g_{\nu\alpha} X_{\mu\beta} + g_{\mu\beta} X_{\alpha\nu} - g_{\nu\beta} X_{\alpha\mu},$$

$$g_{\mu\nu} = g_{\mu\mu}\delta_{\mu\nu}, \quad g_{00} = 1, \quad g_{ii} = -1, \quad i = 1,2,3.$$

Here X_{01}, X_{02}, X_{03} denote the boost generators and X_{12}, X_{13}, X_{23} are the generators of the rotation subalgebra so(3). The complexification of so(3,1) is spanned by

$$h_+ = X_{13} - iX_{23}, \quad h_- = -X_{13} - iX_{23}, \quad h_3 = iX_{12},$$

$$p_+ = -X_{02} - iX_{01}, \quad p_- = X_{02} - iX_{01}, \quad p_3 = iX_{03},$$

with the following nonvanishing Lie products, in the form used in physical applications:

$$[h_3, h_+] = h_+, \quad [h_3, h_-] = -h_-,$$

$$[h_3, p_+] = p_+, \quad [h_3, p_-] = -p_-,$$

$$[h_+, h_-] = 2h_3, \quad [p_+, p_-] = -2h_3,$$

$$[h_+, p_-] = [p_+, h_-] = 2p_3,$$

$$[p_3, p_+] = -h_+, \quad [p_3, p_-] = h_-,$$

$$[p_3, h_+] = p_+, \quad [p_3, h_-] = -p_-.$$

If we replace in these relations the boost operators $X_{0\mu}$ by the new elements $X'_{0\mu} = iX_{0\mu}$, then the elements $X'_{0\mu}$, X_{ik}, $i < k$, $i,k = 1,2,3$, form a basis for the real compact Lie algebra so(4). In the so(4) basis

$$g_{\mu\nu} = -\delta_{\mu\nu}$$

and

$$[p'_3, p'_+] = h_+, \quad [p'_3, p'_-] = -h_-, \quad [p'_+, p'_-] = 2h_3.$$

In what follows we will drop the prime in these relations. This should not cause confusion, since in what follows we will use only the so(4) basis.

Let us consider an angular momentum basis for the so(4) Verma modules of highest weight Λ in analogy to the so(3,1) Verma modules as discussed in Ref. 9: $y_N^n = h_-^n y_N$, $n,N \in \mathbb{N}$, where the y_N are h_+-extremal vectors defined by

$$y_N = \sum_{k=0}^{N} c_k p_-^{N-k} h_-^k, \quad h_+ y_N = 0.$$

Keeping in mind that the elements p_+, p_-, p_3 are actually primed elements in the so(4) basis ($p \to ip$) one obtains from Eq. (4.18) of Ref. 9 for the action of so(4) on d_Λ,

$$h_3 \, y_N^n = (\Lambda_1 - N - n) y_N^n,$$

$$h_+ y_N^n = n(2\Lambda_1 - 2N + 1 - n) y_N^{n-1},$$

$$h_- y_N^n = y_N^{n+1},$$

$$p_3 \, y_N^n = -\alpha_N (2\Lambda_1 - 2N + 1 - n) y_{N-1}^{n+1} + \beta_N (\Lambda_1 - N - n) y_N^n - n y_{N+1}^{n-1},$$

$$p_+ y_N^n = \alpha_N (2\Lambda_1 - 2N + 1 - n) \times (2\Lambda_1 - 2N + 2 - n) y_{N-1}^n + \beta_N n (2\Lambda_1 - 2N + 1 - n) y_N^{n-1} - n(n-1) y_{N+1}^{n-2},$$

$$p_- y_N^n = -\alpha_N y_{N-1}^{n+2} + \beta_N y_N^{n+1} + y_{N+1}^n,$$

where

$$\alpha_N = (1/D)(\Lambda_2^2 + (\Lambda_1 + 1 - N)^2)N(2\Lambda_1 + 2 - N),$$
$$\beta_N = (-i/D^*)\Lambda_2(\Lambda_1 + 1),$$

and

$$D = (\Lambda_1 + 1 - N)^2(2\Lambda_1 - 2N + 3)(2\Lambda_1 - 2N + 1),$$
$$D^* = (\Lambda_1 - N)(\Lambda_1 + 1 - N),$$
$$\Lambda_1 = \Lambda(h_3), \quad \Lambda_2 = \Lambda(p_3) \in C.$$

Now we consider a conjugation σ_0 of the complex algebra D_2 with respect to the compact real algebra so(4). We obtain

$$\sigma_0(h_+) = -h_-, \quad \sigma_0(h_-) = -h_+, \quad \sigma_0(h_3) = -h_3,$$
$$\sigma_0(p_+) = -p_-, \quad \sigma_0(p_-) = -p_+, \quad \sigma_0(p_3) = -p_3,$$

since, for example,

$$\sigma_0(p_+) = \sigma_0(i(-X_{02} - iX_{01})) = \sigma_0(-(iX_{02}) - i(iX_{01}))$$
$$= -(iX_{02}) + i(iX_{01}) = -p_-.$$

Later we will use $\eta_0 = -\sigma_0$.

We now calculate the sesquilinear form S_0 corresponding to so(4) on the Verma module d_Λ given above in an angular momentum basis. The calculation will be presented in a series of steps. Thus,

$$S_0(y_0,y_0) = S_0(1,1) = 1,$$
$$S_0(y_N,y_N) = \xi_\Lambda\left(\eta_0\left(\sum_{k=0}^{N} c_k h_-^k p_-^{N-k}\right)y_N\right)$$
$$= \xi_\Lambda\left(\sum_{k=0}^{N} c_k^* p_+^{N-k} h_+^k y_N\right)$$
$$= \xi_\Lambda(p_+^N y_N), \quad \text{since } h_+y_N = 0,$$
$$\text{and with } c_0^* = 1.$$

Now

$$p_+y_N = \alpha_N(2\Lambda_1 - 2N + 1)(2\Lambda_1 - 2N + 2)y_{N-1}.$$

Hence

$$S_0(y_N,y_N) = \prod_{T=1}^{N} \alpha_T(2\Lambda_1 + 1 - 2T)(2\Lambda_1 + 2 - 2T).$$

Moreover, for $M \geqslant N$,

$$S_0(y_M,y_N) = \delta_{MN}S_0(y_N,y_N), \quad \text{since } \alpha_0 = 0,$$
$$S_0(y_N,y_M) = 0, \quad \text{since } \xi_\Lambda(y_{N-M}) = 0,$$

due to the properties of the projection π (see Sec. II). Next,

$$S_0(y_N^n,y_N^n) = S_0(h_-^n y_N, h_-^n y_N) = \xi_\Lambda(\eta_0(h_-^n y_N), h_-^n y_N)$$
$$= \xi_\Lambda(\eta_0(y_N)h_+^n h_-^n y_N)$$
$$= \xi_\Lambda(\eta_0(y_N)\prod_{k=1}^{n}(2k(\Lambda_1 - N) - k(k-1))y_N)$$
$$= \prod_{k=1}^{n} k(2\Lambda_1 - 2N + 1 - k)S_0(y_N,y_N)$$
$$= \prod_{k=1}^{n} k(2\Lambda_1 - 2N + 1 - k) \times \prod_{T=1}^{N} \alpha_T(2\Lambda_1 + 1 - 2T)(2\Lambda_1 + 2 - 2T).$$

Finally,

$$S_0(y_M^m,y_N^n) = \delta_{NM}\delta_{mn}S_0(y_N^n,y_N^n),$$

with $h_+y_N = 0$.

Now we define new parameters: $l = \Lambda_1 - N$, $m = \Lambda_1 - N - n$, $l_0 = -i\Lambda_2$, $l_1 = \Lambda_1 + 1$. Then α_N,β_N go over into

$$\alpha_l = \{(l+1)^2(4(l+1)^2 - 1)\}^{-1} \times ((l+1)^2 - l_0^2)(l_1^2 - (l+1)^2),$$
$$\beta_l = l_0 l_1 / l(l+1),$$

and we obtain, with $N = l_1 - 1 - l$, $n = l - m$, $N,n \in \mathrm{N}$ ($y_0^0 = y_0 = 1$),

$$S_0(y_{l_1-1},y_{l_1-1}) = 1, \quad N = 0,$$
$$S_0(y_M,y_N) = \delta_{MN}\prod_{s=l}^{l_1-2}\alpha_s(2s+2)(2s+1), \quad M + N > 0,$$
$$S_0(y_0^k,y_0^k) = \delta_{kn}\prod_{t=l}^{m+1}(l+t)(l-t+1), \quad k + n > 0,$$
$$S_0(y_M^k,y_N^n) = \delta_{MN}\delta_{kn}\prod_{t=l}^{m+1}(l+t)(l-t+1) \times \prod_{s=l}^{l_1-2}\alpha_s(2s+1)(2s+2),$$
$$k + n > 0, \quad M + N > 0.$$

The values for the parameters l,m are

$$l = l_1 - 1, l_1 - 2, l_1 - 3, \ldots, \quad \text{for } N = 0,1,2,\ldots,$$
$$m = l, l-1, l-2, \ldots, \quad \text{for } n = 0,1,2,\ldots,$$

and thus

$$y_l^m \leftrightarrow y_N^n.$$

It follows for so(4) on $d_{l_1l_0} = d_\Lambda$, where $\Lambda_1 = l_1 - 1$, $\Lambda_2 = il_0$,

$$h_3 y_l^m = m y_l^m, \quad h_+ y_l^m = (l-m)(l+m+1)y_l^{m+1},$$
$$h_- y_l^m = y_l^{m-1},$$
$$p_3 y_l^m = -\alpha_l(l+m+1)y_{l+1}^m + \beta_l m y_l^m - (l-m)y_{l-1}^m,$$
$$p_+ y_l^m = \alpha_l(l+m+1)(l+m+2)y_{l+1}^{m+1} + \beta_l(l-m)(l+m+1)y_l^{m+1} - (l-m)(l-m-1)y_{l-1}^{m+1},$$
$$p_- y_l^m = -\alpha_l y_{l+1}^{m-1} + \beta_l y_l^{m-1} + y_{l-1}^{m-1},$$
$$l = l_1 - 1 - N, \quad m = l - n, \quad n,N \in \mathrm{N}, \quad l_1,l_0 \in \mathrm{C},$$

with α_l,β_l given above. This representation on $d_{l_1l_0}$ becomes indecomposable for values $l_1 = \frac{1}{2}k$, $l_0 = \frac{1}{2}k'$, $l_1 = l_0 + s$, $k,k' \in \mathrm{N}^+$, $s \in \mathrm{N}$, since $\alpha_{l_0-1} = 0$ and $h_+ y_l^{-l-1} = 0$. These two properties define an infinite-dimensional invariant subspace $\tilde{\pi}_{l_1l_0}$ of the Verma module $d_{l_1l_0}$. The quotient space $\pi_{l_1l_0} \sim d_{l_1l_0}/\tilde{\pi}_{l_1l_0}$ is finite dimensional and its basis states y_l^m are labelled by the parameter values

$$l = l_1 - 1, l_1 - 2, \ldots, l_0, \quad l_1 = l_0 + s, \quad l_0 = \tfrac{1}{2}k,$$
$$k,s \in \mathrm{N},$$
$$m = l, l-1, \ldots, -l.$$

On the quotient space $\pi_{l_1l_0}$ the sesquilinear form S_0 induces a scalar product, since it is nondegenerate and positive definite. Defining an orthonormal basis

$$|l_1,l_0;l,m\rangle \equiv |l,m\rangle = \{S_0(y_l^m,y_l^m)\}^{-1/2} y_l^m,$$

$$S_0(y_l^m,y_l^m)$$

$$= \frac{(2l)!}{(l+m)!}(l-m)! \prod_{s=l}^{l_1-2} \alpha_s(2s+1)(2s+2)$$

$$= \frac{(2l)!(l-m)!}{(l+m)!} \frac{(l_1-(l+1))!(l!)^2\Gamma(l+\frac{3}{2})}{(l_1+l)!(l-l_0)!(l+l_0)!\Gamma(l+1)}$$

$$\times \frac{\Gamma(l_1)(l_1-1-l_0)!(l_1-1+l_0)!(2l_1-1)!}{((l_1-1)!)^2\Gamma(l_1+\frac{1}{2})},$$

where Γ denotes the gamma function, one obtains the familiar form for the (infinitesimally) unitary representations of so(4) on the finite-dimensional quotient module π_{l_1,l_0}, $l_0=\frac{1}{2}k$, $l_1=l_0+s$, $k,s\in\mathrm{N}$,

$$h_3|l,m\rangle = m|l,m\rangle,$$

$$h_+|l,m\rangle = \sqrt{(l-m)(l+m+1)}\,|l,m+1\rangle,$$

$$h_-|l,m\rangle = \sqrt{(l+m)(l-m+1)}\,|l,m-1\rangle,$$

$$p_3|l,m\rangle = -\sqrt{(l+1)^2-m^2}\sqrt{\alpha_l}\,|l+1,m\rangle + \beta_l m|l,m\rangle - \sqrt{l^2-m^2}\sqrt{\alpha_{l-1}}\,|l-1,m\rangle,$$

$$p_+|l,m\rangle = \sqrt{(l+m+1)(l+m+2)}\sqrt{\alpha_l}\,|l+1,m+1\rangle + \sqrt{(l-m)(l+m+1)}\beta_l|l,m+1\rangle - \sqrt{(l-m)(l-m-1)}\sqrt{\alpha_{l-1}}\,|l-1,m+1\rangle,$$

$$p_-|l,m\rangle = -\sqrt{(l-m+1)(l-m+2)}\sqrt{\alpha_l}\,|l+1,m-1\rangle + \sqrt{(l+m)(l-m+1)}\beta_l|l,m-1\rangle + \sqrt{(l+m)(l+m-1)}\sqrt{\alpha_{l-1}}\,|l-1,m-1\rangle.$$

It is thus seen that the Harish-Chandra sesquilinear form S_0 on U[so(4)] induces an inner product on the (finite-dimensional) quotient module $\pi_{l_1l_0}\sim d_{l_1l_0}/\bar{\pi}_{l_1l_0}$ of the Verma module $d_{l_1l_0}$ with respect to the invariant submodule $\bar{\pi}_{l_1l_0}$. This inner product unitarizes the finite-dimensional representation of so(4), that is, brings these representations into the familiar form as needed for physical applications. Moreover, all irreducible unitarizable representations of so(4) are obtained in this manner.

Via analytic continuation in the parameters one could now proceed to obtain the unitary (infinite-dimensional) representations of the noncompact real (Lorentz) algebra so(3,1). This, however, corresponds to negative integer values of n,N, and thus cannot be carried out on the Verma modules that were chosen for the present analysis. That is, our goal of relating the physically relevant representations of the Lorentz algebra so(3,1) to an inner product $(\ ,\ )_1$ induced by a Harish-Chandra sesquilinear form S_1 on U(so(3,1)) cannot be achieved on the modules chosen in this article. We will discuss this situation, and related matters, in another article.

Note added in proof: Scalar products and unitarization for the algebras SU(2), SU(1,1) and the Heisenberg–Wehl algebra H are also discussed in Ref. 10.

[1] Harish-Chandra, "Representations of Semisimple Lie Groups," Am. J. Math. **77**, 743 (1955).

[2] M. Lorente and B. Gruber, "Construction of Extremal Vectors for Verma Submodules of Verma Modules," J. Math. Phys. **25**, 1674 (1984).

[3] B. Gruber and A. U. Klimyk, "Matrix Elements for Indecomposable Representations of Complex su(2)," J. Math. Phys. **25**, 755 (1984).

[4] B. Gruber and A. U. Klimyk, "Multiplicity Free and Finite Multiplicity Indecomposable Representations of the Algebra su(1,1)," J. Math. Phys. **19**, 2009 (1978).

[5] I. M. Gel'fand and A. A. Kirillov, "The Structure of the Lie Field Connected with a Split Lie Algebra," Funct. Anal. Appl. **3**, 6 (1969).

[6] N. Shapovalov, "On a Bilinear Form on the Universal Enveloping Algebra of a Semisimple Lie Algebra," Funct. Anal. Appl. **6**, 65 (1972).

[7] H. Jakobsen, "The Last Possible Place for Certain Highest Weight Modules," Math. Ann. **256**, 439 (1981).

[8] H. Jakobsen, "Hermitian Symmetric Spaces and their Unitary Highest Weight Modules," J. Funct. Anal. **52**, 385 (1983).

[9] B. Gruber and R. Lenczewski, "Indecomposable Representations of the Lorentz Algebra in an Angular Momentum Basis," J. Phys. A **16**, 3703 (1983).

[10] H. D. Doebner, B. Gruber, and M. Lorente, "Boson Operator Realizations of SU(2) and SU(1,1) and Unitarization," J. Math. Phys. **30**, 594 (1989).

Integral bounds for radar ambiguity functions and Wigner distributions

Elliott H. Lieb
Departments of Mathematics and Physics, Princeton University, P. O. Box 708, Princeton, New Jersey 08544

(Received 10 November 1989; accepted for publication 22 November 1989)

An upper bound is proved for the L^p norm of Woodward's ambiguity function in radar signal analysis and of the Wigner distribution in quantum mechanics when $p>2$. A lower bound is proved for $1\leqslant p<2$. In addition, a lower bound is proved for the entropy. These bounds set limits to the sharpness of the peaking of the ambiguity function or Wigner distribution. The bounds are best possible and equality is achieved in the L^p bounds if and only if the functions f and g that enter the definition are both Gaussians.

I. INTRODUCTION

The *ambiguity function* introduced by Woodward[1] is important in radar signal analysis. It is a function of two real variables, τ (the time) and ω (with $2\pi\omega$ being the frequency), and is defined as follows in terms of two given functions f and g of one variable:

$$A_{f,g}(\tau,\omega)=\int f\left(t-\frac{1}{2}\tau\right)g^*\left(t+\frac{1}{2}\tau\right)e^{-2\pi i\omega t}\,dt. \tag{1.1}$$

(Our conventions will be that * as a superscript denotes complex conjugate and all integrals are from $-\infty$ to $+\infty$.) Strictly speaking, $A_{f,g}$ is called the *cross-ambiguity function* of f and g while $A_{f,f}$ is the *proper ambiguity function* of f. Usually, one assumes that f and g are square integrable, which guarantees that the integrand of (1.1) is a summable function of t for every τ. The summability can also be guaranteed by Hölder's inequality and the alternative assumption that $f\in L^a$ and $g\in L^b$ (with $1/a+1/b=1$ and $1\leqslant a,b\leqslant\infty$), as in Definition 2 below, and this generalized hypothesis will often be made in this paper.

There is a simple relation between $A_{f,g}$ and $W_{f,g}$, the (*cross*) *Wigner distribution of f and g* used in quantum mechanics and defined by

$$W_{f,g}(\tau,\omega)=\int f\left(\tau+\frac{1}{2}s\right)g^*\left(\tau-\frac{1}{2}s\right)e^{-2\pi i\omega s}\,ds. \tag{1.2}$$

The relation is

$$W_{f,g}(\tau,\omega)=2A_{f,g^-}(-2\tau,2\omega), \tag{1.3}$$

where f^- denotes the function given by

$$f^-(t)\equiv f(-t). \tag{1.4}$$

$W_{f,f}$ is called *the Wigner distribution* (*or density*) *of f*. Because of (1.3) the bounds obtained here for $A_{f,g}$ apply *mutatis mutandis* to $W_{f,g}$.

Ideally, one would like to choose f and g so that $A_{f,g}$ is sharply peaked around some point (τ_0,ω_0) but, as is well known, there are severe limitations to the peaking that can be achieved. These limitations are inherent in the definition (1.1). Let us define, for $p>0$,

$$I_{f,g}(p)=\int\int|A_{f,g}(\tau,\omega)|^p\,d\tau\,d\omega. \tag{1.5}$$

If $A_{f,g}$ were highly peaked then $I_{f,g}(p)$ would be very large for large p and very small for small p. The dividing line is $p=2$ since, by the Parseval's inversion formula, we have the identity

$$I_{f,g}(2)=\int|f(t)|^2dt\int|g(t)|^2\,dt. \tag{1.6}$$

In this paper, limitations on the sharpness of $A_{f,g}$ will be established by proving that $I_{f,g}(p)$ is universally bounded above when $p>2$ (Theorem 1) and universally bounded below when $1\leqslant p<2$ (Theorem 2). For $1<p<2$ and $2<p<\infty$ the bounds will be shown to be saturated if and *only* if f and g are Gaussians. It is remarkable that Gaussians both maximize and minimize $I_{f,g}(p)$, depending on the value of p.

When $p=2$ the identity (1.6) holds for any f and g, so the obvious quantity to consider is the derivative with respect to p of $I_{f,g}(p)$ at $p=2$ under the normalization assumption that the right side of (1.6) is unity. This derivative, multiplied by -2, is the *entropy* given by

$$S_{f,g}=-\int\int|A_{f,g}(\tau,\omega)|^2\ln|A_{f,g}(\tau,\omega)|^2\,d\tau\,d\omega, \tag{1.7}$$

with $0\ln 0\equiv 0$. It will be proved that when the right side of (1.6) is unity the integral in (1.7) is well defined and (Theorem 3)

$$S_{f,g}\geqslant 1. \tag{1.8}$$

This constant is sharp since it is achieved by Gaussians.

To state the theorems precisely it is first necessary to make some definitions.

Definition 1: $f(t)$ is said to be a *Gaussian* if

$$f(t)=\exp\left[-\alpha t^2+\beta t+\gamma\right], \tag{1.9}$$

with α, β, and γ being complex numbers and with $\mathrm{Re}(\alpha)>0$; $f(t)$ is a *real Gaussian* if α, β, and γ are real numbers with $\alpha>0$. Two functions f and g are said to be a *matched Gaussian pair* if they are both Gaussians with the same α but with possibly different β's and γ's.

Definition 2: For $0<p<\infty$

$$\|f\|_p\equiv\left\{\int|f(t)|^p dt\right\}^{1/p} \tag{1.10}$$

and for $p=\infty$

$$\|f\|_\infty\equiv\text{ess sup}|f(t)|. \tag{1.11}$$

 0022-2488/90/030594-06$03.00

We say that $f \in L^p$ if and only if the right side of (1.10) or (1.11) is finite.

Definition 3: Let $0<p\leqslant\infty$ and define q by $1/q+1/p=1$, i.e., $q=p/(p-1)$. Note that $\infty\geqslant q\geqslant 1$ if $1\leqslant p\leqslant\infty$, and $0>q>-\infty$ if $0<p<1$. Then C_p is defined to be

$$C_p = p^{1/2p}|q|^{-1/2q}, \tag{1.12}$$

for $p\neq 1$ or ∞ while

$$C_1 = C_\infty = 1. \tag{1.13}$$

Note that $C_2 = 1$.

Definition 4: Let p and q be as in Definition 3 with $1<p<\infty$ and let a and b satisfy $1/a+1/b=1$ with $1\leqslant a\leqslant\infty, 1\leqslant b\leqslant\infty$. We define $H(p,a,b)\geqslant 0$ by

$$H(p,a,b)^2 = abp^{-2}|p-2|^{2-p} \times|p-a|^{-1+p/a}|p-b|^{-1+p/b}, \tag{1.14}$$

with the convention that $0^0\equiv 1$. When a or $b=p$

$$H(p,a,b)^2 = \frac{1}{p}\left(\frac{p}{p-1}\right)^{p-1}$$

and

$$H(1,1,\infty) = H(1,\infty,1) = 1. \tag{1.15}$$

We also define $K(p,a,b)\geqslant 0$ by

$$K(p,a,b)^2 = p^{-2}2^{2-p}a^{p/a}b^{p/b}$$

and

$$K(1,1,\infty) = \sqrt{2}. \tag{1.16}$$

The following relations (with $1/p+1/q=1$) are noteworthy for $p>1$:

$$H(p,a,b) = C_q^p\{C_{a/q}C_{b/q}/C_{p/q}\}^{p/q}, \tag{1.17}$$

$$H(p,a,b)^{1/p}H(q,a,b)^{1/q} = K(p,a,b)^{1/p}K(q,a,b)^{1/q} = a^{1/a}b^{1/b}p^{-1/p}q^{-1/q}. \tag{1.18}$$

Theorem 1: *Let $p>2$ and assume that f and $g\in L^2$. Then*

(a) $I_{f,g}(p)\leqslant(2/p)\{\|f\|_2\|g\|_2\}^p$. (1.19)

(b) Equality is achieved in (1.19) if and only if f and g are a matched Gaussian pair.

(c) More generally, if $f\in L^a, g\in L^b$ with $1/a+1/b=1$ and with $p/(p-1)\leqslant a\leqslant p$ and $p/(p-1)\leqslant b\leqslant p$ then

$$I_{f,g}(p)\leqslant H(p,a,b)\{\|f\|_a\|g\|_b\}^p. \tag{1.20}$$

When both a and $b>p/(p-1)$ equality is achieved in (1.20) if and only if f and g are Gaussians that satisfy

$$\begin{aligned} f(t) &= \exp[-(\alpha m' + iA)t^2+\beta t+\gamma],\\ g(t) &= \exp[-(\alpha n' + iA)t^2+\tilde\beta t+\tilde\gamma], \end{aligned} \tag{1.21}$$

with α, A real, $\alpha>0$, and $\beta,\tilde\beta,\gamma,\tilde\gamma$ complex and with $m'=a(p-1)/(ap-a-p)$ and $n'=b(p-1)/(bp-b-p)$. [Note that $(p-1)/(p-2)<m',n'<\infty$ under the stated conditions.] When a or $b=p/(p-1)$, (1.20) is best possible, but equality is never achieved.

(d) If the additional condition that $g=f$ is imposed (which means that the proper ambiguity function $A_{f,f}$ is being considered) or else that $g=f^-$ (which means that the proper Wigner distribution $W_{f,f}$ is being considered) then (1.20) can be improved. In these cases (and with a and b restricted as before)

$$I_{f,f^-}(p) \text{ and } I_{f,f}(p)\leqslant K(p,a,b)\{\|f\|_a\|f\|_b\}^p. \tag{1.22}$$

Equality is achieved in (1.22) if and only if f is any Gaussian. Note that a or $b=p/(p-1)$ is allowed here.

Remarks: (1) Even if f and g are Gaussians, it is not possible to have equality in (1.20) for *all a and b simultaneously*, as (1.21) shows.

(2) In view of the symmetry of $A_{f,g}$ between the pair f,g and the Fourier transforms $\hat f,\hat g$ expressed by (2.4) below, Theorem 1 remains true if f and g are replaced by $\hat f$ and $\hat g$ on the right side of (1.19) *et seq.*

In the case that p is an even integer, Theorem 1(a) and (b) [under the additional assumption for (b) that f and g are twice continuously differentiable and never vanish] was proved by Price and Hofstetter[2] by an ingenious application of the Cauchy–Schwarz inequality. They conjectured Theorem 1(a) and (b) for all $p>2$ in their footnote 10. The Price–Hofstetter bounds have found application in the work of Janssen[3] for example.

The next theorem gives reversed inequalities for $p<2$.

Theorem 2: *Assume that $t\mapsto f(t-\frac{1}{2}\tau)g^*(t+\frac{1}{2}\tau)$ is in L^1 for every τ, so that the definition (1.1) of $A_{f,g}(\tau,\omega)$ makes sense. (This L^1 condition can be satisfied, for example, by assuming that $f\in L^\alpha$ and $g\in L^\beta$ for some $1\leqslant\alpha,\beta\leqslant\infty$ with $1/\alpha+1/\beta=1$.) Let $1\leqslant p<2$ and assume that $0<I_{f,g}(p)<\infty$. Then f and $g\in L^r$ for every $p\leqslant r\leqslant q\equiv p(p-1)$. Moreover, for every pair a,b with $p\leqslant a,b\leqslant q$ and with $1/a+1/b=1$ we have that*

$$I_{f,g}(p)\geqslant H(p,a,b)\{\|f\|_a\|g\|_b\}^p. \tag{1.23}$$

In particular,

$$I_{f,g}(p)\geqslant(2/p)\{\|f\|_2\|g\|_2\}^p. \tag{1.24}$$

If $g=f$ or $g=f^-$ [as in Theorem 1(d)] then (1.23) can be improved to

$$I_{f,g}(p)\geqslant K(p,a,b)\{\|f\|_a\|g\|_b\}^p. \tag{1.25}$$

If $1<p<2$ equality is achieved in (1.23) if and only if f and g satisfy (1.21) et seq., but with m' and n' replaced by $|m'|$ and $|n'|$, respectively. Equality in (1.25) occurs if and only if f is any Gaussian. If $p=1$ and $a,b>1$ equality occurs in (1.23) if f and g are given by (1.21), but $\alpha|m'|$ and $\alpha|n'|$ have to be interpreted as αa and αb, respectively (since $|m'|/|n'|\to a/b$ but $m',n'\to 0$ as $p\to 1$).

Remarks: (3) When $p=1$ and $a,b>1$ the Gaussians referred to in the last part of Theorem 2 are, in fact, the only functions for which equality holds in (1.21). A proof can be constructed by using ideas in Ref. 4, but it will not be given here. The uniqueness of Gaussian minimizers for $p=1$ and $a=b=2$ is closely related to and can be inferred from a theorem of Hudson[5] (see also Ref. 6) which says that the only way in which the function $A_{f,g}(\tau,\omega)$ can be a non-negative function of τ and ω is when $f=\lambda g$ for some $\lambda>0$ and f is a Gaussian. (Actually, Hudson does this in the context of the Wigner distribution, but that is immaterial; also he proves the theorem only for $W_{f,f}$ but his method, extends to the general case.) The connection is established by first noting the relation for summable $A_{f,g}$ (which is easy to derive—at least formally)

$$\int A_{f,g}(\tau,\omega)dt\,d\omega = 2\int f(t)g^*(t)dt. \tag{1.26}$$

On the other hand, by Theorem 2(a) with $p=1$,

$$\int |A_{f,g}(\tau,\omega)|d\tau\,d\omega \geqslant 2\|f\|_2\|g\|_2. \tag{1.27}$$

If $A_{f,g}\geqslant 0$, the left sides of (1.26) and (1.27) are identical, which then requires that $f=\lambda g$ and that equality holds in (1.27). Thus $A_{f,g}\geqslant 0$ is equivalent to equality in (1.24) for $p=1$.

(4) Theorem 2(c) is striking when $p=a=1$ and $b=\infty$. Then

$$\int |A_{f,g}(\tau,\omega)|d\tau\,d\omega \geqslant \|f\|_1\|g\|_\infty. \tag{1.28}$$

This says that if f is fixed and $g\to 0$ in all L^p norms except $p=\infty$, then $\int|A|$ does not go to zero. [For example, $g(t)=\exp[-\lambda t^2]$ with $\lambda\to\infty$.] The Fourier transform also has this property [cf. (2.9)] and it is inherited by $A_{f,g}$.

A tempting conjecture is that inequality (1.24), at least, should hold if $0<p<1$. Our proof fails in this case because Lemma 1 below requires $p\geqslant 1$.

It is instructive to compare Theorems 1(a) and (1.24) by considering Gaussians $f(t)=\exp(-\alpha t^2)$ and $g(t)=\exp(-\beta t^2)$ with $\operatorname{Re}\alpha$ and $\operatorname{Re}\beta>0$. Then one finds

$$\begin{aligned} &I_{f,g}(p)\|f\|_2^{-p}\|g\|_2^{-p} \\ &\quad = (2/p)[\operatorname{Re}\alpha\operatorname{Re}\beta\,]^{p/4-1/2}|(\alpha+\beta^*)/2|^{1-p/2}. \end{aligned} \tag{1.29}$$

Since $\operatorname{Re}\alpha\operatorname{Re}\beta\leqslant\frac{1}{4}|\alpha+\beta^*|^2$ one sees, for Gaussians, that (1.19) holds for $p\geqslant 2$ and that the reverse inequality holds for *all* $0<p<2$, and that equality requires $\alpha=\beta$ in both cases.

Theorem 3: *Assume that f and $g\in L^2$ with $\|f\|_2\|g\|_2=1$. Then*

$$S_{f,g}\geqslant 1.$$

Equality is achieved iff f and g are a matched Gaussian pair.

Remarks: (5) It is possible to show that equality is achieved in Theorem 3 only when f and g are matched Gaussians. The proof is complicated and will not be given; the reader is invited to find a simple proof.

The method of proof of these three theorems follows closely the methods used in Ref. 7 to prove L^p bounds of coherent state transforms. The coherent state transform of f is $A_{f,g}(-\tau,-\omega)\exp(i\pi\omega\tau)$ with g being the fixed Gaussian $g(t)=\tau^{-1/4}\exp(-t^2/2)$. From the mathematical point of view there is, however, a genuinely new development in the present paper, namely the proof that Gaussians uniquely saturate the bounds. This uses Ref. 4.

II. PRELIMINARY LEMMAS

The following convention for the Fourier transform $\hat f$ of a function f will be employed:

$$\hat f(\omega) = \int f(t)e^{-2\pi i\omega t}\,dt, \tag{2.1}$$

so that

$$f(t) = \int \hat f(\omega)e^{2\pi i\omega t}\,d\omega \tag{2.2}$$

and Parseval's relation is

$$\|f\|_2 = \|\hat f\|_2. \tag{2.3}$$

The equality (1.6) follows from (2.3). Some other important facts about $A_{f,g}$ which follow easily from (2.3), the Cauchy–Schwarz inequality and a change of integration variables are

$$A_{f,g}(\tau,\omega) = A_{\hat f,\hat g}(-\omega,\tau), \tag{2.4}$$

$$A^*_{f,g}(\tau,\omega) = A_{g,f}(-\tau,-\omega), \tag{2.5}$$

$$|A_{f,g}(\tau,\omega)|\leqslant\|f\|_2\|g\|_2. \tag{2.6}$$

More generally, if $f\in L^a$, $g\in L^b$ with $1/a+1/b=1$ and $a\geqslant 1, b\geqslant 1$, as in Theorems 1 and 2, Hölder's inequality yields the pointwise bound

$$|A_{f,g}(\tau,\omega)|\leqslant\|f\|_a\|g\|_b. \tag{2.7}$$

Inequality (2.6) is important because it implies that $\ln|A_{f,g}(\tau,\omega)|^2\leqslant 0$ when $\|f\|_2\|g\|_2=1$ and hence $S_{f,g}$ is always well defined by the right side of (1.7) (although it might be $+\infty$).

Three inequalities in Fourier analysis will be needed. The first fact is the sharp constant in the Hausdorff–Young inequality (2.8) proved by Beckner.[8] The criterion for equality is due to Lieb.[4]

Lemma 1: *Let $2\leqslant p\leqslant\infty$ and $1/q=1-1/p=(p-1)/p$. If $f\in L^q$ then $\hat f\in L^p$ and*

$$\|\hat f\|_p\leqslant C_q\|f\|_q. \tag{2.8}$$

Conversely, let $1\leqslant p\leqslant 2$ and assume $f\in L^r$ for some $1\leqslant r\leqslant 2$, in which case $\hat f$ exists by (2.8) (with $q\equiv r$ there.) If $\hat f\in L^p$ then $f\in L^q$ with $q=(p-1)/p$ and

$$\|\hat f\|_p\geqslant C_q\|f\|_q. \tag{2.9}$$

Equality is achieved in (2.8) when $2<p<\infty$ and in (2.9) when $1<p<2$ if and only if f is any Gaussian with α real and β,γ complex in (1.9).

Proof: Inequality (2.8) is Beckner's result, and the condition for equality when $2<p<\infty$ is proved in Ref. 4. For (2.9), let $g\equiv\hat f$. Since $f\in L^r$, $g\in L^s$ [with $s=r/(r-1)\geqslant 2$]. Therefore, $g\in L^p\cap L^s$ and hence, by convexity, $g\in L^2$. Thus $\hat g$ exists and, by the L^2 Fourier inversion formula, $\hat g=f^-$. By (2.8), $f^-\in L^q$ and (using $C_qC_p=1$) $C_q\|f\|_q=C_q\|f^-\|_q\leqslant\|g\|_p=\|\hat f\|_p$. Obviously, the condition for equality when $1<p<2$ follows from the $2<p<\infty$ result. Q.E.D.

Remark: (6) The classical Hausdorff–Young inequality is (2.8) but with C_q replaced by the larger value 1.

The next inequality is the sharp constant in Young's inequality, which was found simultaneously by Beckner[8] and by Brascamp and Lieb.[9] The uniqueness part (b) is due to Brascamp and Lieb.[9] In the following a midline asterisk denotes convolution

$$(f*g)(t) = \int f(t-s)g(s)ds. \tag{2.10}$$

Lemma 2: *Let $1/m+1/n=1+1/r$ with $1\leqslant m\leqslant\infty$, $1\leqslant n\leqslant\infty$, $1\leqslant r\leqslant\infty$. Then, when $f\in L^m$ and $g\in L^n$, $f*g\in L^r$ and*
(a)

$$\|f*g\|_r\leqslant(C_m C_n/C_r)\|f\|_m\|g\|_n. \qquad (2.11)$$

(b) When $m>1$ and $n>1$, equality holds in (2.11) if and only if

$$f(t)=\exp[-\alpha m't^2+\beta t+\gamma],$$
$$g(t)=\exp[-\alpha n't^2+\tilde\beta t+\tilde\gamma], \qquad (2.12)$$

with $\alpha>0$ real, $\beta,\gamma,\tilde\beta,\tilde\gamma$ complex but with $\mathrm{Im}(\beta)=\mathrm{Im}(\tilde\beta)$. Here, $m'=m/(m-1)$ and $n'=n/(n-1)$. If $m=1$ or $n=1$ and $r>1$, (2.11) is at best possible but equality is never achieved. If $m=n=r=1$, equality is achieved when f and g are any pair of non-negative, real valued functions.
(c) If $g^*=f$ or $g^*=f^-$

$$\|f*g\|_r\leqslant\tfrac12 r^{-1/2r}(2m)^{1/2m}(2n)^{1/2n}\|f\|_m\|f\|_n. \qquad (2.13)$$

For all $m\geqslant1$ and $n\geqslant1$ and $r>1$ equality is achieved in (2.13) if and only if f is a Gaussian given by (1.9) with α real and with β real (if $g^*=f$) or β complex (if $g^*=f^-$).

Remarks: (7) The classical inequality of Young is (2.11) but with C_mC_n/C_r replaced by the larger value 1.

(8) Lemma 2(c) was not given in Ref. 9 because it did not occur to us at the time that it might be useful. It is however, a simple consequence of the analysis in Ref. 9.

The third inequality is the converse of Young's inequality. It was first proved by Leindler[10] with 1 in place of C_mC_n/C_r. The sharp form below is due to Brascamp and Lieb.[9]

Lemma 3: *Let f(t) and g(t) be non-negative, real-valued functions that are not identically zero and assume that $f*g\in L^r$. Let $1/m+1/n=1+1/r$ with $0<m\leqslant1$, $0<n\leqslant1$. Note that $0<r\leqslant1$. Then $f\in L^m$ and $g\in L^n$ and*
(a)

$$\|f*g\|_r\geqslant(C_mC_n/C_r)\|f\|_m\|g\|_n. \qquad (2.14)$$

Equality holds in (2.14) when $m<1$ and $n<1$ if and only if

$$f(t)=\exp[\alpha m't^2+\beta t+\gamma],$$
$$g(t)=\exp[\alpha n't^2+\tilde\beta t+\tilde\gamma], \qquad (2.15)$$

with $\alpha>0$ real and $\beta,\gamma,\tilde\beta,\tilde\gamma$ real. Here, $m'=m/(m-1)<0$ and $n'=n/(n-1)<0$.

(b) If $g^*=f$ or $g^*=f^-$ (2.14) can be improved to

$$\|f*g\|_r\geqslant\tfrac12 r^{-1/2r}(2m)^{1/2m}(2n)^{1/2n}\|f\|_m\|f\|_n, \qquad (2.16)$$

with equality (for all m and n) if and only if f is a real Gaussian.

Remark: (9) Lemma 3(b) was not given in Ref. 9 but it is a simple consequence of the analysis given there.

The next lemma is an extension of the Cauchy functional equation to quadratics. [One form of Cauchy's equation is $\xi(t-\frac12\tau)\eta(t+\frac12\tau)=\rho(\tau)$ with ξ and η being Lebesque measurable functions; the only solution is $\xi(t)=be^{-At},\eta(t)=ce^{At}$, and $\rho(\tau)=bce^{A\tau}$ for some constants A, b, c.]

Lemma 4: *Let ξ and η be complex valued, Lebesgue measurable functions on **R** that satisfy $|\xi(t)|=|\eta(t)|=1$ for all t. Suppose there are real valued functions, μ and ν, on **R** (which are not a priori measurable) such that for almost every τ the following holds for almost every t:*

$$\xi(t-\tfrac12\tau)\eta(t+\tfrac12\tau)=\exp[i\mu(\tau)t+i\nu(\tau)]. \qquad (2.17)$$

Then there are real constants, A, α, β, γ and δ such that

$$\xi(t)=\exp[iAt^2+i\alpha t+i\gamma],$$
$$\eta(t)=\exp[-iAt^2-i\beta t-i\delta]. \qquad (2.18)$$

Proof: Let $\mathscr{B}$ denote the set of τ such that (2.17) holds for almost all t. Let $X(t)=\xi(t)\exp(-t^2)$ and $Y(t)=[1/\eta(t)]\exp(-t^2)$. Using the definition (2.1) of the Fourier transform, it is a simple matter to use the Gaussian bound on $X(t)$ to deduce that $\hat X$ is an entire analytic function of order at most 2, i.e., $|\hat X(\omega)|\leqslant\exp[C+D|\omega|^2]$ for suitable $C,D>0$ and all $\omega\in\mathbf{C}$. (In fact, $|\hat X(\omega)|\leqslant\sqrt{\pi}\exp[\pi^2(\mathrm{Im}\,\omega)^2]$.) The same is true of $\hat Y(\omega)$. From (2.17), for every $\tau\in\mathscr{B}$ the following holds for almost every t:

$$X(t-\tfrac12\tau)=Y(t+\tfrac12\tau)\exp\{t[i\mu(\tau)+2\tau]+i\nu(\tau)\}. \qquad (2.19)$$

Taking Fourier transforms of (2.19) with respect to t we find that

$$\hat X(\omega)\exp(-\pi i\omega\tau)$$
$$=\hat Y\Big(\omega-\frac{\mu(\tau)}{2\pi}+\frac{i\tau}{\pi}\Big)$$
$$\times\exp\Big[\pi i\omega\tau-\frac12 i\mu(\tau)\tau+i\nu(\tau)-\tau^2\Big]. \qquad (2.20)$$

We claim that $\hat X(\omega)$ has no zeros, for otherwise suppose that $\hat X(\omega_0)=0$. Then $\hat Y(\omega)=0$ whenever ω satisfies

$$\omega=\omega_0-(1/2\pi)\mu(\tau)+(i/\pi)\tau, \qquad (2.21)$$

for some $\tau\in\mathscr{B}$. As τ ranges over the uncountable set $\mathscr{B}$, the right side of (2.21) ranges over an uncountable set in the complex plane. [Note that $\mu(\tau)$ is real and $i\tau$ is imaginary so there can be no cancellation in (2.21).] The only entire function with uncountably many zeros is the zero function, so $\hat Y(\omega)\equiv0$. This implies that $Y(t)=0$, which is a contradiction. By reversing the roles of X and Y we find that $\hat Y(\omega)$ has no zeros. Because $\hat X$ and $\hat Y$ are entire analytic and zero free they have analytic logarithms, e.g., $\hat X(\omega)=\exp[\phi(\omega)]$ for some entire analytic function ϕ. Since $\hat X$ has order at most 2, $|\phi(\omega)|\leqslant C|\omega|^2+D$ for suitable $C,D>0$. But then ϕ must be a polynomial of order 2, i.e., $\hat X$ is a Gaussian. The same is true of $\hat Y$. By taking the inverse Fourier transform, we have that X and Y are Gaussians, which, by inspection, proves (2.18). Q.E.D.

III. PROOF OF THEOREM 1

Step 1: Fix $\tau\in\mathbf{R}$. Since $f\in L^a$ and $g\in L^b$ with $1/a+1/b=1$, the function $t\to f(t-\frac12\tau)g(t+\frac12\tau)$ is in L^1. Since $A_{f,g}$ is the Fourier transform of this L^1 function, we can use Lemma 1 with $q=p/(p-1)<2$ in place of p there and obtain

$$\int |A_{f,g}(\tau,\omega)|^p\, d\omega$$
$$\leqslant C_q^p \left\{\int \left| f\left(t-\frac{1}{2}\tau\right)\right|^q \left| g\left(t+\frac{1}{2}\tau\right)\right|^q dt\right\}^{p/q}. \qquad (3.1)$$

Note that the right-hand integral may be finite or infinite—depending on τ. If it is infinite then (3.1) is trivially true; if it is finite then the use of Lemma 1 is justified. We shall see in step 2 that this integral is finite for almost every τ.

Step 2: The integral on the right side of (3.1) is just the convolution

$$J(\tau) \equiv (|f^-|^q * |g|^q)(\tau). \qquad (3.2)$$

Integrating (3.1) over τ and applying Lemma 2 to $J(\tau)$ with $r = p/q > 1$ and $m = a/q \geqslant 1, n = b/q \geqslant 1$, we have

$$\begin{aligned} I_{f,g}(p) &\leqslant C_q^p \|J\|_r^r \\ &\leqslant C_q^p (C_m C_n/C_r)^r \{\||f^-|^q\|_m \||g|^q\|_n\}^r \\ &= C_q^p (C_m C_n/C_r)^r \{\|f\|_a \|g\|_b\}^p. \end{aligned} \qquad (3.3)$$

The inequalities (1.19) and (1.20) are obtained by using (1.17).

Step 3: It is an elementary exercise to show that Gaussians of the form (1.21) give equality in (3.1) and (3.3), and hence that $H(p,a,b)$ is the sharp constant in (1.19) and (1.20). We want to prove that these Gaussians *uniquely* saturate the bounds. Assume that $m > 1$ and $n > 1$. If there is equality in (1.19) or (1.20) then (3.1) must be an equality for almost every τ and (3.3) must be an equality. By Lemma 1, the following must be true for almost every τ:

$$f(t-\tfrac{1}{2}\tau)\, g^*(t+\tfrac{1}{2}\tau) = D(\tau)\exp[-\sigma(\tau)t^2 + \delta(\tau)t] \qquad (3.4)$$

for almost every t, with $\sigma(\tau) \in \mathbf{R}$ and $D(\tau), \delta(\tau) \in \mathbf{C}$. By Lemma 2, equality in (3.3) requires

$$\begin{aligned} |f(t)| &= \exp[-\alpha m' t^2 + \beta t + \gamma], \\ |g(t)| &= \exp[-\alpha n' t^2 + \beta t + \tilde{\gamma}], \end{aligned} \qquad (3.5)$$

with $m' = m/(m-1)$, $n' = n/(n-1)$, $\alpha > 0$, and $\beta, \gamma, \tilde{\gamma} \in \mathbf{R}$.

Let us define $\xi(t) = f(t)/|f(t)|$ and $\eta(t) = g^*(t)/|g(t)|$, which makes sense since $f(t)$ and $g(t)$ never vanish by (3.5). Then, comparing (3.4) and (3.5), we find that ξ and η satisfy the hypotheses of Lemma 4. The conclusion of Lemma 4, together with (3.5), gives (1.21).

Step 4: When $a \to p/(p-1)$ then $m' \to \infty$ and $n \to r$. By taking limits of Gaussians in (1.21) with $m' \to \infty$ we see that (1.20) is best possible in this case. Equality is never achieved, however. An informal way to see this is to note tht m' must be infinity. A formal proof is to note that (2.11) or (3.3) cannot be an equality when $m = 1$ and $n = r$ [as is stated in Lemma 2(b)] because of the strict convexity of the L^r norm.

Step 5: When $g = f$ or $g = f^-$ we proceed as in steps 1 to 3, making the appropriate changes and using lemma 2(c). From this we infer (1.22) and conclude that f must be a Gaussian in order to have equality. Upon inserting a Gaussian (1.9) for f and g (or g^-) in (1.1), one finds by inspection that equality in (1.22) does not impose any restriction on the Gaussian. Q.E.D.

IV. PROOF OF THEOREM 2

Before proving this theorem, it is perhaps worth noting a proof strategy that works when $a = p$ or $b = p$, but otherwise yields a weaker result. This strategy does not require Lemma 3. From Parseval's relation one has the identity

$$\begin{aligned} R_{f,g,h,j} &\equiv \int A_{f,g}(\tau,\omega) A^*_{h,j}(\tau,\omega)\, d\tau\, d\omega \\ &= \int f(t) h^*(t)\, dt \int g^*(t) j(t)\, dt, \end{aligned} \qquad (4.1)$$

for any four functions f,g,h, and j. Let $f = |f| e^{i\phi}$ and $g = |g| e^{i\psi}$ and choose $h(t) = |f(t)|^{a-1} e^{i\phi(t)}$ and $j(t) = |g(t)|^{b-1} e^{i\psi(t)}$. Then

$$R_{f,g,h,j} = \|f\|_a^a \|g\|_b^b. \qquad (4.2)$$

On the other hand, by Hölder's inequality,

$$|R_{f,g,h,j}| \leqslant I_{f,g}(p)^{1/p} I_{h,j}(q)^{1/q}. \qquad (4.3)$$

If $1 < p < 2$, then $q = p/(p-1) > 2$ and we can use Theorem 1(c) for the right-most factor in (4.3):

$$\{I_{h,j}(q)\}^{1/q} \leqslant H(q,b,a)^{1/q} \|f^{a-1}\|_b \|g^{b-1}\|_a. \qquad (4.4)$$

Since $\|f^{a-1}\|_b = \|f\|_a^{a-1}$ and $\|g^{b-1}\|_a = \|g\|_b^{b-1}$, we can combine (4.1)–(4.4) with (1.18) to obtain

$$I_{f,g}(p) \geqslant H(p,a,b) L(p,a,b)^p \{\|f\|_a \|g\|_b\}^p, \qquad (4.5)$$

where

$$L(p,a,b) = p^{1/p} q^{1/q} a^{-1/a} b^{-1/b}. \qquad (4.6)$$

If a or $b = p$ then $L(p,a,b) = 1$ and (4.5) is the desired inequality. Unfortunately, if $p < a$ then $b < q$, $L(p,a,b) < 1$ and (4.5) is too weak.

We now turn to a proof of Theorem 2 which makes use of Lemma 3. When $p > 1$, which is the case we consider first, the proof is virtually the same, *mutatis mutandis*, as for Theorem 1.

Step 1: Using inequality (2.9) (with $r = 1$) we have that (3.1) holds, but with the reversed inequality. Note that the left side of (3.1) is finite for almost every τ since $\int\{\int |A_{f,g}| d\omega\} d\tau < \infty$ by assumption.

Step 2: By (3.2) and Lemma 3, (3.3) holds with the reversed inequality. In particular, $f \in L^a$ and $g \in L^b$. This proves (1.23). Similarly, Lemma 3(b) leads to (1.25). The cases of equality for $1 < p < 2$ are handled in the same way as in step 1 of the proof of Theorem 1.

Finally we turn to the case $p = 1$.

Step 3: Suppose $p = 1 < a \leqslant b < \infty$. Then (1.23) holds for every $p \leqslant a$. As p decreases from a to 1, $H(p,a,b)$ converges to $H(1,a,b)$. On the other hand, $B(\tau,\omega) \equiv |A_{f,g}(\tau,\omega)|/\|f\|_a \|g\|_b \leqslant 1$ by (2.7) so $B(\tau,\omega)^p$ increases monotonically as p decreases. Therefore, by Lebesgue's monotone convergence theorem, $\int B(\tau,\omega)^p\, d\tau\, d\omega$ converges to $\int B(\tau,\omega)\, d\tau\, d\omega$ as $p \downarrow 1$ and this, together with (1.23) for $p > 1$, establishes (1.23) for $p = 1$. A similar proof holds for (1.25).

Step 4: Suppose $p = a = b = 1$. For each $a,b > 1$ such that $1/a + 1/b = 1$ inequality (1.23) holds by step 3. As $a \downarrow 1$ and $b \uparrow \infty$ we have that $H(1,a,b) \to H(1,1,\infty)$. Also, it is a standard fact that $\|f\|_a \to \|f\|_1$ and $\|g\|_b \to \|g\|_\infty$. A similar proof works for Eq. (1.25). Q.E.D.

V. PROOF OF THEOREM 3

It is assumed that f and $g \in L^2$ and $\|f\|_2\|g\|_2 = 1$. By (1.6), $I_{f,g}(2) = 1$ and, by (2.6), $|A_{f,g}(\tau,\omega)| \leqslant 1$ for all τ and ω. Let $p > 2$ whence, by Theorem 1, $I_{f,g}(p) \leqslant 2/p$. If we define, for $\epsilon > 0$,

$$K(\epsilon) \equiv \epsilon^{-1}\{I_{f,g}(2) - I_{f,g}(2+2\epsilon)\}, \tag{5.1}$$

we have that

$$K(\epsilon) \geqslant (1+\epsilon)^{-1}. \tag{5.2}$$

Assume now that $S_{f,g}$ defined by (1.7), is finite; otherwise the inequality (1.8) is trivial. (Note that $|A_{f,g}| \leqslant 1$ implies that $0 \leqslant S_{f,g} \leqslant \infty$.) We claim that

$$\lim_{\epsilon \downarrow 0} K(\epsilon) = S_{f,g}, \tag{5.3}$$

which, in view of (5.2), proves the inequality. Since $|A_{f,g}| \leqslant 1$ we have, for each τ and ω, that

$$0 \leqslant \epsilon^{-1}|A_{f,g}|^2(1 - |A_{f,g}|^{2\epsilon}) \leqslant -|A_{f,g}|^2 \ln|A_{f,g}|^2. \tag{5.4}$$

(The last inequality is simply $1 + \epsilon \ln X \leqslant X^\epsilon$ for all $X > 0$.) Now $K(\epsilon)$ is just the integral of the middle function in (5.4) (which is non-negative), and we see that this function is uniformly dominated by an integrable function. Furthermore, as $\epsilon \downarrow 0$ the middle function in (5.4) converges pointwise to the right-hand function. Equation (5.3) then follows by Lebesgue's dominated convergence theorem. Q.E.D.

ACKNOWLEDGMENTS

I thank E. Carlen, I. Daubechies, P. Flandrin, and A. Grossman for helpful discussions and A. J. E. M. Janssen for a helpful correspondence and for encouraging me to write this paper. In fact, the results in this paper have already been quoted and used by Janssen.[11]

This work was partially supported by U. S. National Science Foundation Grant PHY 85-15288-A03.

[1] P. M. Woodward, *Probability and Information Theory with Applications to Radar* (McGraw-Hill, New York, 1953), p. 120.

[2] R. Price and E. M. Hofstetter, "Bounds on the volume and height distributions of the ambiguity function," IEEE Trans. Inf. Theory **IT-11**, 207–214 (1965).

[3] A. J. E. M. Janssen, "Positivity properties of phase-plane distribution functions," J. Math. Phys. **25**, 2240–2252 (1984).

[4] E. H. Lieb, "Gaussian kernels have only Gaussian maximizers," to be published in Invent. Math. (1990).

[5] R. L. Hudson, "When is the Wigner quasi-probabilty density non-negative?," Rep. Math. Phys. **6**, 249–252 (1974).

[6] A. J. E. M. Janssen, "Bilinear phase-plane distribution functions and positivity", J. Math. Phys. **26**, 1986–1994 (1985).

[7] E. H. Lieb, "Proof of an entropy conjecture of Wehrl," Commun. Math. Phys. **62**, 35–41 (1978).

[8] W. Beckner, "Inequalities in Fourier analysis," Ann. Math. **102**, 159–182 (1975).

[9] H. J. Brascamp and E. H. Lieb, "Best constants in Young's inequality, its converse, and its generalization to more than three functions," Adv. Math. **20**, 151–173 (1976).

[10] L. Leindler, "On a certain converse of Hölder's inequality. II," Acta Math. Szeged. **33**, 217–223 (1972).

[11] A. J. E. M. Janssen, "Wigner weight functions and Weyl symbols of non-negative definite linear operators," Philips J. Res. **6**, 7–42 (1989).

On Hamiltonian systems in two degrees of freedom with invariants quartic in the momenta of form $p_1^2p_2^2\cdots$

N. W. Evans
School of Mathematical Sciences, Queen Mary College, Mile End Road, London E1 4NS, England and King's College, Cambridge CB2 1ST, England

(Received 5 July 1988; accepted for publication 1 November 1989)

A search is made for autonomous Hamiltonian systems in two degrees of freedom which admit a second invariant quartic in the momenta with leading term $p_1^2p_2^2/2$. A sufficient condition for the resulting functional equation to possess solutions is deduced and a family of integrable systems is identified, which under the equivalence class of linear transformations reduce to a simpler integrable system found originally by Bozis. The method of Lax pairs is used to find further solutions to the functional equation and give new classes of integrable but nonseparable Hamiltonians.

I. INTRODUCTION

The subject of this paper is autonomous Hamiltonian systems in two degrees of freedom in flat space. Taking (q_1,q_2) and (p_1,p_2) as canonical position and momenta coordinates, the Hamiltonian is

$$H=\tfrac{1}{2}p_1^2+\tfrac{1}{2}p_2^2+V(q_1,q_2). \tag{1}$$

A problem of great interest in physics is to find what forms of potential V possess an additional invariant besides the energy. The first systematic investigation was made by Darboux[1] and reproduced in Whittaker's[2] well-known book. All Hamiltonians of form (1) admitting second invariants linear or quadratic in the momenta were deduced. Such integrable systems necessarily possess separable solutions to the Hamilton–Jacobi equation.[3] By contrast, Hamiltonians with a second constant that is a polynomial in the momenta of degree greater than two correspond to integrable but possibly nonseparable systems. The most famous example is the Toda lattice,[4] with a constant of the motion that is cubic in the momenta. Recently, Holt[5] extended the methods of Darboux and Whittaker to find new Hamiltonians possessing invariants cubic in the momenta, while some n-dimensional integrable systems identified by the methods of Lax pairs have two-dimensional versions with cubic or quartic second invariants.[6] All these results are summarized in the excellent review of Hietarinta.[7]

Here, we look for second invariants quartic in the momenta of form

$$I=\Delta p_1^2p_2^2+g_0p_1^2+g_1p_1p_2+g_2p_2^2+g_3, \tag{2}$$

where Δ, g_0, g_1, g_2, and g_3 are functions of position. This ansatz was first introduced by Bozis,[8] who found a number of examples of Hamiltonians with such an invariant. The vanishing of the Poisson bracket of H and I leads to the equation

$$\begin{aligned}\sum_{i=1}^{2}p_i\left(p_1^2p_2^2\frac{\partial\Delta}{\partial q_i}+p_1^2\frac{\partial g_0}{\partial q_i}+p_1p_2\frac{\partial g_1}{\partial q_i}+p_2^2\frac{\partial g_2}{\partial q_i}+\frac{\partial g_3}{\partial q_i}\right)&=\frac{\partial V}{\partial q_1}(2\Delta p_1^2p_2^2+2g_0p_1+g_1p_2)\\&\quad+\frac{\partial V}{\partial q_2}(2\Delta p_1^2p_2^2+2g_2p_2+g_1p_1).\end{aligned} \tag{3}$$

Equating coefficients of terms of the same degree in (3), we obtain the following system:

$$\frac{\partial\Delta}{\partial q_1}=0,\quad \frac{\partial\Delta}{\partial q_2}=0, \tag{4}$$

$$\frac{\partial g_0}{\partial q_1}=0,\quad \frac{\partial g_2}{\partial q_2}=0, \tag{5}$$

$$\frac{\partial g_2}{\partial q_1}+\frac{\partial g_1}{\partial q_2}-2\Delta\frac{\partial V}{\partial q_1}=0, \tag{6}$$

$$\frac{\partial g_1}{\partial q_1}+\frac{\partial g_0}{\partial q_2}-2\Delta\frac{\partial V}{\partial q_2}=0, \tag{7}$$

$$\frac{\partial g_3}{\partial q_1}-2g_0\frac{\partial V}{\partial q_1}-g_1\frac{\partial V}{\partial q_2}=0, \tag{8}$$

$$\frac{\partial g_3}{\partial q_2}-2g_2\frac{\partial V}{\partial q_2}-g_1\frac{\partial V}{\partial q_1}=0. \tag{9}$$

From (4), Δ is found to be a constant, which without loss of generality is taken as $\frac{1}{2}$. Equations (5) integrate to give $g_0=v_2(q_2)$ and $g_2=v_1(q_1)$. Differentiating (6) with respect to q_2 and (7) with respect to q_1, then eliminating the term in $\partial^2V/\partial q_1\,\partial q_2$, we have

$$\frac{\partial^2g_1}{\partial q_2^2}-\frac{\partial^2g_1}{\partial q_1^2}=0. \tag{10}$$

This is the two-dimensional wave equation, the general solution of which is well known to be $g_1=v_4(q_1+q_2)-v_3(q_1-q_2)$. Substituting this into (6) and (7) and integrating gives

$$V=v_1(q_1)+v_2(q_2)+v_3(q_1-q_2)+v_4(q_1+q_2). \tag{11}$$

Finally, the integrability condition for g_3 yields the equation

$$\begin{aligned}g_1\left(\frac{\partial^2V}{\partial q_1^2}-\frac{\partial^2V}{\partial q_2^2}\right)+2(g_2-g_0)\frac{\partial^2V}{\partial q_1\,\partial q_2}-3\frac{\partial g_0}{\partial q_2}\frac{\partial V}{\partial q_1}&\\+3\frac{\partial g_2}{\partial q_1}\frac{\partial V}{\partial q_2}=0.&\end{aligned} \tag{12}$$

Using the derived forms for g_1, g_2, and g_3, this equation can be recast as

$$\begin{aligned}&[v_4(q_1+q_2)-v_3(q_1-q_2)][v_2''(q_2)-v_1''(q_1)]\\&\quad+2[v_4''(q_1+q_2)-v_3''(q_1-q_2)][v_2(q_2)-v_1(q_1)]\end{aligned}$$

$$+3v_4'(q_1+q_2)[v_2'(q_2)-v_1'(q_1)]+3v_3'(q_1-q_2)$$
$$\times[v_2'(q_2)+v_1'(q_1)]=0, \qquad (13)$$

where primes indicate differentiation with respect to the argument. This functional equation is stated in Hietarinta.[7]

II. THE METHOD OF DIRECT SEARCH

The general solution to the functional equation (13) is not known. It is possible to deduce a sufficient condition for solutions to exist by casting (13) into the form

$$\frac{\partial}{\partial q_1}\left(\frac{1}{v_4(q_1+q_2)-v_3(q_1-q_2)}\frac{\partial}{\partial q_1}([v_4(q_1+q_2) - v_3(q_1-q_2)]^2 v_1(q_1))\right)$$
$$=\frac{\partial}{\partial q_2}\left(\frac{1}{v_4(q_1+q_2)-v_3(q_1-q_2)} \times\frac{\partial}{\partial q_2}([v_4(q_1+q_2)-v_3(q_1-q_2)]^2 v_2(q_2))\right). \qquad (14)$$

This demonstrates that solutions with the structure $v_1 = v_1(q_1)$ and $v_2=v_2(q_2)$ can exist if

$$v_4(q_1+q_2)-v_3(q_1-q_2)=f_1(q_1)f_2(q_2), \qquad (15)$$

as the equation is then separable. It must be stressed that this is a sufficient but *not* necessary condition, as can be seen from the solutions given in Sec. III found by the method of Lax pairs.

Supposing (15) to be satisfied, (14) can be separated to give

$$\frac{1}{f_1}\frac{d}{dq_1}\left(\frac{1}{f_1}\frac{d}{dq_1}(f_1^2 v_1)\right)=\frac{1}{f_2}\frac{d}{dq_2}\left(\frac{1}{f_2}\frac{d}{dq_2}(f_2^2 v_2)\right)=k, \qquad (16)$$

where k is the separation constant. Solving for v_1 and v_2, we obtain

$$v_1(q_1)=\frac{k_1}{f_1^2}\int^{q_1} f_1(u)du+\frac{k}{2f_1^2}\left[\int^{q_1} f_1(u)du\right]^2+\frac{k_2}{f_1^2}, \qquad (17)$$

$$v_2(q_2)=\frac{k_3}{f_2^2}\int^{q_2} f_2(u)du+\frac{k}{2f_2^2}\left[\int^{q_2} f_2(u)du\right]^2+\frac{k_4}{f_2^2}, \qquad (18)$$

where k_1, k_2, k_3, and k_4 are all constants. Finally, substituting (17) and (18) into (8) and (9) gives g_3 as

$$g_3=\tfrac{1}{2}[v_4(q_1+q_2)-v_3(q_1-q_2)]^2+2v_1(q_1)v_2(q_2) + k\int^{q_1} f_1(u)du\int^{q_2} f_2(u)du+k_3\int^{q_1} f_1(u)du + k_1\int^{q_2} f_2(u)du. \qquad (19)$$

It remains to deduce the solutions to (15) for f_1 and f_2. Applying the two-dimensional wave operator annihilates the left-hand side to give

$$f_2(q_2)f_1''(q_1)-f_1(q_1)f_2''(q_2)=0. \qquad (20)$$

This is easily solved; there are only two possibilities, namely,

$$f_1=Aq_1+B,\quad f_2=Cq_2+D, \qquad (21)$$

or

$$f_1=A\exp(lq_1)-B\exp(lq_1),$$
$$f_2=C\exp(lq_2)-D\exp(lq_2), \qquad (22)$$

where A, B, C, D, and l are arbitrary (possibly complex) constants. In fact, (21) leads to the super-integrable potential[9]

$$V=a(q_1^2+q_2^2)+b/q_1^2+c/q_2^2. \qquad (23)$$

This is obviously separable in Cartesian, elliptic, and plane polar coordinates. The system has three functionally independent isolating quadratic integrals, from which the quartic invariant is constructed.

From the standpoint of this paper, (22) is more interesting as it leads to a family of integrable systems that do not separate in any coordinate system. The general solution has the form

$$V=v_1(q_1)+v_2(q_2)+v_3(q_1-q_2)+v_4(q_1+q_2), \qquad (24)$$

where

$$v_1(z)=[k_1'(A\exp(lz)+B\exp(-lz))+k'(A\exp(lz) + B\exp(-lz))^2+k_2]/[A\exp(lz) - B\exp(lz)]^2,$$
$$v_2(z)=[k_3'(C\exp(lz)+D\exp(-lz))+k'(C\exp(lz) + D\exp(-lz))^2+k_4]/[C\exp(lz) - D\exp(-lz)]^2,$$
$$v_3(z)=AD\exp(lz)+BC\exp(-lz),$$
$$v_4(z)=AC\exp(lz)+BD\exp(-lz). \qquad (25)$$

Integrability is preserved under scaling, rotational, and translational transformations and this can be exploited to simplify (25). The constant k' may be eliminated by subtracting $2k'$ from the potential and redefining k_2 and k_4. A real translation in q_1 and q_2 can be used to obtain $|A|=|B|$ and $|C|=|D|$. Finally, a complex translation can be performed to reduce (25) to

$$v_1(z)=[a+b\sin(lz)]/\cos^2(lz),$$
$$v_2(z)=[c+d\sin(lz)]/\cos^2(lz),$$
$$v_3(z)=-f\cos(lz),$$
$$v_4(z)=f\cos(lz), \qquad (26)$$

where a, b, c, d, f, and l are arbitrary complex constants. The second invariant is given by

$$I=\frac{1}{2}p_1^2p_2^2+\frac{[c+d\sin(lq_2)]}{\cos^2(lq_2)}p_1^2+\frac{[a+b\sin(lq_1)]}{\cos^2(lq_1)}p_2^2 +\frac{2[a+b\sin(lq_1)][c+d\sin(lq_2)]}{\cos^2(lq_1)\cos^2(lq_2)}+2f\cos(lq_1) \times\cos(lq_2)p_1p_2+2fd\sin(lq_1)+2fb\sin(lq_2) +2f^2\cos^2(lq_1)\cos^2(lq_2). \qquad (27)$$

In the form (26), the potential was shown to be integrable by Bozis.[8]

There are four coordinate systems in which the two-dimensional Hamilton–Jacobi equation separates, namely

Cartesian, plane polar, parabolic, or elliptic coordinates. To show (26) does not separate in any coordinate system in any Euclidean frame, it is necessary and sufficient to show that there does not exist an integral quadratic in the canonical momenta.[3] Let us assume then an integral J exists of the form

$$J = h_1 p_1^2 + h_2 p_2^2 + h_3 p_1 p_2 + h_4, \tag{28}$$

where the h_i are all functions of position. Equating coefficients in the vanishing Poisson bracket of H and J leads to the system of coupled partial differential equations:

$$\frac{\partial h_1}{\partial q_1} = 0, \quad \frac{\partial h_2}{\partial q_2} = 0, \quad \frac{\partial h_2}{\partial q_1} + \frac{\partial h_3}{\partial q_2} = 0, \quad \frac{\partial h_1}{\partial q_2} + \frac{\partial h_3}{\partial q_1} = 0, \tag{29}$$

$$\frac{\partial h_4}{\partial q_1} - 2h_1[v_1'(q_1) + v_3'(q_1 - q_2) + v_4'(q_1 + q_2)] - h_3[v_2'(q_2) - v_3'(q_1 - q_2) + v_4'(q_1 + q_2)] = 0, \tag{30}$$

$$\frac{\partial h_4}{\partial q_2} - 2h_2[v_2'(q_2) - v_3'(q_1 - q_2) + v_4'(q_1 + q_2)] - h_3[v_1'(q_1) + v_3'(q_1 - q_2) + v_4'(q_1 + q_2)] = 0. \tag{31}$$

It is easy to integrate the set of equations (29) to obtain solutions taken without loss of generality as

$$h_1 = q_2^2 + k_1 q_2 + k_2, \quad h_2 = q_1^2 + k_3 q_1 + k_4, \quad h_3 = -(2q_1 q_2 + k_1 q_1 + k_3 q_2 + k_5), \tag{32}$$

where the k_i are all constants. The integrability condition for h_4 yields the equation

$$\begin{aligned} &2((q_2^2 - q_1^2) + k_1 q_2 - k_3 q_1 + k_2 - k_4)[v_4''(q_1 + q_2) - v_3''(q_1 - q_2)] - 2(q_1 q_2 + k_1 q_1 + k_3 q_2 + k_5)[v_2''(q_2) - v_1''(q_1)] \\ &\quad = 3(2q_1 + k_3)[v_2'(q_2) - v_3'(q_1 - q_2) + v_4'(q_1 + q_2)] - 3(2q_2 + k_1)[v_1'(q_1) + v_3'(q_1 - q_2) + v_4'(q_1 + q_2)], \end{aligned} \tag{33}$$

which must hold everywhere. Now v_1, v_2, v_3, and v_4 given by (26) may be expanded as Taylor polynomials in the neighborhood of the origin. By equating coefficients of powers of q_1 and q_2, it is straightforward to show that the k_i must all vanish, excepting the trivial cases $v_3 = v_4 = 0$ or $v_1 = v_2 = 0$. So, there is no invariant quadratic in the canonical momenta and no separable solution to the Hamilton–Jacobi equation.

III. THE METHOD OF LAX PAIRS

In this section, further solutions to (13) that do not satisfy the sufficient condition (15) are constructed by the method of Lax pairs. This was originally devised as a means of representing the Kortweg–de Vries equation.[10] It was first applied to a finite-dimensional Hamiltonian system by Flachska[11] for the Toda lattice. Shortly afterwards, Moser[12] used the method to demonstrate the integrability of the Calogero system. A Lax pair comprises two $m \times m$ matrices L and M that define the equations of motion through

$$\frac{dL}{dt} = [L, M]. \tag{34}$$

Although L is t dependent, the spectrum of L is not. The invariants of the dynamical system can be taken as $\mathrm{Tr}(L^k)$, for $k = 0, 1, \ldots$.

Inozemtsev[13] constructed a number of integrable systems with invariants of form (2) by choosing the p-dependency of the Lax matrix L to be of form $\mathrm{diag}(\mathbf{p}, -\mathbf{p})$. Let us modify his ansatz and use the 4×4 matrices

$$L_1 = \begin{pmatrix} p_1 & R(q_1 - q_2) & -P(q_1) & S(q_1 + q_2) \\ R(q_1 - q_2) & p_2 & iS(q_1 + q_2) & iQ(q_2) \\ -P(q_1) & -iS(q_1 + q_2) & -p_1 & -iR(q_1 - q_2) \\ S(q_1 + q_2) & -iQ(q_1) & iR(q_1 - q_2) & -p_2 \end{pmatrix}, \tag{35}$$

$$M_1 = \begin{pmatrix} 0 & R'(q_1 - q_2) & -P'(q_1)/2 & S'(q_1 + q_2) \\ -R'(q_1 - q_2) & 0 & iS'(q_1 + q_2) & iQ'(q_2)/2 \\ P'(q_1)/2 & iS'(q_1 + q_2) & 0 & iR'(q_1 - q_2) \\ -S'(q_1 + q_2) & iQ'(q_2)/2 & iR'(q_1 - q_2) & 0 \end{pmatrix}, \tag{36}$$

where the functions P, Q, R, and S are connected with the potentials v_1, v_2, v_3, and v_4 by

$$v_1(z) = \tfrac{1}{2}P^2(z), \quad v_2(z) = \tfrac{1}{2}Q^2(z), \quad v_3(z) = R^2(z), \quad v_4(z) = S^2(z). \tag{37}$$

The matrices L_1 and M_1 are a Lax pair if P, Q, R, and S satisfy the functional equations

$$\frac{R'(q_1 - q_2)}{R(q_1 - q_2)} = \frac{Q'(q_2) - P'(q_1)}{2(P(q_1) + Q(q_2))}, \tag{38}$$

$$\frac{S'(q_1+q_2)}{S(q_1+q_2)} = \frac{Q'(q_2)-P'(q_1)}{2(P(q_1)-Q(q_2))}. \tag{39}$$

By applying the operator $\partial/\partial q_1 + \partial/\partial q_2$ to (38) and $\partial/\partial q_1 - \partial/\partial q_2$ to (39), we find that

$$\frac{P''(q_1)}{P(q_1)} = \frac{Q''(q_2)}{Q(q_2)} = \text{const.} \tag{40}$$

From this, it follows that there are only two distinct solutions to (38) and (39), namely,

$$P(z) = Az + B,\ Q(z) = -Az + D,$$
$$R(z) = E/(Az+B+D),\ S(z) = F/(Az+B-D), \tag{41}$$

or

$$P(z) = A\cosh(Bz+C), \quad Q(z) = A\cosh(Bz+D),$$
$$R(z) = \frac{E}{\cosh(\frac{1}{2}[Bz+C-D])},$$
$$S(z) = \frac{F}{\sinh(\frac{1}{2}[Bz+C+D])} \tag{42}$$

Here, A, B, C, D, E, and F are arbitrary complex constants. The solutions may be extended by defining the 8×8 Lax matrices

$$L = \begin{pmatrix} L_1 & L_2 \\ L_2 & -L_1 \end{pmatrix}, \tag{43}$$

$$M = \begin{pmatrix} M_1 & M_2 \\ -M_2 & M_1 \end{pmatrix}, \tag{44}$$

where

$$L_2 = \mathrm{diag}(\widetilde{P}(q_1), \widetilde{Q}(q_2), \widetilde{P}(q_1), \widetilde{Q}(q_2)), \tag{45}$$

$$M_2 = \tfrac{1}{2}\,\mathrm{diag}(\widetilde{P}'(q_1), \widetilde{Q}'(q_2), -\widetilde{P}'(q_1), -\widetilde{Q}'(q_2)). \tag{46}$$

The functions $\widetilde{P}$ and $\widetilde{Q}$ are given in terms of the potentials v_1 and v_2 by

$$v_1(z) = \tfrac{1}{2}(P^2(z) + \widetilde{P}^2(z)), \quad v_2(z) = \tfrac{1}{2}(Q^2(z) + \widetilde{Q}^2(z)). \tag{47}$$

The matrices L and M satisfy the Lax equation (34) if

$$\frac{R'(q_1-q_2)}{R(q_1-q_2)} = \frac{\widetilde{P}'(q_1)+\widetilde{Q}'(q_2)}{2(\widetilde{Q}(q_2)-\widetilde{P}(q_1))}, \tag{48}$$

$$\frac{S'(q_1+q_2)}{S(q_1+q_2)} = \frac{\widetilde{Q}'(q_2)-\widetilde{P}'(q_1)}{2(\widetilde{P}(q_1)-\widetilde{Q}(q_2))}. \tag{49}$$

The distinct solutions of (48) and (49) are

$$\widetilde{P}(z) = (Az+B)^2 + G, \quad \widetilde{Q}(z) = (Az-D)^2 + G, \tag{50}$$

and

$$\widetilde{P}(z) = G\sinh(Bz+C) + H,$$
$$\widetilde{Q}(z) = -G\sinh(Bz+D) + H. \tag{51}$$

The Hamiltonian is $\mathrm{Tr}(L^2)/8$ while the quartic invariant is conveniently taken as

$$I = \tfrac{1}{64}[\mathrm{Tr}(L^2)]^2 - \tfrac{1}{16}\mathrm{Tr}(L^4), \tag{52}$$

which is calculated to be

$$I = \tfrac{1}{2}p_1^2p_2^2 + v_2(q_2)p_1^2 + v_1(q_1)p_2^2 + (v_4(q_1+q_2)$$
$$- v_3(q_1-q_2))p_1p_2 + 2v_1(q_1)v_2(q_2) + v_4(q_1+q_2)$$
$$\times[\widetilde{P}(q_1)\widetilde{Q}(q_2) + P(q_1)Q(q_2)] + v_3(q_1-q_2)$$
$$\times[\widetilde{P}(q_1)\widetilde{Q}(q_2) - P(q_1)Q(q_2)] + \tfrac{1}{2}(v_4(q_1+q_2)$$
$$- v_3(q_1-q_2))^2. \tag{53}$$

Simplifying by use of linear transformations, the integrable systems may be taken without loss of generality to be

$$V = v_1(q_1) + v_2(q_2) + v_3(q_1-q_2) + v_4(q_1+q_2), \tag{54}$$

where

$$v_1(z) = v_2(z) = z^4 + az^2,$$
$$v_3(z) = b/z^2, \tag{55}$$
$$v_4(z) = c/z^2,$$

or

$$v_1(z) = a\cosh(4lz) + b\sinh(2lz),$$
$$v_2(z) = a\cosh(4lz) - b\sinh(2lz),$$
$$v_3(z) = c/\cosh^2(lz), \tag{56}$$
$$v_4(z) = d/\sinh^2(lz).$$

Here, a, b, c, d, and l are arbitrary complex constants. It is straightforward to show that (55) and (56) cannot satisfy (33) and do not separate in any coordinate system. They are related (but are inequivalent) to the systems identified by Inozemtsev[13] and furnish us with further examples of integrable but nonseparable potentials.

A systematic and complete search for finite-dimensional integrable Hamiltonians with invariants quartic in the momenta is important for two reasons. First, as integrability is rare, it is useful to construct lists of all known cases (under the equivalence class of linear transformations). Second, the theory of invariants quadratic in the momenta is very complete; all such invariants arise from separability of the Hamilton–Jacobi equation in the elliptic coordinates and their degenerations. It would be valuable to generalize this classification theory to higher degree polynomial invariants. So, a point of significance is to find every solution to the functional equation (13). To extend the work of this paper, the necessary and sufficient conditions for (13) to possess solutions must be discovered. The method of Lax pairs seems to offer the most promising way of constructing further solutions and perhaps classifying them.

ACKNOWLEDGMENTS

The author acknowledges financial support from the Science and Engineering Research Council of the U. K. and King's College, Cambridge.

It is a pleasure to thank D. Lynden-Bell and J. L. Collett for valuable discussions. The suggestions and comments of the (anonymous) referee greatly improved the presentation and contents of the paper.

[1]G. Darboux, Arch. Neerland. (ii) **6**, 371 (1901).
[2]E. T. Whittaker, *A Treatise on the Analytical Dynamics of Particles and Rigid Bodies* (Cambridge U. P., London, 1927), p. 325.
[3]P. Winternitz, Ya. A. Smorodinsky, M. Uhlir, and I. Fris, Sov. J. Nucl. Phys. **4**, 625 (1966).
[4]M. Toda, J. Phys. Soc. Jpn. **23**, 501 (1967); J. Ford, S. P. Stoddard, and J. S. Turner, Prog. Theor. Phys. **50**, 1547 (1973).
[5]C. Holt, J. Math. Phys. **23**, 1037 (1982).
[6]F. Calogero, Lett. Nuovo Cimento **13**, 383 (1975); M. A. Olshanetsky and A. M. Perelomov, Invent. Math. **37**, 93 (1976).
[7]J. Hietarinta, Phys. Rev. **147**, 87 (1987).

[8]G. Bozis, Cel. Mech. **28**, 367 (1982).
[9]The first systematic investigation into super-integrability in two degrees of freedom seems to have been given in J. Fris, V. Mandrosov, Ya. A. Smorodinsky, M. Uhlir, and P. Winternitz, Phys. Lett. **16**, 354 (1965). Subsequently, this was extended to three degrees of freedom in A. A. Makarov, Ya. A. Smorodinsky, K. Valiev, and P. Winternitz, Nuovo Cimento **52**, 1061 (1967), where all potentials separable in spherical polar and one other coordinate system were found. The remaining possibilities remained uninvestigated and were given recently in N. W. Evans, "Separability and Integrability in Stellar Dynamics," Ph.D. thesis, Cambridge University, 1988.
[10]P. D. Lax, Commun. Pure Appl. Math. **21**, 467 (1968).
[11]H. Flashka, Phys. Rev. B **9**, 1924 (1974). There is a review of the method of Lax pairs in J. Moser, *Various Aspects of Integrable Hamiltonian Systems*, Proceedings of C. I. M. E. Conference, Bressanone, Italy (1978); Prog. Math. **8**, 251 (1980).
[12]J. Moser, Adv. Math. **16**, 197 (1975).
[13]V. I. Inozemtsev, J. Phys. A: Math. Gen. **17**, 815 (1984); V. I. Inozemtsev, Phys. Scr. **29**, 518 (1984); V. I. Inozemtsev and D. V. Meshcheryakov, Lett. Math. Phys. **9**, 13 (1985).

On the Painlevé classification of partial differential equations

Ladislav Hlavatý
Institute of Physics, Czechoslovak Academy of Sciences, Na Slovance 2, 180 40 Prague 8, Czechoslovakia

(Received 1 December 1988; accepted for publication 1 November 1989)

The classification of partial differential equations from the point of view of the singular point analysis is suggested. The general form of the equation for the Painlevé resonances is derived. The Painlevé-type classification of semilinear second-order polynomial partial differential equations in two independent variables is performed.

I. INTRODUCTION

The interest in the Painlevé equations was revived a few years ago due to the ARS conjecture[1] that relates the integrability of partial differential equations (PDEs) to the Painlevé property of the similarity reductions of the equations.

The singular point analysis[2,3] was developed for testing the Painlevé property and many equations were tested by this procedure. The physically relevant equations that are usually autonomous, i.e., without explicit dependence on the independent variables, were investigated first. Recently their nonautonomous counterparts became the focus of attention,[4–6] but in a rather nonsystematic way, namely introducing coefficients dependent on the independent variables into the integrable equations.

In my previous paper[7] I have considered the possibility of using the singular point analysis for the classification of Painlevé-type ODEs. Extension of this method to the PDEs provides us with a tool for the systematic search for the (nonautonomous) PDEs with the Painlevé property. In this paper the extension is presented and applied to the Painlevé-type classification of the (first degree) second-order polynomial PDEs in two independent variables.

The definition as well as the investigation of the Painlevé property for PDEs is rather complicated.[8] Nevertheless, we can investigate which equations pass the Painlevé test defined in Ref. 3 and thus are good candidates on integrable PDEs.

The test checks whether a general solution of the considered PDE in the d-dependent variables $z_1,z_2,\dots,z_d$ has the expansion

$$u(z_1,z_2,\dots,z_d) = \sum_{m=0}^{\infty} u_n(z_1,z_2,\dots,z_d)F(z_1,z_2,\dots,z_d)^{n+p} \qquad (1.1)$$

in a neighborhood of a movable noncharacteristic singular manifold given by the equation $F(z_1,z_2,\dots,z_d)=0$. Both F and u_n are assumed to be analytic functions of $z_1,z_2,\dots,z_d$.

Inserting the expansion (1.1) into the investigated equation yields a recurrent formula that determines $u_n(z_1,z_2,\dots,z_d)$ for all $n\geqslant 0$ except for a finite number of $r_1,\dots,r_K\geqslant 0$ called resonances. One must check that the number of resonances is sufficient, i.e., $K=M-1$ for an equation of the Mth order, and that the recurrent formula for the resonances does not impose supplementary conditions on the arbitrary functions F and u_{r_i}. These two conditions are required in order that (1.1) gives the expansion of the general solution. The equations that pass this test will be called Painlevé admissible.

The classification scheme suggested in Ref. 7 follows the basic steps of this test. In the first step the types of equations that admit the expansion (1.1) with negative integer p are determined. Afterwards, the resonance analysis and the compatibility conditions of the recursion formula for u_n, the coefficients of the expansion (1.1), are used to specify the coefficients of these equations. The general scheme is described in Sec. II and in the next sections it is applied to polynomial second-order PDEs in two independent variables.

It would be useful to extend the classification to the rational equations. The number of types of equations is much bigger in this case (cf. Ref. 9) and there are some special problems connected with the finite point analysis, nevertheless I hope to publish results on the rational equations in the future.

II. THE DOMINANT TRUNCATIONS AND THE RESONANCE FORMULAE

The generalization of concepts and formulae used in Ref. 7 to the PDEs is rather straightforward. In this paper we restrict ourselves to PDEs in two independent variables $z_1=x$, $z_2=y$. We consider the PDEs that are real for real x,y even though the variables are considered complex during the analysis. The extension to more independent variables is self-evident. Similarly, the formulas can be extended to more dependent variables $u_1,u_2,\dots,u_N$. However, in this paper we shall deal only with $N=1$.

Every polynomial or even rational PDE of order N can be written in the form

$$E(\mathscr{K},g):=\sum_{K\in\mathscr{K}} g_K(x,y)[u]^K(x,y)=0, \qquad (2.1)$$

where $u=u(x,y)\in\mathbb{C}$ denotes the dependent variable, $\mathscr{K}$ is a set of multiindices

$$K:=(k_{00};k_{10},k_{01};k_{20},k_{11},k_{02};\dots;k_{N0},k_{N-1,1},\dots,k_{0N}), \qquad (2.2)$$

$$[u]^K:=\prod_{\substack{0<i,j\\ i+j<N}} u_{ij}^{\,k_{ij}}, \qquad (2.3)$$

$$u_{nm}:=u_{nx,my}:=\partial_x^n\partial_y^m u(x,y),\quad n,m\in N_0 \qquad (2.4)$$

 0022-2488/90/030605-05$03.00

(N_0 denotes the natural numbers including zero), and g_K are analytic functions.

Example: The set $\mathscr{E}$ for the Burgers equation

$$u_y + u_{xx} + u_x u = 0$$

is

$$\mathscr{E} = \{(0;0,1;0,0,0),\ (0;0,0;1,0,0),\ (1;1,0;0,0,0)\},$$

$g_K = 1$ for all K.

For investigation of the leading order terms of Eqs. (2.1) it is useful to introduce the concept of the so-called p dominance[7] (p integer) of the term $[u]^K$. For PDEs in two variables it is defined as

$$D(p,K) := \sum_{\substack{i,j \geqslant 0 \\ i+j \leqslant N(K)}} (p-i-j)k_{ij}, \tag{2.5}$$

where $N(K)$ is the order of the term $[u]^K$. Actually, $D(p,k)$ is the power of F occurring in $[u]^K$ when the leading order behavior F^p is assumed for u.

For a given p, the dominant truncation of the equation (2.1) is

$$T(p,\mathscr{E},g) := \sum_{\substack{K\in\mathscr{E}, \\ D(p,K) = \mu(p,\mathscr{E})}} g_K(x,y)\,[u]^K(x,y), \tag{2.6}$$

where $\mu(p,\mathscr{E})$ is the p dominance of Eq. (2.1) defined as

$$\mu(p,\mathscr{E}) := \min_{K\in\mathscr{E}} D(p,K). \tag{2.7}$$

The just defined dominant truncation is that part of the equation that for a given p contributes to the leading terms in F as $F\to 0$.

The general form of the dominant truncations with the p-dominance equal to m is

$$T(p,m,g) := \sum_{K\in M(p,m)} g_K(x,y)\,[u]^K(x,y), \tag{2.8}$$

where $M(p,m)$ denotes the set of all K (of *a priori* unspecified order) with the p dominance equal to m. The dominant truncations with $p<0$ containing terms of different orders are the most important for the singular point analysis. Several simplest ones are displayed in Table I.

Next we are going to derive the formulas necessary for the resonance analysis of $T(p,m,g)$. It consists[2,3] in the substitution of

$$u(x,y) = a(x,y)F(x,y)^p[1 + b(x,y)F(x,y)^r] \tag{2.9}$$

into (2.8) and collecting terms up to the first order in b. By this way we get

$$T(p,m,g) \approx A(p,m,g,a)F^m + bR(p,m,g,a,r)F^{m+r}, \tag{2.10}$$

where

$$A(p,m,g,a) := \sum_{K\in M(p,m)} g_K[p]^K a^{d(K)} F_x^{I(K)} F_y^{J(K)}, \tag{2.11}$$

$$R(p,m,g,a,r) := \sum_{K\in M(p,m)} g_K[p]^K a^{d(K)} F_x^{I(K)} F_y^{J(K)}\,Y(p,K,r), \tag{2.12}$$

and

$$d(K) := \sum_{i,j} k_{ij} \tag{2.13}$$

represents the degree of $[u]^K$,

$$I(K) := \sum_{i,j} ik_{ij}, \tag{2.14}$$

$$J(K) := \sum_{i,j} jk_{ij}, \tag{2.15}$$

$$[p]^K := \prod_{i,j} [p]_{i+j}{}^{k_{ij}}, \tag{2.16}$$

$$[p]_0 := 1, \quad [p]_j := p(p-1)\cdots(p-j+1), \quad j\in N, \tag{2.17}$$

$$Y(p,K,r) := \sum_{i,j} k_{ij}[r+p]_{i+j}/[p]_{i+j}. \tag{2.18}$$

The indices in the sums and the product (2.13)–(2.18) run over $i,j \geqslant 0$, $i+j \leqslant N(K)$, the order of the term $[u]^K$.

The equation for the leading order term $u_0(x,y)$ of the solution expansion (1.1) is then

$$A(p,m,g(x,y),u_0(x,y)) = 0, \tag{2.19}$$

and the equation for the values of the resonances r is

$$R(p,m,g(x,y),u_0(x,y),r) = 0, \tag{2.20}$$

where u_0 is a solution of (2.19).

Similarly as in the case of ODEs, one can prove that[7]

$$R(p,m,g,a,r=-1) = mA(p,m,g,a)/p, \tag{2.21}$$

so that $r=-1$ is always one of the resonances.

TABLE I. Dominant truncations with p dominances equal to m. The coefficients g_K denoted here as A,B,C,... are functions of x and y.

p	m	$T(p,m,g)$
-1	-2	$Au_x + Bu_y + Cu^2$,
-1	-3	$Au_{xx} + Bu_{xy} + Cu_{yy} + Du_x u + Eu_y u + Gu^3$,
-1	-4	$Au_{xxx} + Bu_{xxy} + Cu_{xyy} + Du_{yyy} + Eu_{xx}u + Gu_{xy}u + Hu_{yy}u + Lu_x^2 + Mu_x u_y + Nu_y^2 + Pu_x u^2 + Qu_y u^2 + Ru^4$,
-2	-4	$Au_{xx} + Bu_{xy} + Cu_{yy} + Du^2$,
-2	-5	$Au_{xxx} + Bu_{xxy} + Cu_{xyy} + Du_{yyy} + Eu_x u + Gu_y u$,
-3	-6	$Au_{xxx} + Bu_{xxy} + Cu_{xyy} + Du_{yyy} + Eu^2$.

III. THE CLASSIFICATION OF THE DOMINANT PARTS OF THE SECOND-ORDER PDEs

In this section we shall apply the formalism described in the preceding section to the investigation of the semilinear second-order polynomial PDEs.

Namely, we are going to deal with equations of the form

$$Au_{xx} + Bu_{xy} + Cu_{yy} + P(u_x,u_y,u,x,y) = 0, \tag{3.1}$$

where P is polynomial in u_x,u_y,u and the functions P, A, B, C, are analytic in x,y. Let us explore which equations of the form (3.1) are Painlevé admissible.

The dominant truncations of (3.1) must contain the terms with second derivatives in order that the expansion (1.1) may contain two arbitrary functions necessary for representation of a general solution of the PDE in the vicinity of the singularity. From Table I it is obvious that the only possible dominant truncations are $T(-1,-3)$ and $T(-2,-4)$. Thus the Painlevé admissible Eqs. (3.1) must be of the form

$$Au_{xx}+Bu_{xy}+Cu_{yy}+Du_xu+Eu_yu + Lu_y+Mu_x+Gu^3+Qu^2+Pu+S=0, \tag{3.2}$$

where $A,B,...,S$ are analytic functions of x and y (some of them can be zero).

As the next step we must perform the resonance analysis of the dominant truncations $T(-1,-3)$ and $T(-2,-4)$. Applying the formulas (2.11)–(2.20) from the preceding section to $T(-1,-3)$ we get the equations for the leading order coefficient $u_0(x,y)$ and resonances

$$2Z-Wu_0+Gu_0^2=0, \tag{3.3}$$

$$Z(r-1)(r-2)+W(r-2)u_0+3Gu_0^2=0, \tag{3.4}$$

where

$$Z=Z(x,y):=AF_x^2,+BF_xF_y+CF_y^2, \tag{3.5}$$

$$W=W(x,y):=DF_x+EF_y. \tag{3.6}$$

Equations (3.3) and (3.4) are nearly identical with those for the second order ODEs so we can use the results of Ref. 7 that imply

$$(2-r)W^2(x,y)=(4-r)^2Z(x,y)G(x,y), \quad r=1,2,3,4. \tag{3.7}$$

Indeed, as explained in Ref. 7 [Eqs. (4.1)–(4.10)] the only admissible resonances are -2, 1, 2, 3, 4, 6 and the pairs $(-2,1)$ and (3,6) yield the same relations.

It follows from (3.7) that there are just four types of Painlevé admissible equations with the dominant truncation $T(-1,-3)$ having the following leading order coefficients and resonances.

Type b: $G=0$, $E\neq 0$, or $D\neq 0$.

$$u_0(x,y)=2Z/W, \quad r=-1,2. \tag{3.8}$$

Type c: $G\neq 0$, $A=D^2/(9G)$, $B=2DE/(9G)$, $C=E^2/(9G)$.

$$u_0(x,y)=W/(3G), \quad r=-1,1, \tag{3.9}$$

$$u_0(x,y)=2W/(3G), \quad r=-1,-2. \tag{3.10}$$

Type d: $G\neq 0$, $A=-D^2/G$, $B=-2DE/G$, $C=-E^2/G$.

$$u_0(x,y)=-W/G, \quad r=-1,3, \tag{3.11}$$

$$u_0(x,y)=2W/G, \quad r=-1,6. \tag{3.12}$$

Type e: $G\neq 0$, $D=E=0$,

$$u_0(x,y)^2=-2Z/G, \quad r=-1,4. \tag{3.13}$$

These types are PDE analogs of the Painlevé subcases i(b)–i(e) in Ref. 10.

Investigating similarly the dominant truncation $T(-2,-4)$ we get the analog of i(a)

Type a: $G=D=E=0$, $Q\neq 0$,

$$u_0(x,y)=-6Z/Q, \quad r=-1,6. \tag{3.14}$$

IV. THE CLASSIFICATION OF REMAINING NONDOMINANT PARTS OF THE SECOND-ORDER PDEs

The terms of an investigated equation that do not belong to its dominant part form the so-called recessive part of the equation. They are the terms with higher p dominance (i.e., lower in absolute value because the p dominance is negative for $p<0$). In the previous section we have determined the Painlevé admissible dominant truncations of (3.1). The goal of this section is to specify the corresponding recessive parts. The tool for that will be the compatibility conditions that follow from the recursion formula for the coefficients of the expansion (1.1).

By the transformation of independent variables

$$X=\phi(x,y), \quad Y=\psi(x,y) \tag{4.1}$$

Eq. (3.2) can always be (locally) transformed to either parabolic or hyperbolic form (as $x,y\in\mathbb{C}$ there is no difference between hyperbolic and elliptic equations)

$$u_{xx}=P(u_x,u_y,u,x,y) \quad \text{or} \quad u_{xy}=P(u_x,u_y,u,x,y). \tag{4.2}$$

Besides we can use the transformation of the function u

$$U(x,y)=\lambda(x,y)u(x,y)+\rho(x,y) \tag{4.3}$$

for further simplification of the investigated equations.

Below we will assume that $A=0$ or -1, $B=-1$ or 0, and G is a constant. [For the type a $G=0$ and $Q(x,y)=\text{const.}$]

We are going to use the compatibility condition for analysis of the various types of equations obtained in the preceding section. We shall use the Kruskal ansatz

$$F(x,y)=x+f(y). \tag{4.4}$$

This simplifies the calculations substantially because then we can consider the coefficients u_n in (1.1) as functions only of y. On the other hand, we must be aware that the coefficients of Eq. (3.2) may depend on x and y so that they must also be expanded in the vicinity of $F(x,y)=0$. For example,

$$D(x,y)=\sum_{n=0}^{\infty}D_n(y)(x+f(y))^n. \tag{4.5}$$

This way we get the recursion formula for $n=1,2,...$:

$$\begin{aligned}-u_n(n+1)(n-R)(A+Bf_y) &= Bu_{n-1,y}(n+p-1)+G\sum_{i+j+k=n}u_iu_ju_k+\sum_{i+j+k=n-p-2}Q_iu_ju_k \\ &+\sum_{i+j+k=n-1}E_iu_ju_{k,y}+\sum_{r+j+k=n}(D_r+E_rf_y)u_ju_k(k+p) \\ &+\sum_{r+j=n-2}L_ru_{j,y}+\sum_{r+j=n-1}(L_rf_y+M_r)u_j(j+p)+\sum_{r+j=n-2}P_ru_j+S_{n-2+p},\end{aligned} \tag{4.6}$$

where $p = -2$ for the type a, $p = -1$ for the types b, c, d, e, and R is the corresponding resonance value different from -1. The summation indices satisfy $0<i,j,k<n-1$, $0<r<n$ and the conditions under the summation signs. The coefficients u_i, Q_i, D_i, E_i, L_i, M_i, P_i, S_i are functions of y.

For $n = R$ this formula becomes a compatibility condition that must be satisfied for arbitrary F in order that (3.2) may be Painlevé admissible. This condition puts further restrictions on the coefficients of Eq. (3.2).

A. Type a equations

They are the equations with the dominant truncation $T(-2,-4)$. The transformations (4.1) and (4.3) can be transformed to the form

$$u_{xz} = 6u^2 + Lu_y + Mu_x + S, \quad \text{where } z = x \text{ or } y. \tag{4.7}$$

Let us analyze the parabolic case $z = x$ first. From the recurrence formula one gets

$$u_n(n+1)(n-6) = k_n f_y^n L^n + \text{terms with smaller powers of } f_y, \quad n = 1,2,..., \tag{4.8}$$

where $k_1 = -2$, $k_2 = 1/25,...$, $k_6 = 72/75000,...$. As $f(y)$ is assumed to be arbitrary the compatibility condition for $n = 6$ requires that $L = 0$. Thus any Painlevé admissible equation of this type can be transformed to an ODE (with the Painlevé property).

Similarly, for the hyperbolic case the recurrence formula yields

$$u_1 = (-f_{yy}f_y^{-1} + L_0 f_y + M_0)/5, \tag{4.9}$$

$$u_2 = f_{yyy}/(60f_y^2) + \text{terms independent of } f_{yyy}, \tag{4.10}$$

$$u_3 = f_{yyy}[f_{yy}f_y^{-4} + 4L_0 f_y^{-2} - M_0 f_y^{-3}]/300 + \text{terms independent of } f_{yyy}, \tag{4.11}$$

and for $n \geqslant 4$

$$u_n(n+1)(n-6) = k_n \frac{d^n f}{dy^n}[f_{yy}f_y^{-n-1} - L_0 f_y^{1-n} - M_0 f_y^{-n}] + \text{terms independent of } \frac{d^n f}{dy^n}, \tag{4.12}$$

$k_4 = -1/300$, $k_5 = -1/150$, $k_6 = -1/300$ so that the compatibility condition cannot be satisfied for arbitrary f. It means that there is no Painlevé admissible hyperbolic equation of type a.

B. Type b equations

These equations, as well as types c, d, e, have the dominant truncation $T(-1,-3)$. The type b equations are reducible by (4.1) to

$$u_{xz} = Du_x u + Eu_y u + Lu_y + Mu_x + Qu^2 + Pu + S, \tag{4.13}$$

where $z = x$ or y. The recurrence formula yields

$$u_1 = [u_0(E_0 u_{0,y} - L_0 f_y - M_0 + Q_0 u_0 - D_1 - E_1 f_y) - Bu_{0,y}](2A + 2Bf_y)^{-1}, \tag{4.14}$$

where due to (3.8)

$$u_0 = 2(A + Bf_y)/(D_0 + E_0 f_y). \tag{4.15}$$

Let us investigate the terms proportional to f_{yyy} for $n = R = 2$. We get

$$0 = E_0 f_{yyy}(BD_0 - E_0 A)(2AE_0 - BD_0 + E_0 Bf_y) \tag{4.16}$$

for arbitrary $f(y)$, hence the equations of type b (both parabolic and hyperbolic) are Painlevé admissible only if $E(x,y) = 0$ or $D(x,y) = 0$. By the transformations (4.1), (4.3) and if necessary by the interchange $x \leftrightarrow y$ they can be reduced to

$$u_{xz} = -2u_x u + Lu_y + Qu^2 + Pu + S, \quad z = x \text{ or } y. \tag{4.17}$$

The compatibility condition for the hyperbolic form then reads

$$f_{yy}(Q_0 - L_0) + f_y^2[Q_0(Q_0 - L_0) + L_1 - Q_1] - f_y P_0 = 0, \tag{4.18}$$

that implies $Q = L$ and $P = 0$, so that the Painlevé admissible hyperbolic equation of type b is

$$u_{xy} = -2u_x u + Q(x,y)[u_y + u^2] + S(x,y), \tag{4.19}$$

where Q and S are analytic functions.

The compatibility condition for the parabolic case of (4.17) reads

$$f_y(L_1 - Q_0 L_0) - Q_1 + Q_0^2 - P_0 = 0, \tag{4.20}$$

where from we get $Q = L_x/L$ and $P = Q_x - Q^2$ due to the arbitrariness of f, so that the most general Painlevé admissible parabolic PDE of type b is

$$u_{xx} = -2u_x u + Lu_y + (L_x/L)u^2 + (2L_x^2/L^2 - L_{xx}/L)u + S. \tag{4.21}$$

C. Types c and d

The relations between A, B, C and D, E imply

$$B^2 = 4AC, \tag{4.22}$$

which means that these types are always transformable to the parabolic equation

$$u_{xx} + (r-4)u_x u + (2-r)u^3 + Lu_y + Mu_x + Qu^2 + Pu + S = 0, \quad r = 1,3. \tag{4.23}$$

The compatibility conditions corresponding to $R = 1$ or $R = 3$, namely,

$$(f_y L_0)^R + \text{terms with smaller power of } f_y = 0, \tag{4.24}$$

require that $L = 0$ so that the Painlevé admissible equations of types *c* and *d* are always transformable to the ODEs classified by Painlevé.

D. Type e equations

The last equations that must be investigated are reducible to

$$u_{xz} = 2u^3 + Lu_y + Mu_x + Pu + S, \quad z = x \text{ or } y. \tag{4.25}$$

The analysis of these equations is similar to type a. There is no Painlevé admissible hyperbolic equation of type e because the compatibility condition requires

$$0 = f_{yyyy} f_{yy} f_y^{-9/2} + \text{terms independent of } f_{yyyy}. \qquad (4.26)$$

The recurrence formula for the parabolic form gives

$$(n+1)(n-4)u_n = k_n (f_y L_0)^n + \text{terms with smaller powers of } f_y, \qquad (4.27)$$

where $k_1 = -1$, $k_2 = 1/6$, $k_3 = -2/27$, $k_4 = 1/12, \ldots$. The compatibility condition for $r = 4$ requires $L(x,y) = 0$ hence all Painlevé admissible parabolic equations of the type e are reducible to ODEs.

V. CONCLUSIONS

We have extended the formulas for the Painlevé-type classification procedure suggested in Ref. 7 to PDEs in two variables. Their further extension to more variables is easily deducible.

The method was used for the classification of the first degree second-order polynomial PDEs. They belong to the dominance classes $T(-1, -3)$ or $T(-2, -4)$ defined in the Sec. II. The resonance analysis and the compatibility conditions imply that most of the Painlevé admissible PDEs from the investigated class are transformable to the ODEs classified by Painlevé. The only genuine Painlevé admissible PDEs are those reducible by (4.2), (4.4) to the form

$$u_{xy} = -2u_x u + Q[u_y + u^2] + S, \qquad (5.1)$$

or

$$u_{xx} = -2u_x u + Lu_y + (L_x/L)u^2 + (2L_x/L^2 - L_{xx}/L)u + S, \qquad (5.2)$$

where L, Q, S are analytic functions of x and y.

The former equation is immediately integrable. Its first integral is of the Riccati type

$$u_y = u^2 - V, \qquad (5.3)$$

where

$$V_x = QV + S. \qquad (5.4)$$

The latter one, i.e., Eq. (5.2) is the most general Painlevé admissible extension of the Burgers equation (cf. Refs. 4 and 6) and should be, therefore, integrable.

The procedure presented in this paper can be adapted for classification of other classes of equations, e.g., PDEs in many variables, rational PDEs (where the zero point analysis must be included), systems of equations, etc. The calculations obviously become much more elaborate and extensive (cf. Refs. 9 and 11) even though it seems that the essence of the method remains unchanged.

[1] M. J. Ablowitz, A. Ramani, and H. Segur, Lett. Nuovo Cimento **38**, 333 (1978).
[2] M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715 (1980).
[3] J. Weiss, M. Tabor, and G. Carnevale, J. Math. Phys. **24**, 522 (1983).
[4] W. H. Steeb and J. A. Louw, Phys. Lett. A **113**, 61 (1985).
[5] N. Joshi, Phys. Lett. A. **125**, 456 (1987).
[6] L. Hlavatý, Phys. Lett. A **128**, 335 (1988).
[7] L. Hlavatý, J. Phys. A. **21**, 2855 (1988).
[8] R. S. Ward, Phys. Lett. A **102**, 279 (1984).
[9] F. J. Bureau, Ann. Mat. Pura Appl. **64**, 229 (1964).
[10] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1944).
[11] F. J. Bureau, Ann. Mat. Pura Appl. **66**, 1 (1964).

The wetted solid—A generalization of the Plateau problem and its implications for sintered materials

Peter Salamon
Department of Mathematical Sciences, San Diego State University, San Diego, California 92182
J. Bernholc
Department of Physics, North Carolina State University, Raleigh, North Carolina 27695-8202
R. Stephen Berry and Martin E. Carrera-Patiño
Department of Chemistry and the James Franck Institute, The University of Chicago, Chicago, Illinois 60637
Bjarne Andresen
Physics Laboratory, University of Copenhagen, Universitetsparken 5, DK-2100 Copenhagen Ø, Denmark

(Received 26 January 1988; accepted for publication 25 October 1989)

A new generalization of the Plateau problem that includes the constraint of enclosing a given region is introduced. Physically, the problem is important insofar as it bears on sintering processes and the structure of wetted porous media. Some primal and dual characterizations of the solutions are offered and aspects of the problem are illustrated in one and two dimensions in order to clarify the combinatorial elements and demonstrate the importance of numerous local minima.

I. THE PROBLEM AND ITS PHYSICAL REALIZATIONS

We present a physically interesting and largely unexplored generalization of the problem of Plateau. Plateau's original problem concerns the surface of least area with a given boundary curve.[1–4] Its solutions, known as minimal surfaces, have zero mean curvature and are usually associated with soap bubbles and wire frames. A well-known generalization is the problem of minimum area with a given boundary and enclosing a given volume.[1,4] The solutions are again surfaces with constant mean curvature and are usually associated with the shape of a liquid–gas interface.

In the present work we introduce still another extension of this classic problem. The extension arises in the analysis of sintering processes and again incorporates a volume constraint while adding the constraint of enclosing a given region.

Problem: Given a region Ω and a positive real number V_Γ, find a region Γ with a smooth boundary $\partial\Gamma$ having the minimum area such that Γ contains Ω and has volume V_Γ.

Physically, this extension represents the problem of the *shape* Γ *of the wetted solid* Ω. Figure 1 illustrates the point. Some problems of this sort have been posed and solved,[5,6] but the general importance of the problem does not seem to have been previously appreciated and the purely geometric formulation given above is new.

The primary raison d'être of this paper rests on the mathematicians' traditional criterion of shedding new light on a classical problem. In addition, we introduce an example which forms the basis of further calculations[7–10] on sintering processes.

We begin by arguing that this problem indeed represents the ideal wetted solid. There are two physical interpretations. In the first, we force the liquid to cover the solid, although perhaps only with an infinitesimal layer. Since the area of the liquid–solid interface is then fixed and the area of the gas–solid interface is zero, the total surface free energy attributable to the shape of the wetted solid is proportional to the total surface area of the liquid, which leads to the conclusion that the shape of Γ minimizes such an area. In this idealization we neglect the effect of layer thickness on the energies.[11]

The second interpretation lies closer to our presently intended applications. We imagine the liquid and the solid to be very similar in their physical properties: In fact, we take them to be identical except for their ability to support a shear stress. Our motivation comes from surface melting models of sintering processes.[12] In such processes the "liquid" merely represents a more mobile form of the solid. Specifically, we assume that the liquid and solid are so similar that the energy per unit area for the gas–solid and gas–liquid interfaces are equal, while the liquid–solid surface tension is negligible, i.e.,

$$\sigma_{gl} = \sigma_{gs}, \quad \sigma_{ls} = 0, \tag{1}$$

where σ_{gl}, σ_{ls}, and σ_{gs} are the respective surface tensions and the subscripts refer to gas, liquid, and solid. Corresponding to a variation in the shape of the total region of liquid and solid Γ, the variation dG in total surface free energy is now the sum

$$dG = \sigma_{gl}\, dA_{gl} + \sigma_{ls}\, dA_{ls} + \sigma_{gs}\, dA_{gs}, \tag{2}$$

where the dA's are variations in the areas of the respective surfaces. Using conditions (1) on the surface tensions, Eq. (2) becomes

$$dG = \sigma_{gl}(dA_{gl} + dA_{gs}) = \sigma_{gl}\, dA_\Gamma, \tag{3}$$

where A_Γ is the area of the outer surface of the condensed phase (solid or liquid) in contact with the gas. Condition

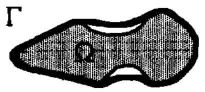

FIG. 1. The shaded solid Ω is wetted with white "liquid" to form the total object Γ, within the heavy outline, which has the given volume V_Γ.

 0022-2488/90/030610-06$03.00

(3), along with $\sigma_{gl} > 0$, implies that the minimum surface free energy coincides with the minimum area for $\partial\Gamma$, the boundary separating Γ from the gas.

Our problem is related to classical capillarity problems.[4,13] The usual condition for the contact angle θ is known as the Young equation:

$$\sigma_{gs} = \sigma_{ls} + \sigma_{gl} \cos \theta. \tag{4}$$

When our assumptions on the surface tensions, Eq. (1), are substituted into Eq. (4), we find that θ should vanish. This is not surprising in light of the fact that Eq. (4) follows from the minimization of the total surface free energy given the relative worths of the areas at the interfaces. Since our formulation counts liquid and solid surfaces equally, the optimal contact angle is zero.

II. STRUCTURE OF THE SOLUTIONS

While existence and regularity are relatively easy to establish for our problem,[14] in the present development we entirely ignore such issues. We assume existence and regularity and focus on other, equally interesting aspects. After deducing some necessary conditions arising from global optimality, we turn to combinatorial aspects which follow from the fact that in general the solutions are far from unique. We show this to be the case by considering the problem first in one and two dimensions.

A. One dimension

In one dimension we are looking for a set Γ which covers (contains) a given set Ω, has given length $L(\Gamma)$, and such that it has the minimum number of endpoints, i.e., such that the cardinality $C = |\partial\Gamma|$ is minimum. While this problem is very easy, it already exhibits highly degenerate solutions with numerous local minima for even moderately complicated Ω. These features introduce the combinatorial considerations which stay with the problem in two and three dimensions. As a concrete one-dimensional example take the following union of intervals [see Fig. 2(a)]:

$$\Omega = [-3,-2] \cup [-1,0] \cup [0.5,1] \cup [1.5,2]. \tag{5}$$

The minimal cardinality C^* as a function of length is

$$C^*(L) = \min|\partial\Gamma| = \begin{cases} \text{undefined}, & \text{if } L < 3, \\ 8, & 3 \leqslant L < 3.5, \\ 6, & 3.5 \leqslant L < 4, \\ 4, & 4 \leqslant L < 5, \\ 2, & 5 \leqslant L, \end{cases} \tag{6}$$

while the number of different ways of achieving C^* is

$$N(C^*(L)) = \begin{cases} 1, & \text{if } \ L \in \{3,4,5\} \\ 2, & \text{if } \ L = 3.5 \\ \infty, & \text{if } \ L \notin \{3,3.5,4,\}. \end{cases} \tag{7}$$

The two possible coverings with length $L = 3.5$ are shown in Fig. 2(b).

One could further resolve the infinite solution set obtained for other values of L by deriving an expression for its volume $V(C^*(L))$. For example, for $L = 4.8$ all solutions are of the form

$$\Gamma = [-3 - x_1, -2 + x_2] \cup [-1 - x_3, 2 + x_4], \tag{8}$$

with $x_i \geqslant 0$ for $i = 1, 2, 3, 4$ and $\Sigma_i x_i = 0.8$, as illustrated in Fig. 2(c). Generalizing, we see that with the parametrization (8), the solution set for the given L is always a simplex $x_i \geqslant 0$ for $i = 1,\ldots,C^*(L)$ and $\Sigma_i x_i = L - L^*$, where $L^* = \min L$ with the given C^*. This simplex has the volume

$$V(C^*(L)) = (L - L^*)^{C^*} \sqrt{C^*}/(C^* - 1). \tag{9}$$

Fixing C at some value above C^*, the solution sets become more complicated, at least in part due to the appearance of local minima. As an example, consider $L = 4$ and $C = 6$; one "locally optimal" solution is given by $\Gamma = [-3,0] \cup [0.5,1] \cup [1.5,2]$, as shown in Fig. 2(d).

B. Two dimensions

The problem in two dimensions is richer. Since a globally optimal solution is also locally optimal, we examine a portion of an optimal configuration for which we may choose a Cartesian coordinate system in which a suitable portion of the boundary $\partial\Gamma$ is given by a smooth curve $f(x)$ on an interval $[x_1,x_2]$, with Γ locally defined by $Y \leqslant f(x)$. We further divide the interval $[x_1,x_2]$ into subintervals according to whether f can have one- or two-sided variations, i.e., according to whether $f(x)$ does or does not coincide with the boundary of Ω.[15] On intervals where two-sided variations are available, the local problem is just the classic isoperimetric problem[1,4,16–18] of minimizing the length

$$L(f) = \int_{x_1}^{x_2} \sqrt{1 + f'^2}\, dx \tag{10}$$

subject to a given area

$$V = \int_{x_1}^{x_2} f(x)\, dx \tag{11}$$

and given endpoints. The classical results assure us that the solutions must be pieces of circles with radius $R = 1/\lambda$, where λ is the Lagrange multiplier from the Lagrangian $L = \sqrt{1 + f'^2} + \lambda f$. Since λ is the Lagrange multiplier corresponding to the area constraint, it must also equal the rate at which perimeter increases per unit change in area.[16–18] From this we can see that a necessary condition for global optimality is that *all such circular arcs have the same radius*! Note that this follows from the sign of the first variation, which transfers some area from one interval to another. While our

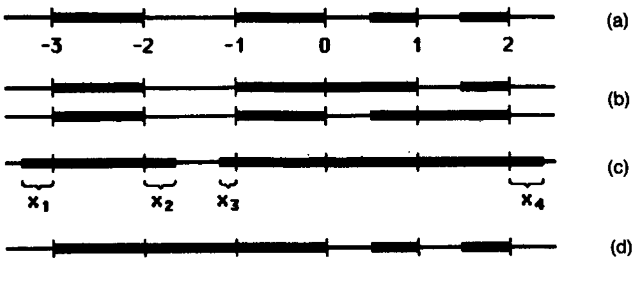

FIG. 2. (a) The one-dimensional region Ω given in Eq. (5). (b) Two possible coverings with length $L = 3.5$. (c) A generic element from the simplex of coverings with $L = 4.8$. (d) A covering with $L = 4$, which represents only a local minimum.

arguments are local, they may be applied to any portion of $\partial\Gamma$; thus we conclude that this boundary is the union of pieces of $\partial\Omega$ and pieces of circles. We see that such circles must be tangent to $\partial\Omega$ by another local argument. Again consider a Cartesian coordinate system and an interval $[x_1,x_2]$, where the boundary $\partial\Omega$ is given by a function $Y_\Omega(x)$ and $\partial\Gamma$ makes contact with the $\partial\Omega$ in the interval. The inclusion of Ω in Γ is then expressed by the inequality

$$f(x) \geqslant Y_\Omega(x). \tag{12}$$

Classical arguments on corner conditions with one-sided variations[19] assure us that f is tangent to $Y_\Omega(x)$ since $L_{f'f'}$ cannot vanish.

As a final global condition, we find that the curvature of the circular arcs on the "wet" portion of $\partial\Gamma$ (i.e., on the portion where it is distinct from $\partial\Omega$) must be less than the curvature of the dry portion where $\partial\Gamma$ and $\partial\Omega$ coincide. This follows from the same first variation argument we used to conclude that the circular arcs all had to have the same radii. This tempts us to attempt the construction of an optimal family of solutions for a given Ω and progressively larger areas V_Γ by "growing" the solutions along the segments of minimum curvature. While this construction gives locally optimal shapes, it can fail to take advantage of topological changes which could improve the objective, i.e., decrease total perimeter (see Fig. 3).

In fact, the physically realized state for a wetted porous medium depends in detail on the fill–drain history of the sample. Accordingly, it is of as much interest to give the density of states at a certain energy and volume as to give the shape which realizes the absolute minimum of the energy at the volume.

C. Three dimensions

The situation in three dimensions is very similar. We again turn first to the local problem, which has been well studied and for which standard arguments guarantee existence and regularity.[1,4] Using a coordinate system, we focus on a portion of $\partial\Gamma$ such that this boundary is defined by a function $z=f(x,y)$ and Γ is locally defined by $z \leqslant f(x,y)$ for (x,y) in an interval $I=[x_1,x_2]\times[y_1,y_2]$. On this interval the problem becomes the well-known obstacle problem with a constraint. We again divide into subregions according to whether or not f coincides with $\partial\Omega$. On subregions where f is distinct from $\partial\Omega$, we are allowed two-sided variations. The problem is then one of minimizing

$$A(f) = \int_I \sqrt{1+f_x^2+f_y^2}\, dx\, dy \tag{13}$$

subject to the constraint of the given volume

$$V_\Gamma = \int_I f(x,y)\, dx\, dy, \tag{14}$$

where f_x and f_y are partial derivatives of f with respect to x and y. This gives the Lagrangian

$$L = \sqrt{1+f_x^2+f_y^2} + \lambda f, \tag{15}$$

whose extremals are surfaces with constant mean curvature:

$$\bar{\kappa} = (1/R_1 + 1/R_2)/2 = \lambda, \tag{16}$$

where R_1 and R_2 are the radii of curvature in two conjugate directions and λ is again the Lagrange multiplier corresponding to the volume constraint. As before, λ represents the rate at which surface area must increase per unit increase in volume; thus the global solution must consist of pieces of $\partial\Omega$ and pieces that are surfaces of constant mean curvature tangent to $\partial\Omega$. By the same first-order variational argument employed above, we find that the mean curvature of all these pieces must be the same and it must be greater than the mean curvature anywhere on the portion of $\partial\Gamma$ which coincides with $\partial\Omega$. The tangency of f where it again meets $\partial\Omega$ follows by standard results on one-sided variations.

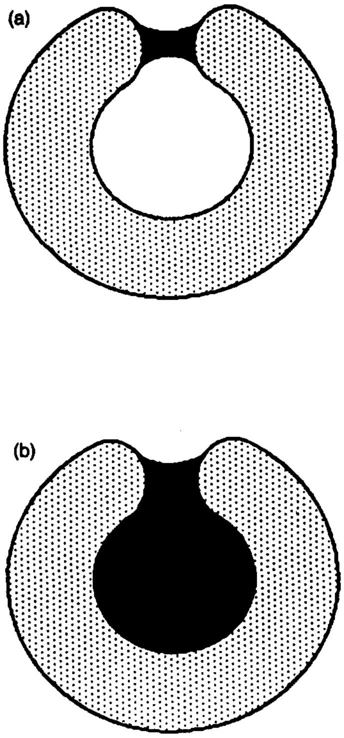

FIG. 3. The shaded solid is wetted by black "liquid" of the same mean curvature in both panels, but the difference in liquid volume is due to the past history of the system: dry in panel (a) and wet in panel (b).

Physically, we can understand the solutions as puddles forming on the solid skeleton provided by Ω. The fact that all the puddles have the same mean curvature results from the familiar equation[20,21]

$$p_1 - p_2 = 2\sigma\bar{\kappa}, \tag{17}$$

which relates the pressure drop $p_1 - p_2$ across an elastic surface to the surface tension σ and mean curvature $\bar{\kappa}$ of the surface. The constancy of $\bar{\kappa}$ then follows from the equilibrium condition that the liquid pressure be the same in all the puddles.

To gain further physical insight, we introduce a dual realization of our problem. We again consider Ω to be the solid skeleton, but rather than covering Ω with a given volume $V_\Gamma - V_\Omega$ of liquid which wets the surface, we cover it with an elastic skin with constant surface energy density σ and envision pumping a gas at a given pressure p_2 into the compartment between Ω and the skin while fixing the external pressure at a level p_1 which is sufficiently large to guarantee that the elastic skin is everywhere pressed firmly against the solid when $p_2 = 0$. We will refer to this problem as the bean-in-a-bag or "Christo" problem. It is clear from this model, which involves pockets of gas, that the separation of the elastic skin will occur first from points of large mean curvature $\bar{\kappa}$, thereby reducing the surface area the most. In fact, the Christo problem helps by providing a physical realization of a dual in which the givens are Ω and $\kappa_{\min}$. In this dual formulation, the problem is to find the region Γ with the minimum volume which contains Ω and whose mean curvature is everywhere greater than or equal to $\kappa_{\min}$. Achieving a given $\kappa_{\min}$ requires pumping the gas under the elastic skin with a given pressure $p_2 = p_1 - 2\sigma\kappa_{\min}$. We could also characterize the problem by asking for the smooth surface containing Ω and volume V_Γ which has the largest value of the minimum mean curvature.

D. Two properties

We conclude this section with two general and yet powerful properties of the class of solutions. We will refer to the first of these as the *layering property*: Let Γ be a solution to the problem given Ω_0 and V and let $\Omega_0 \subset \Omega_1 \subset \Gamma$; then Γ is also a solution to the problem given Ω_1 and V.

The proof is immediate. The layering property derives from the fact that when some of the liquid covering a wetted solid freezes, its freezing does not affect the shape of the liquid above it, i.e., of the new wetted solid (assuming that the liquid does not change its volume upon freezing).

The layering property hints at a universality of structure among solutions to the problem which we pursue a little further here. To see this we define an equivalence relation on the family of solid skeletons. Formally, we say that Ω_1 is equivalent to Ω_2 at volume V and write

$$\Omega_1 =_V \Omega_2, \tag{18}$$

provided that there exists a region Γ which solves the wetted solid problem with Ω_1, V as data and, also, the problem with Ω_2, V as data. That is to say that by the time we have covered up Ω_1 or Ω_2 to a level V, their distinctive jagged features have been covered over by the puddles. This leads to something resembling ultrametricity among the set of states containing a skeleton Ω_0 and having a given volume. The associated "distance" can be thought of as the total fill–drain volume needed to reach Ω_1 from Ω_2.

As the final property, we mention *scale invariance*. Specifically, let Γ be a solution of the problem for Ω and V and let μ be a scale factor for the map sending (x,y,z) to $(\mu x,\mu y,\mu z)$ in some coordinate frame. Then the region $\mu\Gamma$ solves the problem for given $\mu\Omega$ and $\mu^3 V$.

III. APPLICATIONS

Because of its relation to porous media and sintered materials, the case where Ω is a lattice of packed spheres is of great interest.[5–10] For values of V_Γ near V_Ω, solving the problem is equivalent to locating the puddles in the necks surrounding the points of tangency between spheres. Puddling grows until the liquid or mobile "phase" attains a volume $V_1 = V_\Gamma - V_\Omega$ at which these puddles first come into contact. For a lattice of identical spheres, the "liquid" layer becomes connected at this stage, while patches of solid $\partial\Omega$ still show through. In the next stage, Γ has a surface of constant mean curvature with the topology of a three-dimensional lattice: We may suppose for the present that it belongs to the recently announced class of periodic complete surfaces of constant mean curvature.[22] Once the given volume V_Γ increases considerably beyond this value, the solution Γ begins to include filled pockets delineated by four spheres in mutual contact. Note that the Christo representation no longer works in the regime where pockets become filled. While there are many equivalent ways of filling such pockets for identical spheres, the order in which the pores are filled in a real porous material can make small differences and create many local optima. Finding the global optimum is then a problem of the modern "programming" sort and probably best attacked by methods such as simulated annealing.[23] Since the state of the real physical system is to a large extent dependent on its fill–drain history rather than on the true minimum of the free energy, the global optimum is again of secondary interest to counting the number of states at a certain level of suboptimality.

While we have referred to our problem as the problem of the wetted solid, it is important to note that our "liquid" merely represents mobile pools of material which can be redistributed along the surface of the solid. Realizations of interest include sintered materials, wetted porous media, and precipitates from saturated solutions. Nonetheless, in pursuing the example of the structure of a wetted collection of packed spheres, it is convenient to make intuitive arguments which treat the material that has been transported to the "necks" as though it were a liquid. That is not to say that this pool of material *is* a liquid; it is only to say that it is able to respond to surface tension forces (surface free energy differences) faster than the rate at which new material is supplied or transported to the mobile pool. This is certainly valid for sufficiently small neck sizes. It is also an excellent approximation even for large neck sizes for materials that respond quickly to local surface tension. One case in which this is likely is precisely that of a liquid surface.[24] Searching for conditions that give rise to surface melting was in fact the original motivation for investigating this class of problems. The possibility of a solid skeleton coated by a liquid that is identical to the solid in all ways except for its ability to support a shear stress was instrumental in the isolation of the zero-contact-angle case of the classical theory for the distribution of liquids on a solid.

Applications typically involve a one-parameter family of such problems. For the case of sintering, Ω evolves as sintering progresses. For another class of problems, the family of solutions is indeed well parametrized with Ω as the solid matrix which does not change as V_Γ increases and decreases. This could represent the growing together of precipitated particles immersed in a saturated solution which fills the pore spaces or the equilibrium structure of a liquid which wets a porous medium.

IV. TWO IDENTICAL SPHERES—AN EXAMPLE

We illustrate the above discussions with an example involving two identical spheres in point contact. This example

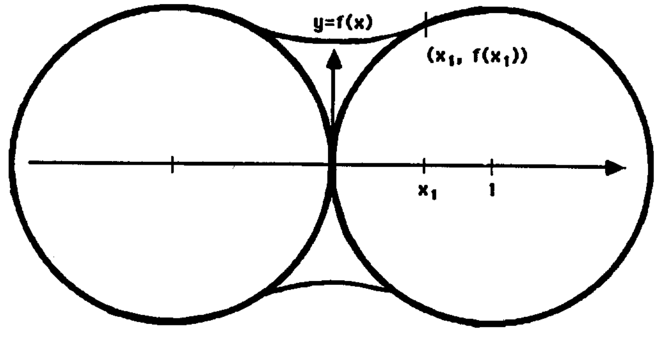

FIG. 4. Notation used for the example of two touching unit spheres, with a small amount of "liquid" filling the neck between them.

is the building block for treating a lattice of spheres in the regime before the puddles in the different necks touch each other. The rotational symmetry of the example allows the problem to be characterized using surfaces of revolution, thus reducing the associated partial differential equations to ordinary differential equations. Rotationally symmetric surfaces of constant mean curvature are known as Delaunay surfaces and have been extensively studied.[25,1,2,4] The problem of fitting them to enclose a given volume around an evolving skeleton Ω represents a new twist appropriate to applications of sintering processes. The one-parameter family of Ω's we consider is where the spheres gradually get smaller, releasing progressively more volume into the mobile pool.

By the scaling property, it is sufficient to solve the problem of two unit spheres in point contact and then scale the results. By symmetry, we may limit our view to the first quadrant. We use the notation of Fig. 4 for the curve $y = f(x)$, which is the generator for the surface in the region of the neck, and let

$$y = y_\Omega(x) \equiv \sqrt{1-(x-1)^2} \tag{19}$$

for y on the circle which defines $\partial\Omega$.

To find the shape of the fluid, we set up the calculus of variations problem to minimize the surface area subject to a fixed volume constraint. Formally, we ask for f, which minimizes the surface area

$$A = 2\pi\int_0^{x_1} f\sqrt{1+f'^2}\,dx + 2\pi\int_{x_1}^{2} y_\Omega\,\sqrt{1+y_\Omega'^2}\,dx \tag{20}$$

subject to constrained total volume

$$V = \pi\int_0^{x_1} f^2\,dx + \pi\int_{x_1}^{2} y_\Omega^2\,dx \tag{21}$$

with $f'(0) = 0$. The conditions for the point x_1, where the boundaries $\partial\Gamma$ and $\partial\Omega$ join, are $f(x_1) = y_\Omega(x_1)$ and $f'(x_1) = y_\Omega'(x_1)$. This yields the Lagrangian

$$\mathbb{L} = f\sqrt{1+f'^2} + \lambda f^2. \tag{22}$$

Letting

$$H = \mathbb{L} - f'\frac{\partial\mathbb{L}}{\partial f'} = \text{const}, \tag{23}$$

we have the Euler–Lagrange equation[16–18]

$$f' = \frac{df}{dx} = \pm\sqrt{\frac{f^2}{(H+\lambda f^2)^2} - 1}, \tag{24}$$

where H and λ are constants which must be determined from the boundary conditions. While Eq. (24) can be integrated to give f in terms of elliptic functions,[26] the evaluation of f corresponding to a situation of interest is more easily achieved numerically. Useful methods for calculation are discussed in Ref. 8. Here we mention only that the family of solutions is obtained most conveniently in terms of x_1 and that it necessitates some shooting method[27] for most ways of specifying the data for the problem. A solution with unit radius can be scaled to radius R to give $f_R(x) = Rf(x/R)$, which satisfies Eq. (24) with $H_R = RH$ and $\lambda_R = \lambda/R$ and leads naturally to $x_{1R} = RX_1$ and volume $V_{\Gamma R} = R^3 V_\Gamma$.

Note that the Euler–Lagrange equation (24) is completely independent of y_Ω. The solution depends on Ω only through the boundary conditions and in fact, it is satisfied for any solid of revolution with constant mean curvature, i.e., any Delaunay surface. The fact that the Lagrange multiplier λ coincides with the mean curvature $\bar{\kappa}$ can be seen by applying the formula for the mean curvature of a surface of revolution generated by $y = f(x)$ revolved about the x axis[28]:

$$\bar{\kappa} = \frac{1}{2\sqrt{1+f'^2}}\left[\frac{f''}{1+f'^2} - \frac{1}{f}\right]. \tag{25}$$

By inserting Eq. (24), Eq. (25) becomes $\bar{\kappa} = \lambda$.

It is interesting to note that the importance of Delauney surfaces have not been previously recognized in sintering studies, although an instance of them can be found in a previous study of porous media.[5]

The building block of two hard spheres in the small neck regime can be used to treat random or close packed arrays of spheres with known distributions of radii. The new aspects are again combinatorial.

V. CONCLUSIONS

In this paper we have introduced a new modification of the problem of Plateau. Assuming existence and regularity, we used standard results concerning the local version of the problem to deduce new global conditions. While these conditions become obvious after some reflection, they are sufficient to assemble global solutions for many physically interesting examples. We also sketched a method for obtaining the solution in a radially symmetric example important for sintering processes.

Our approach provides a realistic model of wetted porous media and is of particular importance for the understanding of sintered materials. To invoke the present solutions for a wetted solid, we have to ignore the nonideality in the form of a thickness-dependent free energy responsible for the disjoining pressure of Deryaguin.[29,30] Ours is a particularly appropriate model for the sintering processes in which some sort of enhanced surface mobility or surface melting occurs, but the formation of the melted layer is the slow variable. The present model should work very well under such conditions. In particular, it is a much better approximation to reality than the traditional models such as the circle approximation for $f(x)$ advanced by Kuczynski[31] in the

1950's and used widely since. Our model differs by as much as 200% relative error (for small neck sizes!) and agrees better with experiment.[7,10]

The present approach stresses the importance of combinatorial methods, local minima, and density of solutions, rather than the absolute minimum. Work in this area has drifted away from an interest in the detailed shape of the surface to models for the value of the thermodynamic potential of the liquid covering Ω.[5,6,30] There remains important information to be gained from microscopic details which can supplement macroscopic phenomenology, including models of thermodynamic potentials.

ACKNOWLEDGMENTS

We would like to thank J. M. Pedersen, J. Jellinek, F. Amar, D. Fisher-Colbrie, A. Zandi, and D. Podber for helpful discussions. The hospitable surroundings of Telluride Summer Research Center and the Aspen Center for Physics are much appreciated.

Author PS acknowledges a grant from the Danish Natural Science Research Council. Authors RSB and MC wish to acknowledge the support of the Department of Energy under Grant No. DE-FG02-86ER13488. Authors JB, RSB, and PS acknowledge partial support from Office of Naval Research, Grant No. N00014-88-K-0491.

[1]R. Osserman, *A Survey of Minimal Surfaces* (Van Nostrand, Amsterdam, 1969).

[2]J. C. C. Nitsche, *Vorlesungen über Minimalflächen* (Springer, Berlin, 1975).

[3]J. L. Troutman, *Variational Calculus with Elementary Convexity* (Springer, New York, 1983).

[4]R. Finn, *Equilibrium Capillary Surfaces* (Springer, New York, 1986).

[5]W. von Engelhardt, in *Proceedings of the Fourth World Petroleum Congress*, edited by A. Barbier (Carlo Colombo, Rome, 1955), Vol. 1, p. 399.

[6]A. Scheidegger, *The Physics of Flow Through Porous Media* (Univ. of Toronto, Toronto, 1974).

[7]F. Amar, J. Bernholc, R. S. Berry, J. Jellinek, and P. Salamon, J. Appl. Phys. **65**, 3219 (1989).

[8]C. Schön, P. Basa, J. Bernholc, R. S. Berry, J. Jellinek, and P. Salamon, "On the Wetting of Radially Symmetric Surfaces" (unpublished).

[9]R. S. Berry, J. Bernholc, and P. Salamon, "The Disappearance of Grain Boundaries in Sintering" (unpublished).

[10]J. Bernholc, P. Salamon, and R. S. Berry, in *The Physics and Chemistry of Small Clusters*, edited by P. Jena, B. K. Rao, and S. N. Khanna (Plenum, New York, 1987), p. 43.

[11]This is in contradistinction to the approach taken in Refs. 29 and 30, where the effect of layer thickness on energy is assumed to be dominant.

[12]Surface melting[20,21] is a phenomenon whereby the surface of a solid melts at a lower temperature than the bulk. Sintering is the process when packed powdered material is heated (and/or pressed) to promote partial cohesion of the powder.

[13]P. G. de Gennes, Rev. Mod. Phys. **57**, 827 (1985).

[14]E. Giusti, *Minimal Surfaces and Functions of Bounded Variation* (Birkhauser, Boston, 1984).

[15]We ignore the possibility that the two sets defined in this fashion may be pathological, i.e., something other than a union of disjoint intervals.

[16]L. E. Elsgolc, *Calculus of Variations* (Pergamon, London, 1962).

[17]I. M. Gelfand and S. V. Fomin, *Calculus of Variations* (Prentice-Hall, Englewood Cliffs, NJ, 1963).

[18]G. Leitman, *The Calculus of Variations and Optimal Control* (Plenum, New York, 1981).

[19]Reference 16, pp. 94–98.

[20]W. J. Moore, *Physical Chemistry* (Prentice-Hall, Englewood Cliffs, NJ, 1972).

[21]M. Prutton, *Surface Physics*, Oxford Physics Series (Clarendon, Oxford, 1975).

[22]N. Kapouleas, Bull. Am. Math. Soc. **17**, 318 (1987).

[23]S. Kirkpatrick, C. D. Gelatt, Jr., and M. P. Vecchi, Science **220**, 671 (1983).

[24]V. V. Nauchitel and A. J. Pertsin, Mol. Phys. **40**, 1341 (1984).

[25]C. Delaunay, J. Math. Pure. Appl. **6**, 309 (1841).

[26]M. Carrera-Patiño, Ph.D. thesis, University of Chicago, 1988.

[27]W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling, *Numerical Recipes* (Cambridge U. P., Cambridge, 1986), pp. 582 ff.

[28]L. P. Eisenhardt, *The Differential Geometry of Curves and Surfaces* (Ginn, Boston, 1909); B. O'Neill, *Elementary Differential Geometry* (Academic, New York, 1966).

[29]J. R. Philip, J. Chem. Phys. **66**, 5069 (1977).

[30]B. V. Deryaguin, J. Phys. Chem. **3**, 29 (1932).

[31]G. C. Kuczynski, Trans. AIME **185**, 169 (1949); see, also, R. L. Coble, J. Am. Ceram. Soc. **41**, 55 (1958).

The Duistermaat–Heckman integration formula on flag manifolds

R. F. Picken
Freie Universität Berlin, Institut für Theorie der Elementarteilchen, FB 20, WE 4, Arnimallee 14, 1000 Berlin 33, Federal Republic of Germany and Grupo Teórico de Altas Energias, Instituto de Física e Matemática, Avenida Prof. Gama Pinto 2, 1699 Lisboa, Portugal [a]

(Received 11 October 1988; accepted for publication 20 September 1989)

An exposition is given of various geometrical properties of flag manifolds and of the Duistermaat–Heckman integration formula as applied to flag manifolds.

I. INTRODUCTION

We start with an informal discussion of the content of this paper and the motivation that led to this work. First of all, what is the Duistermaat–Heckman (DH)[1] integration formula? Put in the simplest possible terms, it is a generalization of the following elementary integral formula on the two-sphere S^2 coordinatized in the usual way by (θ,φ):

$$\int_{\theta=0}^{\pi}\int_{\varphi=0}^{2\pi}\exp(\cos\theta)\sin\theta\, d\theta\, d\varphi = 2\pi(e-e^{-1}). \tag{1.1}$$

One observes that the right-hand side is a sum of contributions from the two critical points of the function $\cos\theta$ at $\theta=0,\pi$. The generalization is as follows. First, one needs a compact manifold M, of dimension $2n$, with a symplectic form ω. (A symplectic form is a two-form ω such that its n-fold wedge product $\omega\wedge\cdots\wedge\omega=\omega^n$ vanishes nowhere on the manifold and such that $d\omega=0$, where d is the exterior derivative.) In the example above, $M=S^2$ and $\omega=\sin\theta\, d\theta\wedge d\varphi$ (the singularity at $\theta=0$ can be made to disappear by changing to better coordinates). From ω one can then form the symplectic volume $\omega^n/n!$. In the example, the symplectic volume is ω itself. The second ingredient is a function H on M called the Hamiltonian function because of links with the geometric formulation of Hamiltonian mechanics (see, for instance, Ref. 2). In the example, $H=\cos\theta$. Assuming for simplicity that H has only isolated critical points $\{m_I\}$, the DH formula states

$$\int_M \exp(H)\frac{\omega^n}{n!}=\sum_{m_I}\frac{\exp(H(m_I))}{D(m_I)}, \tag{1.2}$$

i.e., the right-hand side is again a sum over the critical points of H. The factors $D(m_I)$ are, in fact, precisely the factors one would get from evaluating the integral using a "semiclassical" or Gaussian approximation at each critical point, so that, loosely speaking, one may paraphrase the DH statement by saying that the semiclassical approximation is exact for these integrals. (In fact, as m_I need not be a local maximum of H, it is incorrect to talk of the semiclassical approximation. However, if one evaluates the fluctuation determinant as if m_I were a maximum, the answer obtained is exact.) As was pointed out by Kirwan,[3] semiclassical exactness in a related class of oscillatory integrals is a very special feature that only holds if H is a perfect Morse function.

A better way of viewing the result (1.2) is as an example of "localization": one can show that the integrand on the left-hand side is an exact form everywhere except at the critical points m_I. Thus one may cut out solid spheres of arbitrarily small radius around each critical point and, using Stokes' theorem, the integral reduces to a collection of surface integrals around each critical point. As the radii can be sent to zero this explains why the right-hand side of the DH formula is given solely in terms of local data at the critical points.

The notion of localization lies at the heart of some extremely fertile ideas developed by Witten.[4] Roughly speaking, similar localization arguments are applied to infinite-dimensional manifolds, the configuration spaces of various quantum field theoretical and string models. The relationship of these ideas of Witten with the DH formula was elucidated in Atiyah[5] and Atiyah and Bott.[6] From the physical point of view, supersymmetry plays a key role in localization, and in this context there is a close connection with the "physicist's proofs" of various cases of the Atiyah–Singer index theorem via supersymmetric sigma models.[7] It should be borne in mind, however, that the idea of applying the DH formula to infinite-dimensional manifolds is as yet mainly intuitive and lacks a sound mathematical footing.

The second part of our title also requires some explanation. What are flag manifolds? They are best thought of for the moment as certain types of homogeneous spaces G/H, where G is a Lie group and H is a subgroup of G. These special homogeneous spaces have a number of remarkable geometric properties, one of which is that they are Kähler manifolds. Thus, in particular, they possess a symplectic form ω, a necessary requirement for discussing the DH formula (1.2). Flag manifolds have appeared in the physics literature in a variety of contexts, e.g., as target manifolds for sigma models[8] or in a geometric formulation of harmonic superspace.[9]

Why, in our study of the DH formula, are we restricting ourselves to flag manifolds? After all, the DH formula is applicable in a far larger class of situations, whereas in this particular case it reduces to a well-known and old result, the Harish-Chandra formula.[10] (Although the results coincide, we shall continue to speak of the DH formula on flag manifolds instead of the Harish-Chandra formula, as our whole outlook and methods are based on the more modern approach.) We have, in fact, briefly described a more general setting for the DH formula at the start of Sec. V and the reader mainly interested in getting a flavor of the general result may skip to Sec. V straight away. One reason for choosing the special case of flag manifolds is that they provide a rich class of examples with interesting geometrical features, which are worth studying in their own right. How-

[a] Present address.

 0022-2488/90/030616-23$03.00

ever, our main motivation for concentrating on the DH formula on flag manifolds comes from the following chain of arguments. In order to be able to apply the DH formula in the generalized sense mentioned above with greater confidence, one needs a better understanding of the infinite-dimensional manifolds involved. One such class of manifolds has been the subject of much mathematical interest recently, namely, so-called loop groups, spaces of loops on Lie groups, i.e., of maps from the circle to a Lie group (see the authoritative book by Pressley and Segal[11]). The point is that there are far-reaching analogies between flag manifolds, on the one hand, and loop groups, on the other, so much so that one can state that flag manifolds are *the* finite-dimensional analogs of loop groups. In short, then, our reason for studying the DH formula on flag manifolds is that this is the next best thing to studying the DH formula on loop groups, and is indispensible for a thorough understanding of the infinite-dimensional case.

Loop groups are not just of mathematical interest but are also relevant to physics for a variety of reasons. Both the Hilbert space for the quantum mechanical model of a point particle on a group manifold and the configuration space for strings on a group manifold[12] are loop groups. Generally a theory of maps from the circle to a Lie group is the prototype of a sigma model, from the analysis of which one may be led to more general conclusions about sigma models. Our own interest in the whole subject was originally awakened by the first topic, quantum mechanics on a group, the aim being to link a DH-type localization on the loop group to a different kind of "semiclassical exactness" in this model known from older treatments. For more details, see Ref. 13.

At the risk of stating the obvious it should be pointed out that much of our material is, of course, not new, but drawn from various areas of the mathematical literature. To our knowledge, however, a study of this nature, which puts together the different parts of the mathematical theory in a manner accessible to physicists, has not previously been performed, and thus we hope that the paper fulfills a certain need. We would also like to mention that, en route, we have, in fact, added a number of new results (such as the construction of the symplectic volume via line bundles in Sec. III, the expression for the Hamiltonian in Bruhat coordinates in Sec. IV, and the recursion formula for Duistermaat–Heckman problems in Sec. V).

The outline of our material is as follows. In Sec. II we concentrate on the description of flag manifolds as complex manifolds and discuss their decomposition into complex cells, which also gives rise to a coordinatization in terms of complex coordinates. In Sec. III it is shown how to construct suitable symplectic forms on flag manifolds, using two different methods, one algebraic and one via complex line bundles. A method for obtaining the symplectic volume in the line bundle approach is also described. In Sec. IV a special class of Hamiltonian functions on flag manifolds is introduced, particularly appropriate to the complex coordinates from Sec. II and the symplectic forms of Sec. III. In Sec. V the DH formula is discussed and applied in a number of examples using the formalism developed in the previous sections. Finally various relevant results from Lie algebra structure theory and representation theory are collected in the Appendix.

II. CELL DECOMPOSITION AND COMPLEX COORDINATES FOR FLAG MANIFOLDS

Let us start by saying what we mean by flag manifolds. Let G be a compact connected Lie group, T its torus, and $C(T_0)$ the centralizer in G of some subtorus T_0 of T, i.e., $C(T_0) = \{g\in G \,|\, g^{-1}T_0 g = T_0\}$. Generically, $C(T_0) = T$, but, for some special choices of subtorus T_0, $C(T_0)$ is a larger subgroup of G, containing T. Then the coset space G/T [resp. $G/C(T_0)$] is called a flag manifold. When we do not wish to distinguish between the generic and nongeneric case we will denote the flag manifold by M. The term flag manifold derives from the action of M on "flags," which are nested sequences of linear spaces

$$\{C_1,\dots,C_r \,|\, C_i\subset C_{i+1},\ \dim(C_i) < \dim(C_{i+1})\}.$$

The reason for this nomenclature is made clear in Fig. 1. A simple example of a flag manifold is $SU(2)/U(1)$. $SU(2)$ acts transitively on $\mathbb{C}P^1$ (the space of complex lines in $\mathbb{C}^2$, i.e., a manifold of flags) by left matrix multiplication. The $U(1)$ torus, being the subgroup of diagonal matrices, fixes the flag

$$\mathscr{F}_0 = \left\{\begin{pmatrix} z \\ 0 \end{pmatrix} \middle|\, z\in\mathbb{C}\right\}.$$

Thus $SU(2)/U(1)$ acts effectively and transitively on this flag manifold and we may identify each element of $SU(2)/U(1)$ with its image under the action on $\mathscr{F}_0$.

There exist two further ways of viewing flag manifolds, which highlight several of their special features. First, there is a natural isomorphism of M with the adjoint orbit $\{g\lambda g^{-1} | g\in G\}$, where λ is an element of the torus Lie algebra $\mathfrak{t}$. Under this isomorphism the coset $[g]_T = \{gt\,|\,t\in T\}$ corresponds to $g\lambda g^{-1}$ in the adjoint orbit [in the nongeneric case λ must be chosen such that $C(T_0)$ is the centralizer of λ in G]. Adjoint orbits come equipped with a natural symplectic structure (Kirillov[14]) and thus flag manifolds inherit this structure also. Symplectic structures on flag manifolds will be studied in detail in the next section.

Second, let G_c be the complexification of G, B be a Borel subgroup of G_c, and P be a parabolic subgroup of G_c such that $P\cap G = C(T_0)$. [We refer to the Appendix for definitions, but to fix ideas we take as an example $G_c = SU(3)_c = SL(3,\mathbb{C})$. Then the subgroup of upper triangular matrices is a Borel subgroup of G_c and, in a $(2+1)\times(2+1)$ block decomposition of the elements of G_c, the subgroup of block upper triangular matrices provides an example of a parabolic subgroup.] Then there exist isomorphisms $G/T\cong G_c/B$ and $G/C(T_0)\cong G_c/P$. In one di-

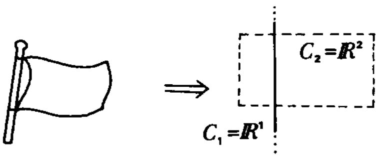

FIG. 1. The origin of the name "flag manifold."

rection $[g]_T$ is mapped to $[g]_B$. In the other direction one uses the so-called Iwasawa decomposition: any element $g_c \in G_c$ may be factorized as

$$g_c = gb, \quad g \in G, \quad b \in B, \tag{2.1}$$

in a unique fashion, up to torus elements, which are common to G and B. Thus, in the other direction, $[g_c]_B = [gb]_B$ is mapped to $[g]_T$. As an example, consider $SU(n)/T \cong SL(n,\mathbb{C})/B$, where B is the Borel subgroup of upper triangular matrices. The Iwasawa decomposition in this case may be proved by means of the Gramm-Schmidt orthonormalization process: regard $g_c \in SL(n,\mathbb{C})$ as the juxtaposition of n column vectors of length n ($e_1,...,e_n$, say). Then one obtains orthogonal vectors $\{e_i'\}$, $i = 1,...,n$, in the usual way:

$$\begin{aligned} e_1' &= e_1, \\ e_2' &= e_2 - ((e_2,e_1')/(e_1',e_1'))e_1', \\ &\vdots \\ e_n' &= e_n - ((e_n,e_{n-1}')/(e_{n-1}',e_{n-1}'))e_{n-1}' - \cdots \\ &\quad - ((e_n,e_1')/(e_1',e_1'))e_1'. \end{aligned} \tag{2.2}$$

Here (e_i,e_j) denotes the Hermitian product $e_i^T \bar{e}_j$. After normalization of $\{e_i'\}$ the juxtaposition of the normalized $\{e_i'\}$ constitutes a unitary matrix g related to g_c by $g = g_c b'$, where b' is upper triangular. As the upper triangular matrices form a group, this implies $g_c = gb$, where $b = (b')^{-1}$ is upper triangular.

Through this second isomorphism flag manifolds not only become complex manifolds but, as will now be described, they also admit a decomposition into complex cells arising from the Bruhat decomposition of G_c. Let W denote the Weyl group of G (a discrete group of permutations acting on the Weyl chambers of $\mathscr{V}$—see the Appendix). Let the action of an element $w \in W$ be represented by a matrix $\omega_w \in G$ acting via the coadjoint action: $w(Z) = \omega_w Z \omega_w^{-1}$, for $Z \in \mathscr{V}$. The Bruhat decomposition of G_c is a disjoint decomposition of G_c into double cosets:

$$G_c = \bigcup_{w \in W} B\omega_w B = \bigcup_{w \in W} \{b_1 \omega_w b_2 | b_1, b_2 \in B\}. \tag{2.3}$$

Thus one also has a disjoint decomposition of G_c/B, namely,

$$G_c/B = \bigcup_{w \in W} [B\omega_w]_B \equiv \bigcup_{w \in W} M_w. \tag{2.4}$$

This is a cell decomposition as the B on the left-hand side factorizes into the complex torus and a nilpotent factor (i.e., the corresponding Lie algebra is nilpotent under the Lie bracket). The torus factor may be permuted through ω_w and dumped in the right-hand B whilst the remaining nilpotent part is affine and may be given a cell structure. In fact, depending on w, it may be possible to dump some of the nilpotent part into the right-hand B as well. More precisely, Atiyah[15] has shown that the partial ordering of the Weyl group elements coming from group theory [$w' \geqslant w$ iff $\langle Z_1, w'(Z_2)\rangle \leqslant \langle Z_1, w(Z_2)\rangle$, for any Z_1,Z_2 in the same Weyl chamber, where $\langle\, ,\, \rangle$ is the Cartan–Killing form] coincides with a geometric partial ordering

$$w' \geqslant w \Leftrightarrow M_{w'} \subset \bar{M}_w. \tag{2.5}$$

As $M_{w'} \cap M_w = \varnothing$, for $w' \neq w$, and all M_w are complex cells, the (real) dimension of $M_{w'}$ is at least 2 less than the dimension of M_w if $w' > w$. (The ordering in the Weyl group W has also been described in terms of "simple reflections in W" in Borel–Tits.[16]) The whole of the above discussion can be carried through in a similar fashion for nongeneric flag manifolds. In this case, one obtains, from the Bruhat decomposition of G_c, the cell decomposition

$$G_c/P = \bigcup_{w \in W/W_0} [B\omega_w]_P, \tag{2.6}$$

where W_0 is the Weyl group of $C(T_0)$.

In order to illustrate the above remarks we proceed to discuss some examples. First, consider the previous example $SU(2)/U(1) \cong SL(2,C)/B$, where B is the subgroup of upper triangular matrices. The Weyl group may be regarded as acting on $\mathscr{V}$, the real traceless diagonal matrices, by permuting the diagonal elements. Thus the Weyl group consists of two elements $\mathbf{1}$ and (12) acting as follows:

$$\mathbf{1}\begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix} = \begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix}, \quad (12)\begin{pmatrix} a & 0 \\ 0 & -a \end{pmatrix} = \begin{pmatrix} -a & 0 \\ 0 & a \end{pmatrix}. \tag{2.7}$$

These two elements are represented by $SU(2)$ matrices ω_1 and $\omega_{(12)}$, where

$$\omega_1 = \begin{pmatrix} 1 & 0 \\ 0 & 1 \end{pmatrix}, \quad \omega_{(12)} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \tag{2.8}$$

acting via the coadjoint action. Then the Bruhat decomposition is

$$SL(2,\mathbb{C})/B = [B\mathbf{1}]_B \cup \left[B\begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}\right]_B \equiv M_1 \cup M_{(12)}. \tag{2.9}$$

The first cell M_1 is just the point $[\mathbf{1}]_B$, whereas the second cell is isomorphic to $\mathbb{C}$, as may be seen by factorizing $b \in B$ as

$$b = \begin{pmatrix} 1 & z \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \zeta & 0 \\ 0 & \zeta^{-1} \end{pmatrix}, \quad z \in \mathbb{C}, \quad \zeta \in \mathbb{C} \setminus \{0\}. \tag{2.10}$$

The right-hand factor lies in the complex torus and may be absorbed in the right B coset after commuting with $\omega_{(12)}$. This leaves the left-hand factor, so that $M_{(12)}$ is identified with the complex plane $\mathbb{C}$. Thus one regains the familiar cell decomposition of $\mathbb{C}P^1 \cong SL(2,\mathbb{C})/B$ into $\mathbb{C} \cup \{\text{point}\}$. For later purposes one would like the identity to lie in the largest cell. This is achieved by an overall multiplication on the left by $\omega_{(12)}^{-1}$ giving

$$SL(2,\mathbb{C})/B = \left[\begin{pmatrix} 0 & -1 \\ 1 & 0 \end{pmatrix}\right]_B \cup \left\{\left[\begin{pmatrix} 1 & 0 \\ -z & 1 \end{pmatrix}\right]_B \middle| z \in \mathbb{C}\right\}. \tag{2.11}$$

Let $C \subset \mathscr{V}$ be the Weyl chamber $\{\mathrm{diag}(a,-a) | a \in \mathbb{R}_+\}$. Then using the Cartan–Killing bracket $\langle Z_1, Z_2 \rangle = 4\,\mathrm{tr}(Z_1 Z_2)$, for $Z_1, Z_2 \in \mathscr{V}$, one obtains the group theoretical ordering $(12) \leqslant \mathbf{1}$. Equivalently one may verify the geometric ordering $M_1 \subset \bar{M}_{(12)} \Rightarrow (12) \leqslant \mathbf{1}$ by writing

$$\begin{aligned} \left[\begin{pmatrix} 1 & 0 \\ -z & 1 \end{pmatrix}\right]_B &= \left[\begin{pmatrix} 1 & 0 \\ -z & 1 \end{pmatrix} \begin{pmatrix} -z^{-1} & -1 \\ 0 & -z \end{pmatrix}\right]_B \\ &= \left[\begin{pmatrix} -z^{-1} & -1 \\ 1 & 0 \end{pmatrix}\right]_B, \end{aligned} \tag{2.12}$$

thereby displaying M_1 as the "point at infinity" of the complex one-cell $M_{(12)}$.

The case $SU(3)/T \cong SL(3,\mathbb{C})/B$, with B the subgroup of upper triangular matrices, may be treated along parallel lines. The Weyl group acts on $\mathscr{V}$ as the group of permutations of the entries of $\mathrm{diag}(a,b,-(a+b))$. In Table I the Weyl group elements are listed together with their SU(3) representatives. After "dumping" as much as possible into the right-hand B and multiplying on the left by $\omega_{(13)}^{-1}$, the corresponding Bruhat cells are found to be

$$M_{(13)} = \left\{ \begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \left[\begin{pmatrix} 1 & z_1 & z_2 \\ 0 & 1 & z_3 \\ 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix} \right]_B \right\}$$

$$= \left\{ \left[\begin{pmatrix} 1 & 0 & 0 \\ -z_3 & 1 & 0 \\ -z_2 & z_1 & 1 \end{pmatrix} \right]_B \right\},$$

$$M_{(123)} = \left\{ \left[\begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ -z_1 & -z_2 & 1 \end{pmatrix} \right]_B \right\},$$

$$M_{(132)} = \left\{ \left[\begin{pmatrix} 1 & 0 & 0 \\ -z_3 & 0 & 1 \\ -z_2 & -1 & 0 \end{pmatrix} \right]_B \right\}, \tag{2.13}$$

$$M_{(23)} = \left\{ \left[\begin{pmatrix} 0 & 1 & 0 \\ 0 & -z_3 & 1 \\ 1 & 0 & 0 \end{pmatrix} \right]_B \right\},$$

$$M_{(12)} = \left\{ \left[\begin{pmatrix} 0 & 0 & -1 \\ -1 & 0 & 0 \\ -z_1 & 1 & 0 \end{pmatrix} \right]_B \right\},$$

$$M_1 = \left\{ \left[\begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \right]_B \right\}.$$

All z_i range over the complex plane. The pattern of partial orderings is given in this case by

$$(13) \leqslant \begin{matrix} (123) \\ (132) \end{matrix} \leqslant \begin{matrix} (12) \\ (23) \end{matrix} \leqslant \mathbf{1}. \tag{2.14}$$

As a third example we discuss a nongeneric case, namely, $SU(3)/C(T_0)$, where T_0 is the one-dimensional subtorus $\{\mathrm{diag}(e^{ia},e^{ia},e^{-2ia}) \mid a \in \mathbb{R}\}$. Let $M(2,1)$ denote the complex $(2+1)\times(2+1)$ block matrices. Then $C(T_0)$ is the group of unitary block-diagonal matrices in $M(2,1)$ with unit determinant, i.e., $C(T_0) = S(U(2)\times U(1))$. $SU(3)/C(T_0)$ is isomorphic to $SL(3,\mathbb{C})/P$, where P is the parabolic subgroup of block upper triangular matrices in $M(2,1)$ with unit determinant. In this case one has to deal with the relative Weyl group, i.e., the Weyl group of SU(3) factored by the Weyl group of $C(T_0)$, which contains just two elements $\mathbf{1}$ and (12). Using the factorizations $(123) = (13)(12)$ and $(132) = (23)(12)$, the relative Weyl group is seen to consist of three elements: $[\mathbf{1}]$, $[(23)]$, and $[(13)]$. In the Bruhat decomposition (2.6) one may use the same matrix representatives for ω_w as before, but because P is a larger subgroup than B it is possible to "dump" more of the left-hand B into the right-hand P. After multiplying on the left by $\omega_{(13)}^{-1}$, one obtains the Bruhat cells $M_{[w]} = \omega_{(13)}^{-1} B \omega_{[\omega]} P$:

$$M_{[(13)]} = \left\{ \left[\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ -z_2 & z_1 & 1 \end{pmatrix} \right]_P \right\},$$

$$M_{[(23)]} = \left\{ \left[\begin{pmatrix} 0 & 1 & 0 \\ 0 & -z_3 & 1 \\ 1 & 0 & 0 \end{pmatrix} \right]_P \right\}, \tag{2.15}$$

$$M_{[(1)]} = \left\{ \left[\begin{pmatrix} 0 & 0 & -1 \\ 0 & 1 & 0 \\ 1 & 0 & 0 \end{pmatrix} \right]_P \right\}.$$

Again, the z_i range over $\mathbb{C}$. The ordering in this case is $[(13)] \leqslant [(23)] \leqslant [\mathbf{1}]$. $SL(3,\mathbb{C})/P$ is isomorphic to $\mathbb{C}P^2$, as may be seen by the following argument. The group P, acting from the left on $\mathbb{C}^3$ as a group of 3×3 matrices, is the stability subgroup of a $\mathbb{C}^2$ in $\mathbb{C}^3$ (namely, the $\mathbb{C}^2$ with bottom component equal to zero). Thus $SL(3,\mathbb{C})/P$ is the Grassmannian $G(2,3)$ of complex two-planes in $\mathbb{C}^3$. Each $\mathbb{C}^2$ in $\mathbb{C}^3$ uniquely defines a $\mathbb{C}^1$ in $\mathbb{C}^3$ by orthogonality and thus $G(2,3) \cong G(1,3) \cong \mathbb{C}P^2$. The Bruhat cell decomposition into $\mathbb{C}^0 \cup \mathbb{C}^1 \cup \mathbb{C}^2$ obtained in (2.15) corresponds to the well-known cell decomposition of $\mathbb{C}P^2$ using homogeneous coordinates $(\zeta_1,\zeta_2,\zeta_3)$, namely,

$$\mathbb{C}P^2 = \{\zeta_1 \neq 0, \zeta_2 = \zeta_3 = 0\} \cup \{\zeta_2 \neq 0, \zeta_3 = 0\} \cup \{\zeta_3 \neq 0\}$$
$$= \mathbb{C}^0 \cup \mathbb{C}^1 \cup \mathbb{C}^2.$$

For the general case, the "dumping" procedure may be described as follows. Prior to multiplying on the left by $\omega_{w_m}^{-1}$ (where w_m is the minimal Weyl group element under the partial ordering) the Bruhat cells are described by

$$M_w = [B\omega_w]_B = \{[b'^{w}(z)\omega_w]_B\}, \tag{2.16}$$

where

$$b'^{w}(z) = \mathbf{1} + \sum_{\alpha \in R_+^w} z^\alpha E_\alpha \tag{2.17}$$

and

$$R_+^w = \{\alpha \in R_+ \mid w^{-1}(\alpha) \in R_-\}. \tag{2.18}$$

This follows from the fact that any $b(z')$ of the form

$$b(z') = \mathbf{1} + \sum_{\alpha \in R_+} z'^{\alpha} E_\alpha \tag{2.19}$$

may be factorized:

TABLE I. The Weyl group elements with their SU(3) representatives.

$\mathbf{1}$	$\omega_1 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	(132)	$\omega_{(132)} = \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 1 \\ -1 & 0 & 0 \end{pmatrix}$
(12)	$\omega_{(12)} = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & 0 \\ 0 & 0 & 1 \end{pmatrix}$	(123)	$\omega_{(123)} = \begin{pmatrix} 0 & 0 & 1 \\ -1 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix}$
(23)	$\omega_{(23)} = \begin{pmatrix} 1 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & -1 & 0 \end{pmatrix}$	(13)	$\omega_{(13)} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}$

$$b(z') = b'^{w}(z)\left(\mathbf{1} + \sum_{\alpha\in R_+\setminus R_+^w} z''^{\alpha} E_\alpha\right), \qquad (2.20)$$

for some choice of z^α, z''^α. The second factor may be permuted through ω_w and dumped in the B on the right using the fact that $E_\alpha \omega_w$ is proportional to $\omega_w E_{w^{-1}(\alpha)}$, and $w^{-1}(\alpha)\in R_+$, for $\alpha\in R_+\setminus R_+^w$. This last statement follows from the calculation

$$\begin{aligned}[][H, w_w^{-1} E_\alpha \omega_w] &= \omega_w^{-1}[w(H), E_\alpha]\omega_w \\ &= \alpha(w(H))\omega_w^{-1} E_\alpha \omega_w \\ &= w^{-1}(\alpha)(H)\omega_w^{-1} E_\alpha \omega_w \end{aligned} \qquad (2.21)$$

and the fact that the root spaces $\mathfrak{g}^\alpha$ spanned by E_α are one dimensional so that $\omega_w^{-1} E_\alpha \omega_w$ must be proportional to $E_{w^{-1}(\alpha)}$. After multiplying on the left by $\omega_{\bar{w}}^{-1}$ one has instead

$$M_w = [\omega_{w_m}^{-1} B\, \omega_w]_B = \{[b^{w}(z)\omega_{\bar{w}}]_B\}, \qquad (2.22)$$

where

$$\bar{w} = w_m^{-1} w \qquad (2.23)$$

and

$$b^{w}(z) = \mathbf{1} + \sum_{\alpha\in R_-^w} z^\alpha E_\alpha, \qquad (2.24)$$

where

$$R_-^w = \{w_m^{-1}(\alpha) \mid \alpha\in R_+^w\}. \qquad (2.25)$$

Notice that $\bar{w}_m = \mathbf{1}$ and $R_-^{w_m} = R_-$. Hence in particular the identity element $[\mathbf{1}]_B$ is contained in the largest Bruhat cell M_{w_m}, as required.

Finally we turn to a discussion of coordinates on flag manifolds. It is a general feature of the Bruhat cell decomposition that the closure of the largest cell is the flag manifold itself. Thus the cell coordinates of the largest Bruhat cell provide a coordinatization of nearly all of the flag manifold missing only a lower-dimensional subspace. A complete atlas of coordinate charts is obtained by moving this coordinate patch around by means of left multiplication with the Weyl group representatives ω_w, $w\in W$. It is easily checked that the transition functions are holomorphic on the overlaps and hence flag manifolds are complex manifolds.

There are, of course, other approaches to coordinatizing flag manifolds, e.g., one could start with a coordinatization of G and restrict to an appropriate subset of coordinates to describe M, or one could use global coordinates, i.e., Pluecker coordinates, which are the generalization to flag manifolds of homogeneous as opposed to local coordinates for $\mathbb{C}P^n$. However, "Bruhat coordinates" are the ideal choice to display the complex structure and thus for the description of holomorphic line bundles over M, which will enter in Sec. III. Furthermore, as will be seen in Sec. IV, Bruhat coordinates are particularly appropriate for describing the flows associated with the special "projection" Hamiltonians, which are studied there. Finally Bruhat coordinates are very convenient for integration purposes as the integration region is simply a complex cell. This is because the result is the same if one integrates over a maximal Bruhat cell or over its closure, the manifold itself.

III. KÄHLER STRUCTURES ON FLAG MANIFOLDS

In the previous section it has already been mentioned that flag manifolds possess symplectic and complex structures, arising from the isomorphisms with the adjoint orbit description and with the complexified description. In fact, a stronger statement may be proved, namely, that flag manifolds are Kähler manifolds, admitting G-invariant Kähler metrics. This means that they possess complex local coordinates z_α (with holomorphic transition functions between coordinate patches), a Hermitian Riemannian metric $ds^2 = g_{\alpha\beta}\, dz_\alpha\, d\bar{z}_\beta$ (the Kähler metric), and a corresponding two-form, the Kähler form $\omega = i g_{\alpha\beta}\, dz_\alpha \wedge d\bar{z}_\beta$, which is closed. Thus ω is simultaneously a symplectic form. The G invariance of these structures means invariance under the transformations induced by left G multiplication: $g\cdot[g']_T = [gg']_T$, $\forall g\in G$, $[g']_T\in M = G/T$.

The objective in studying these structures is to obtain expressions for ω and especially for the symplectic volume, $\omega^n/n!$, where $\dim M = 2n$, in terms of the complex Bruhat coordinates from the previous section. The symplectic volume is one of the two pieces of information required for the DH integration formula, which is to be discussed in Sec. V. The other piece of information is the Hamiltonian function, which will be the subject of the next section.

We will follow two different and complementary paths to achieve the above goal. The first path makes use of a detailed analysis of G-invariant geometric structures on flag manifolds by Bordemann, Forger, and Roemer (BFR).[8] By exploiting the G invariance fully, they obtain simple algebraic formulas for the components of various tensors, including ω, at a preferred point of the manifold (the identity). The only task then remaining is to compute expressions for a suitable basis of G-invariant one-forms in terms of the Bruhat coordinates, as the components of ω with respect to this basis will be constant due to G invariance. The second path uses the fact that it is possible to construct in a natural fashion certain holomorphic line bundles over the flag manifold. These "homogeneous vector bundles" were the subject of a deep study by Bott.[17] The main result of interest for our purpose is that, in suitable circumstances, the two-form representing the first Chern class of such a bundle may be identified with a G-invariant Kähler form ω. In addition, Bott's paper gives an expression for the Kähler potential of this Kähler form (the Kähler potential is a function F defined on each coordinate patch such that $\omega = i\,\partial\bar{\partial}F$, where

$$d = \partial + \bar{\partial} = dz_\alpha \frac{\partial}{\partial z_\alpha} + d\bar{z}_\alpha \frac{\partial}{\partial \bar{z}_\alpha})$$

in terms of a local holomorphic section s of the bundle, namely, $F = \ln |s|^2$. The main task to be performed in this approach is the construction of an Hermitian structure $\mathscr{H}$ on the bundle, as this is needed to define the norm squared of a section $[|s|^2 = \mathscr{H}(s,s)]$. Finally, putting several results together, we show that the symplectic volume is given locally by a function $\det g$, which is most easily obtained in the line bundle language as the norm squared of a section of a special "volume bundle."

We now start with the algebraic method of BFR.[8] Consider first the generic case $M = G/T$. Let $\mathfrak{g} = \mathfrak{t} + \mathfrak{n}$ be the decomposition of the Lie algebra of G into torus and nontorus parts. Then there is a correspondence between the nonzero elements of $\mathfrak{n}$ and nonvanishing G-invariant vector fields on M given by $X\in\mathfrak{n}\leftrightarrow\widetilde{X}_m = (d/dt)(\exp tX\cdot m)|_{t=0}$, $\forall m\in M$ (notice that, for $X\in\mathfrak{t}$, the corresponding vector field would vanish at the identity $[1]_T = e$). Thus there is also a correspondence between nonzero elements of the complexification of $\mathfrak{n}$, $\mathfrak{n}_c = \Sigma_{\alpha\in R}\mathfrak{g}^\alpha$, and nonvanishing complexified vector fields on M. In particular, we may identify the tangent space at e, $T_e(M)$ [resp. the complexified tangent space, $T_e^c(M)$] with $\mathfrak{n}$ [resp. $\mathfrak{n}_c$]. BFR's analysis then leads to the following formula for the components of the Kähler form ω at e: let Z^ω lie in the interior of the positive Weyl chamber C_0 and let $\{E_\alpha|\alpha\in R\}$ be the preferred basis of generators of $\mathfrak{n}_c$ (see the Appendix). Then, regarding ω_e as a skew bilinear map $T_e^c(M)\times T_e^c(M)\to\mathbb{C}$, we have

$$\begin{aligned}\omega_e(E_\alpha,E_{-\alpha}) &= i\alpha(Z^\omega),\\ \omega_e(E_\alpha,E_\beta) &= 0, \quad \alpha+\beta\neq 0.\end{aligned}\tag{3.1}$$

[The evaluation of ω on the real vectors corresponding to the set $\{i(E_\alpha+E_{-\alpha}),\ E_\alpha - E_{-\alpha}\}$ follows from bilinearity.] Thus the space of Kähler structures on M is of dimension l, where $l=\operatorname{rank} G$, and is parametrized by the element Z^ω, which may be chosen freely in the interior of the positive Weyl chamber.

One way to understand the formula (3.1) is to regard the flag manifold as an adjoint orbit $\{g\,iZ\,g^{-1}|g\in G\}$, $iZ\in\mathfrak{t}$. Each nondegenerate orbit intersects the interior of iC_0, where C_0 is the positive Weyl chamber, at a single point which one may take to be iZ itself (the adjoint orbit intersects $\mathfrak{t}$ at a discrete set of points, the Weyl group orbit of any of them). Thus iZ corresponds to the identity e of M. At a general point m of the adjoint orbit, the Kirillov symplectic form is given by the following construction: to each vector X at m one associates a Lie algebra element $a(X)$ with the property

$$X=\frac{d}{dt}(\exp(ta(X))m\exp(-ta(X)))\bigg|_{t=0}.\tag{3.2}$$

This correspondence between $\mathfrak{g}$ and vector fields is analogous to the previous one, except that G now acts on M via the adjoint action. Let X,Y be two vectors at m. Then the Kirillov form ω is given by $\omega_m(X,Y) = \langle m,[a(X),a(Y)]\rangle$. [The Kirillov symplectic form is usually associated with the coadjoint orbit $\{g^{-1}\,iZ\,g|g\in G\}$, which is, of course, also an adjoint orbit. The expression given is unaffected by a simultaneous sign change in $a(X)$ and $a(Y)$.] At iZ one can identify the complexified tangent space with $\mathfrak{n}_c$ via the above correspondence and one obtains

$$\begin{aligned}\omega_{iZ}(E_\alpha,E_{-\alpha}) &= \langle iZ,[E_\alpha,E_{-\alpha}]\rangle = \langle[iZ,E_\alpha],E_{-\alpha}\rangle\\ &= \alpha(iZ)\langle E_\alpha,E_{-\alpha}\rangle = i\alpha(Z).\end{aligned}\tag{3.3}$$

Thus the Kähler form (3.1) is displayed as the Kirillov symplectic form of M, regarded as an adjoint orbit intersecting $iC_0\subset\mathfrak{t}$ at iZ^ω.

Before proceeding we briefly discuss the generalization to the nongeneric case. Let $C(T_0)$ be the centralizer of a torus element iZ_0; then the roots in the formula (3.1) range over the subset $\widehat{R} = \{\alpha\in R\,|\alpha(Z_0) = 0\}$. Assuming $\widehat{R}$ is smaller than R, this means that Z_0 lies on at least one of the hyperplanes $\mathscr{V}_\alpha$ that divide $\mathscr{V} = i\mathfrak{t}$ into Weyl chambers. One then has a reduced space $\widehat{\mathscr{V}} = i\hat{\mathfrak{t}}$, where

$$\hat{\mathfrak{t}} = \{X\in\mathfrak{t}|\alpha(X) = 0, \quad \forall\alpha\in R\setminus\widehat{R}\}.$$

Thus $\widehat{\mathscr{V}}$ may be split into chambers as before, and Z^ω in (3.1) must then be chosen in the positive Weyl chamber $\widehat{C}_0$ of $\widehat{\mathscr{V}}$. Taking $G=\mathrm{SU}(3)$ as an example, one could choose $Z_0 = \operatorname{diag}(1,1,-2)$, which lies on $\mathscr{V}_{12}$. Then $R\setminus\widehat{R}$ consists of $\{12,21\}$ and $\widehat{\mathscr{V}}$ is simply $\mathscr{V}_{12}$, which is separated into two chambers by removing the origin.

Returning to the generic case, the task remaining is to construct G-invariant one-forms e^α dual to the G-invariant vector fields $\widetilde{E}_\alpha$ defined above. Furthermore, the e^α are to be expressed in terms of the complex Bruhat coordinates from the previous section. Let U be a Bruhat coordinate patch of M; then, through the "dumping" method described in Sec. II, one has a preferred element $g_c(z)$ representing the coset $[g_c]_B$. Next one performs an Iwasawa decomposition $g_c(z) = g(z)b(z)$ and in this way one has described a map g: $U\to G$. With this map one can pull back the canonical left-invariant one-forms on G to U, and then one takes e^α to be the form multiplying E_α in this expression, i.e.,

$$g^{-1}(z)dg(z) = \sum_{i=1}^{l} h^iH_i + \sum_{\alpha\in R} e^\alpha E_\alpha\tag{3.4}$$

(where $\{H_i\}$ is a basis of $\mathfrak{h}$). Because G is compact, the E_α only appear on the left-hand side in the compact combinations $i(E_\alpha+E_{-\alpha}), E_\alpha - E_{-\alpha}$ and hence the complex one-forms e^α satisfy the constraint $e^\alpha = -\overline{e^{-\alpha}}$. With respect to the left-invariant basis $\{e^\alpha\}$, ω has constant components and thus, from (3.1), ω is given by

$$\begin{aligned}\omega &= \sum_{\alpha\in R_+}\omega_e(E_\alpha,E_{-\alpha})e^\alpha\wedge e^{-\alpha}\\ &= -i\sum_{\alpha\in R_+}\alpha(Z^\omega)e^\alpha\wedge\overline{e^\alpha}.\end{aligned}\tag{3.5}$$

We now proceed to calculate some examples to demonstrate how the above method works in practice. For $\mathrm{SU}(2)/T\cong\mathrm{SL}(2,\mathbb{C})/B$, the Bruhat coordinatization is

$$z\mapsto\left[\begin{pmatrix}1 & 0\\ -z & 1\end{pmatrix}\right]_B = [g_c(z)]_B.\tag{3.6}$$

Perform the Iwasawa decomposition,

$$\begin{pmatrix}1 & 0\\ -z & 1\end{pmatrix} = (1+z\bar z)^{-1/2}\begin{pmatrix}1 & \bar z\\ -z & 1\end{pmatrix}b(z) = g(z)b(z),\tag{3.7}$$

and obtain

$$g^{-1}(z)dg(z) = (1+z\bar z)^{-1}\times\begin{pmatrix}\frac12(\bar z\,dz - z\,d\bar z) & d\bar z\\ -dz & -\frac12(\bar z\,dz - z\,d\bar z)\end{pmatrix}.\tag{3.8}$$

Thus

$$e^{12} = -\overline{e^{21}} = 2(1+z\bar z)^{-1}d\bar z\tag{3.9}$$

(as $E_{12} = \frac12\binom{0\,1}{0\,0}$). It is convenient to choose $Z^\omega = \mu\operatorname{diag}(\frac18,-\frac18), \mu>0$, i.e., $Z^\omega = \mu H_\lambda$, where H_λ corresponds to the basic weight λ of SU(2) via $\langle H_\lambda, Z\rangle = \lambda(Z)$, $\forall Z \in \mathscr{V}$. Then from (3.1) and (3.2) one has

$$\omega = i\mu(1+|z|^2)^{-2}\, dz \wedge d\bar{z}. \qquad (3.10)$$

Second, we consider $SU(3)/T \cong SL(3,\mathbb{C})/B$. The Bruhat coordinatization (2.7) is

$$(z_1,z_2,z_3) \mapsto \left[\begin{pmatrix} 1 & 0 & 0 \\ -z_3 & 1 & 0 \\ -z_2 & z_1 & 1 \end{pmatrix}\right]_B = [g_c(z)]_B. \qquad (3.11)$$

By a Gramm–Schmidt orthonormalization one obtains the SU(3) factor $g(z)$ of the Iwasawa decomposition as the juxtaposition of column vectors (e_1', e_2', e_3') with

$$e_1' = \frac{1}{(K_1)^{1/2}} \begin{pmatrix} 1 \\ -z_3 \\ -z_2 \end{pmatrix},$$

$$e_2' = \frac{1}{(K_1K_2)^{1/2}} \begin{pmatrix} \bar{z}_3 + z_1\bar{z}_2 \\ 1+|z_2|^2 - z_1\bar{z}_2 z_3 \\ z_1 + z_1|z_3|^2 - \bar{z}_3 z_2 \end{pmatrix}, \qquad (3.12)$$

$$e_3' = \frac{1}{(K_2)^{1/2}} \begin{pmatrix} -\bar{z}_1\bar{z}_3 + \bar{z}_2 \\ -\bar{z}_1 \\ 1 \end{pmatrix},$$

where

$$K_1 = 1 + |z_2|^2 + |z_3|^2,$$
$$K_2 = 1 + |z_1|^2 + |z_2 - z_1 z_3|^2. \qquad (3.13)$$

Next calculate the coefficients e^α of E_α in $g^{-1}\,dg$, giving

$$e^{12} = [\sqrt{6}/K_1(K_2)^{1/2}]\,[(1+|z_2|^2 - z_1\bar{z}_2 z_3)d\bar{z}_3 + (z_1 + z_1|z_3|^2 - \bar{z}_3 z_2)d\bar{z}_2],$$
$$e^{13} = [\sqrt{6}/(K_1K_2)^{1/2}]\,[(-\bar{z}_1\,d\bar{z}_3 + d\bar{z}_2), \qquad (3.14)$$
$$e^{23} = [\sqrt{6}/(K_1)^{1/2}K_2]\,[-K_1\,d\bar{z}_1 + (z_3 + \bar{z}_1 z_2)(d\bar{z}_2 - \bar{z}_1\,d\bar{z}_3)],$$

$[E_{ij} = 6^{-1/2}$ (matrix with 1 in ij position and 0 elsewhere)]. To obtain ω, one still needs to select an element Z^ω in the positive Weyl chamber C_0. It is convenient to express Z^ω as

$$Z^\omega = (\mu/18)\mathrm{diag}(2,-1,-1) + (\nu/18)\mathrm{diag}(1,1,-2)$$
$$= \mu H_{\lambda_1} + \nu H_{\lambda_2}, \qquad (3.15)$$

where $\mu,\nu > 0$, to ensure that Z^ω is in the positive Weyl chamber. Here H_{λ_1} and H_{λ_2} correspond to the basic weights λ_1,λ_2 of SU(3). Then, after some calculation, one obtains

$$\begin{aligned}\omega = {} & (i\mu/K_1^2)[(1+|z_3|^2)dz_2\wedge d\bar{z}_2 - \bar{z}_2 z_3\, dz_2 \wedge d\bar{z}_3 \\ & - z_2\bar{z}_3\, dz_3\wedge d\bar{z}_2 + (1+|z_2|^2)dz_3\wedge d\bar{z}_3] \\ & + (i\nu/K_2^2)[K_1\, dz_1\wedge d\bar{z}_1 - (z_3+\bar{z}_1 z_2)dz_1 \\ & \wedge (d\bar{z}_2 - \bar{z}_1\, d\bar{z}_3) \\ & - (\bar{z}_3 + z_1\bar{z}_2)(dz_2 - z_1\, dz_3)\wedge d\bar{z}_1 \\ & + (1+|z_1|^2)(dz_2 - z_1\, dz_3)\wedge(d\bar{z}_2 - \bar{z}_1\, d\bar{z}_3)].\end{aligned} \qquad (3.16)$$

For the purpose of integration one is especially interested in the symplectic volume $\omega^n/n!$. The calculation is made easier by noticing that e^{12} and e^{13} do not have any dz_1 or $d\bar{z}_1$ dependence, thus forcing the contribution from $e^{23}\wedge e^{32}$ to be the $dz_1\wedge d\bar{z}_1$ part only. The expressions simplify considerably, and one obtains

$$\frac{\omega^3}{3!} = \mu\cdot\nu\cdot(\mu+\nu)\frac{1}{(K_1K_2)^2}\, i^3 \prod_{j=1}^{3} dz_j\wedge d\bar{z}_j. \qquad (3.17)$$

As we will see shortly, the line bundle approach leads to a much quicker derivation of this result.

We now turn to the second method of obtaining the geometric quantities we are interested in, namely, via holomorphic line bundles. Let M be $G/T \cong G_c/B$ or $G/C(T_0) \cong G_c/P$ and let χ be a character of T [resp. $C(T_0)$], i.e., a group homomorphism from T [resp. $C(T_0)$] to the circle group U(1). This character extends uniquely to a group homomorphism χ: $B \to \mathbb{C}^*$ (resp. χ: $P \to \mathbb{C}^*$), where the multiplicative group $\mathbb{C}^*$ is the complexification of U(1), i.e., $\mathbb{C}^* = GL(1,\mathbb{C}) = \mathbb{C}\setminus\{0\}$. For example, when $G = SU(2)$ any element $t \in T$ may be written as $t = \mathrm{diag}(e^{ia}, e^{-ia})$. The characters of T are in 1–1 correspondence with the integers: $(\lambda_1)^m(t) = e^{ima}$, for $m \in \mathbb{Z}$. If B is a Borel subgroup of SL(2, $\mathbb{C}$), then for $b \in B$ the characters $(\lambda_1)^m$ extend to $(\lambda_1)^m(b) = (b_{11})^m \in \mathbb{C}^*$.

Given a character χ, and thus its extension to B or P, one obtains a complex line bundle $L_\chi \xrightarrow{\pi} M$, where

$$\begin{aligned} L_\chi &= (G_c \times_\chi \mathbb{C})/B \\ &= \{[(g_c,c)]\,|\, g_c \in G_c,\ c \in \mathbb{C}, \\ &\quad (g_c,c) \sim (g_c b, \chi(b^{-1})c)\}, \end{aligned} \qquad (3.18)$$

and $\pi([(g_c,c)]) = [g_c]_B$. This bundle may be viewed as the line bundle associated to the canonical principal B bundle $G_c \xrightarrow{\pi} M$, where $\pi(g_c) = [g_c]_B$, via the one-dimensional representation of B given by χ. For the nongeneric case, replace B by P.

We remark that the bundles L_χ exhaust the topological possibilities for complex line bundles over $M = G/T$, assuming G is simply connected. These bundles are classified topologically by $H^2(M,\mathbb{Z})$. From the cohomology exact sequence for the exact sequence $0 \to T \to G \to M \to 0$,

$$\cdots \to H^p(T,\mathbb{Z}) \to H^p(G,\mathbb{Z}) \to H^p(M,\mathbb{Z}) \to H^{p-1}(T,\mathbb{Z}) \to \cdots, \qquad (3.19)$$

and, using the assumption $H^1(G,\mathbb{Z}) = 0$ [$H^2(G,\mathbb{Z})$ is always zero], one derives $H^2(m,\mathbb{Z}) \cong H^1(T,\mathbb{Z}) = \mathbb{Z}^l$ and $\mathbb{Z}^l$ corresponds precisely to the lattice of characters of T.

Next we turn to a brief discussion of holomorphic line bundles and holomorphic sections. Rather than developing a general theory, we prefer to illustrate the ideas by the example $SU(2)/T \cong SL(2,\mathbb{C})/B$. SL(2, $\mathbb{C}$)/B is covered by two Bruhat coordinate patches U_1 ($z \mapsto [(\begin{smallmatrix} 1 & 0 \\ z & 1\end{smallmatrix})]_B$) (previously we had $z \mapsto [(\begin{smallmatrix} & 0 \\ -z & 1\end{smallmatrix})]_B$; this change of sign is insignificant) and U_2 ($z' \mapsto [(\begin{smallmatrix} z' & -1 \\ 1 & 0\end{smallmatrix})]_B$). On the overlap $U_1 \cap U_2$ the coordinates are related by $z = (z')^{-1}$. The bundle $L_\chi \xrightarrow{\pi}$ SL(2, $\mathbb{C}$)/B is trivialized over each patch by means of local trivializations h_i: $\pi^{-1}(U_i) \to U_i \times \mathbb{C}$ defined by

$$h_1\left(\left[\left(\begin{pmatrix}1 & 0\\ z & 1\end{pmatrix},c\right)\right]\right)=\left(\left[\begin{pmatrix}1 & 0\\ z & 1\end{pmatrix}\right]_B,c\right),$$

$$h_2\left(\left[\left(\begin{pmatrix}z' & -1\\ 1 & 0\end{pmatrix},c'\right)\right]\right)=\left(\left[\begin{pmatrix}z' & -1\\ 1 & 0\end{pmatrix}\right]_B,c'\right). \quad (3.20)$$

With respect to these local trivializations, the transition function g_{12}: $U_1\cap U_2\to \mathrm{GL}(1,\mathbb{C})$ is defined to be the function satisfying

$$h_1\circ h_2^{-1}([g_c]_B,c')=([g_c]_B,g_{12}c')$$

on $U_1\cap U_2$. If χ is chosen to be $(\lambda_1)^m$, then one finds $g_{12}(z')=z'^m$ or $g_{12}(z)=z^{-m}$. A bundle is said to be holomorphic if the transition functions g_{ij} are holomorphic functions on each overlap $U_i\cap U_j$. Thus for arbitrary m the bundles described above are holomorphic.

Given a subset U of M, a (local) section s is a continuous map s: $U\to\pi^{-1}(U)$ satisfying $\pi\circ s=\mathbf{1}_U$. If $U=M$ we speak of a global section. A section is said to be holomorphic if its representatives f_i with respect to the local trivializations h_i are holomorphic functions on $U\cap U_i$. These representatives are defined by

$$h_i(s([g_c]_B))=([g_c]_B,f_i([g_c]_B)). \quad (3.21)$$

In our example, a holomorphic section s is described by two holomorphic functions $f_1(z)$ and $f_2(z)$, where

$$s\left(\left[\begin{pmatrix}1 & 0\\ z & 1\end{pmatrix}\right]_B\right)=\left[\left(\begin{pmatrix}1 & 0\\ z & 1\end{pmatrix},f_1(z)\right)\right] \quad \text{on } U\cap U_1, \quad (3.22)$$

$$s\left(\left[\begin{pmatrix}z' & -1\\ 1 & 0\end{pmatrix}\right]_B\right)=\left[\left(\begin{pmatrix}z' & -1\\ 1 & 0\end{pmatrix},f_2(z')\right)\right] \quad \text{on } U\cap U_2. \quad (3.23)$$

The two representatives are connected by the transition function g_{12}: in z coordinates, $f_1(z)=g_{12}(z)f_2(z^{-1})$ or, in z' coordinates, $f_1(z'^{-1})=g_{12}(z')f_2(z')$. This poses stringent conditions on the existence and number of global holomorphic sections. In fact, global holomorphic sections exist only if $m\leqslant 0$ [where $g_{12}(z)=z^{-m}$], as f_1 and f_2 must be Taylor series in their respective arguments. By matching the Taylor series, the number of linearly independent global holomorphic sections is easily seen to be $-m+1$.

The above features generalize to arbitrary flag manifolds. In particular, all line bundles over flag manifolds, constructed by means of a character as above, are holomorphic. The statements in our example concerning the existence and number of global sections are generalized by the Borel–Weil theorem (see Ref. 11, Chap. 1, and Ref. 17): let R_+ be the set of positive roots of G with respect to some ordering and let B be the Borel subgroup of G_c with Lie algebra $\mathfrak{b}=\mathfrak{h}+\Sigma_{\alpha\in R_+}\mathfrak{g}^\alpha$. Let L_χ be a line bundle over G_c/B as above. Then the following statements hold.

(a) If χ is not a lowest weight (with respect to the above ordering) of any representation (i.e., if χ is not antidominant), then there are no global holomorphic sections.

(b) If χ is the lowest weight (with respect to the above ordering) of some representation φ, then the space of global sections of L_χ forms an irreducible representation of G with lowest weight χ, and as a consequence there are $(\dim\varphi)$ independent global holomorphic sections.

In order to be able to define the norm squared, $|s|^2$, of a section s of L, one requires a further structure, namely, a Hermitian structure on L. A Hermitian structure on a complex vector space V is a map $\mathscr{H}$: $V\times V\to\mathbb{C}$ satisfying

$$(1)\quad \mathscr{H}(\mu_1v_1+\mu_2v_2,v)=\mu_1\mathscr{H}(v_1,v)+\mu_2\mathscr{H}(v_2,v),$$

$$(2)\quad \overline{\mathscr{H}(v_1,v_2)}=\mathscr{H}(v_2,v_1),\quad \forall v_1,v_2,v\in V,\quad \mu_1,\mu_2\in\mathbb{C}. \quad (3.24)$$

A Hermitian structure on a complex vector bundle is a continuous assignment to each base point of a Hermitian structure on the fiber over that base point. (As an aside, we remark that each fiber of L_χ has the structure of a one-dimensional complex vector space via the following definition of addition and scalar multiplication:

$$[(g_c,c_1)]+\mu[(g_c,c_2)]=[(g_c,c_1+\mu c_2)].)$$

In our examples SU(2)/T and SU(3)/T, a Hermitian structure can be found by inspection. We will discuss a systematic procedure for SU(N)/T shortly. Let $L_{\lambda_1^m}$ be the bundle over SU(2)/T associated to the character $\lambda_1^m(t)=(t_{11})^m$, $\forall t\in T$. [Here λ_1 is the basic weight of SU(2).] Then a Hermitian structure $\mathscr{H}_{\lambda_1^m}(\cdot,\cdot)$ is given by

$$\mathscr{H}_{\lambda_1^m}([(g_c,c_1)],[(k_c,c_2)])=c_1\bar{c}_2(k_c^\dagger g_c)_{11}^m, \quad (3.25)$$

on the fiber over $[g_c]_B=[k_c]_B$. It is a simple matter to check that $\mathscr{H}_{\lambda_1^m}(\cdot,\cdot)$ is well defined and satisfies (1) and (2) above. Similarly, for SU(3)/T we have the bundles $L_{\lambda_1^m\lambda_2^n}$ associated to the character $\lambda_1^m\lambda_2^n$, where λ_1,λ_2 are the basic weights given by $\lambda_1(t)=t_{11}$, $\lambda_2(t)=t_{11}t_{22}$, $\forall t\in T$. Then a Hermitian structure on $L_{\lambda_1^m\lambda_2^n}$ is given by

$$\mathscr{H}_{\lambda_1^m\lambda_2^n}([(g_c,c_1)],[(k_c,c_2)])$$
$$=c_1\bar{c}_2(k_c^\dagger g_c)_{11}^m\left(\sum_j(\hat{g}_c)_{j3}\,\overline{(\hat{k}_c)_{j3}}\right)^n, \quad (3.26)$$

on the fiber over $[g_c]_B=[k_c]_B$, where $(\hat{g}_c)_{j3}$ is the determinant of the minor of g_c with row j and column 3 deleted. This choice is suggested by the properties

$$(g_cb)_{i1}=(g_c)_{i1}\lambda_1(b)\quad\text{and}\quad(\widehat{g_cb})_{j3}=(\hat{g}_c)_{j3}\lambda_2(b),$$

which ensure that $\mathscr{H}_{\lambda_1^m\lambda_2^n}(\cdot,\cdot)$ is well defined.

Following Bott,[17] the Chern form c_χ of the bundle L_χ is given, after choosing a nonvanishing holomorphic section s_i in each coordinate patch U_i, by

$$c_\chi=(i/2\pi)\partial\bar\partial\ln|s_i|_\chi^2\equiv(i/2\pi)\partial\bar\partial\ln\mathscr{H}_\chi(s_i,s_i). \quad (3.27)$$

The actual choice of nonvanishing section in each patch is irrelevant as, if $s_1=fs_1'$ with f a holomorphic function, we have

$$i\,\partial\bar\partial\ln|s_1'|_\chi^2-i\,\partial\bar\partial\ln|s_1|_\chi^2=i\,\partial\bar\partial\ln f\bar f=0. \quad (3.28)$$

By a similar argument one shows that the above patchwise construction of c_χ leads to a globally defined form. The main result of interest for our purpose, Proposition 10.1 in Bott,[14] may be phrased as follows: $2\pi c_\chi$ may be identified with a Kähler form ω (i.e., the components $g_{\alpha\beta}$ of $\omega=ig_{\alpha\beta}\,dz_\alpha\wedge d\bar{z}_\beta$ form a positive definite matrix), if χ lies in the positive Weyl chamber. We proceed to verify this result in our examples SU(2)/T, SU(3)/T: choosing the section

$$s([g_c(z)]_B) = [(g_c(z),1)], \tag{3.29}$$

where $g_c(z)$ is the coordinatized representative of the coset $[g_c(z)]_B$ in (3.6) [resp. (3.11)], and using (3.25) [resp. (3.26)] one obtains, for $SU(2)/T$,

$$\omega = i\,\partial\bar\partial \ln(1+|z|^2)^m, \quad m>0, \tag{3.30}$$

and, for $SU(3)/T$,

$$\omega = i\,\partial\bar\partial \ln\,(1+|z_3|^2+|z_2|^2)^m(1+|z_1|^2+|z_2-z_1z_3|^2)^n,$$
$$m,n>0. \tag{3.31}$$

A short calculation shows that these results agree with (3.10) and (3.16) if we set $\mu = m$ (resp. $\mu = m$, $\nu = n$).

It is instructive to reflect for a moment on the relationship between the two approaches we have used to obtain ω. The input in the BFR approach was an element of the positive Weyl chamber (continuous parameters), whereas in the line bundle approach, the input was a character in the positive Weyl chamber (discrete parameters). Thus the line bundle approach picks out a special discrete subset of all the possible ω's; the relationship between the approaches for this subset may be stated concisely as follows: if ω is obtained by the line bundle approach using the character χ, then the same ω is obtained from the BFR method using $Z^{\omega} = H_\chi$.

Up to now, in the line bundle approach, we have only defined Hermitian structures by inspection in two special cases. In order to obtain systematic results from the line bundle approach, we must first discuss the notion of the "product of two bundles." Given two bundles (either principal or vector bundles) with transition functions g_{ij} (resp. g'_{ij}) on the overlaps $U_i \cap U_j$ of open sets $\{U_i\}$ covering the base space, it is natural to define the product of the two bundles as the bundle that has transition functions $g''_{ij} = g_{ij}g'_{ij}$ on $U_i\cap U_j$. Clearly the product bundle obtained in this way is not uniquely prescribed, and thus the product is really a product of isomorphism classes of bundles. The line bundles L_χ we have been considering form a group with respect to this product: if the representatives of $[g_c]_B$ on U_i (resp. U_j) are connected by $b_{ij}(z)\in B$, then the transition function over $U_i\cap U_j$ is $\chi(b_{ij}(z))$; thus by the multiplicative properties of characters, $L_\chi = L_{\chi_1}L_{\chi_2}$, for $\chi = \chi_1\chi_2$.

In general, unlike in the above example, an explicit description of the product bundle is lacking. However, for principal U(1) bundles, a construction has been given by Kobayashi[18]: let P_1 and P_2 be two principal U(1) bundles over M, with projections π_1 (resp. π_2); then define $P_1P_2 = \Delta(P_1\times P_2)/\sim$, where

$$\Delta(P_1\times P_2) = \{(p_1,p_2)\in P_1\times P_2 | \pi_1(p_1) = \pi_2(p_2)\}$$

and $(p_1,p_2)\sim(u\cdot p_1, u^{-1}\cdot p_2)$, where $\cdot$ is the action of U(1) on P_i. The projection to M is the restriction to either factor followed by projection with the appropriate π_i, and the U(1) action on P_1P_2 is defined by $u\cdot[(p_1,p_2)] = [(u\cdot p_1, p_2)]$.

To each principal U(1) bundle P over M there is associated a complex line bundle L over M by the usual construction $L = (P\times\mathbb{C})/\sim$, where $(p,c)\sim(u\cdot p, u^{-1}c)$, $\forall u\in U(1)$. The vector space structure of addition and scalar multiplication is defined on each fiber by $[(p,c_1)] + \mu[(p,c_2)] = [(p, c_1+\mu c_2)]$. Analogously to Kobayashi's procedure,[18] we define the product L_1L_2 of two associated line bundles as $L_1L_2 = \Delta(L_1\times L_2)/\sim$, where

$$\Delta(L_1\times L_2) = \{(v_1,v_2)\in L_1\times L_2|\pi_1(v_1) = \pi_2(v_2)\}$$

and $(v_1,v_2)\sim(zv_1, z^{-1}v_2)$, $\forall z\in\mathbb{C}^*$. Then L_1L_2 is isomorphic to the line bundle associated to P_1P_2, i.e., $(P_1P_2\times\mathbb{C})/\sim$ via the identification

$$[([(p_1,c_1)],[(p_2,c_2)])]\leftrightarrow[([(p_1,p_2)],c_1c_2)].$$

Through this identification the vector space structure on L_1L_2 is given by

$$[([(p_1,c_1)],[(p_2,c_2)])] + \mu[([(p_1,c'_1)],[(p_2,c'_2)])]$$
$$= [([(p_1,c''_1)],[(p_2,c''_2)])],$$

where $c''_1c''_2 = c_1c_2 + \mu c'_1c'_2$. This vector space structure may also be defined directly in terms of the vector bundles by

$$[(v_1,v_2)] + \mu[(v_1,v'_2)] = [(v_1, v_2+\mu v'_2)].$$

Given Hermitian structures $\mathscr{H}_1(\cdot,\cdot)$, $\mathscr{H}_2(\cdot,\cdot)$ on L_1, L_2, respectively, a Hermitian structure $\mathscr{H}$ on L_1L_2 is obtained by the following definition:

$$\mathscr{H}([(v_1,v_2)],[(v'_1,v'_2)]) = \mathscr{H}_1(v_1,v'_1)\mathscr{H}_2(v_2,v'_2). \tag{3.32}$$

It is easily checked that $\mathscr{H}$ is well defined and satisfies the properties (1) and (2) from before. Thus, given the factorization of a general line bundle in terms of a set of basic line bundles, the problem of finding a Hermitian structure on the general line bundle is reduced to that of finding Hermitian structures on the basic line bundles.

For the case of the bundles L_χ over G_c/B one first establishes that $L_{\chi_1}L_{\chi_2} = L_{\chi_1\chi_2}$ in the above sense also (by checking that the correspondence

$$[([(g_c,c_1)]_{\chi_1},[(g_c,c_2)]_{\chi_2})]\leftrightarrow[(g_c,c_1c_2)]_{\chi_1\chi_2}$$

is an isomorphism), and thus it is sufficient to find Hermitian structures for the bundles L_{λ_i}, where the λ_i are the basic weights. For the examples $SU(2)/T$ and $SU(3)/T$ this was done by inspection, and formulas (3.25) and (3.26) give the Hermitian structure according to the rule (3.32) for a general weight in terms of the Hermitian structures for the basic weights [being the expressions (3.25) and (3.26) with $m=1$ (resp. $m=1$, $n=0$ and $m=0$, $n=1$)].

Next we look at a different way of viewing the basic line bundles over $M = SU(N)/T$ which leads to a systematic method of obtaining a Hermitian structure for this general case. First, we discuss the determinant bundle of a Grassmannian: let $G(k,n)$ be the Grassmannian of complex k-planes in $\mathbb{C}^n$; thus each point of $G(k,n)$ is a k-plane, which we assume to be spanned by the k independent vectors $e_1,\ldots,e_k$ in $\mathbb{C}^n$. There is a natural k-plane bundle over $G(k,n)$ which has as fiber over each point the k-plane represented by that point. Thus a general point of the total space E_k of this bundle is (e', V_k), where e' is an arbitrary vector in the k-plane V_k. Now the determinant bundle of this k-plane bundle is by definition $\det = \wedge^k E_k$, i.e., the k-fold antisymmetric tensor product of E_k. Thus a general point of det is $(e'_1\wedge e'_2\wedge\cdots\wedge e'_k, V_k)$, where $e'_i\in V_k$, $\forall i$, and $\wedge$ is the antisymmetric product satisfying $e_1\wedge e_2 = -e_2\wedge e_1$. It is easy to check that, for any choice of $\{e'_i\}$, $e'_1\wedge\cdots\wedge e'_k$ (which is called a k vector) is proportional to the basic k vector $e_1 \wedge \cdots \wedge e_k$. Thus det is a complex line bundle. Now, following Chern,[19] we define a Hermitian structure on det by

$$\mathscr{H}_{\det}((e_1' \wedge \cdots \wedge e_k', V_k),(e_1' \wedge \cdots \wedge e_k'', V_k)) = \det[(e_i', e_j'')], \tag{3.33}$$

where the rhs denotes the determinant of the $k \times k$ matrix with (ij)th element (e_i', e_j''), and (,) is the canonical Hermitian structure on $\mathbb{C}^n$: $(e_1, e_2) = e_1^T \bar{e}_2$.

This construction may be extended naturally to the flag manifolds $M = SU(N)/T = G(1,2,...,N)$. A general point of M is a one-plane C_1 contained in a two-plane C_2,..., contained in a $(N-1)$-plane $C_{N-1} \subset \mathbb{C}^N$. Thus, for $1 \leqslant k \leqslant N-1$, there exists a k-plane bundle E_k over M with general point $(e', C_1 \subset C_2 \subset \cdots \subset C_{N-1})$, where $e' \in C_k$. The determinant bundle $\wedge^k E_k$ we will denote $\det_k$. A Hermitian structure $\mathscr{H}_{\det_k}$ on each $\det_k$ is defined as above, apart from replacing C_k by the flag $C_1 \subset \cdots \subset C_{N-1}$.

We now show that $\det_k$ is precisely the bundle L_{λ_k} corresponding to the k th basic weight λ_k. Suppose on two overlapping coordinate patches the representatives $[g_c(z)]_B$ and $[g_c(z')]_B$ are related by $g_c(z) = g_c(z')b(z')$. The fiber over $[g_c]_B$ in E_k is spanned by the first k columns of g_c and thus the transition function between the two patches is the principal $k \times k$ minor of $b(z')$. The transition function in $\det_k$ is the determinant of this minor, i.e.,

$$\prod_{i=1}^{k} b_{ii}(z'),$$

which corresponds to the character of T defined by

$$\chi_k(t) = \prod_{i=1}^{k} t_{ii}, \quad \forall t \in T.$$

To see that χ_k is the basic weight λ_k, we go to the additive language and use formula (A9), which expresses the basic weights in terms of the simple roots. For convenience we replace $\langle \cdot,\cdot \rangle$ by $\operatorname{tr}(\cdot,\cdot)$:

$$\delta_{ij} = 2 \operatorname{tr}(H'_{\lambda_i}, H'_{\lambda_j})/\operatorname{tr}(H'_{\alpha_j}, H'_{\alpha_j}), \tag{3.34}$$

where $\lambda_i(Z) = \operatorname{tr}(H'_{\lambda_i}, Z)$, $\forall Z \in \mathscr{V}$, and similarly for α_j. The roots are $ij = \varepsilon_i - \varepsilon_j$, where $\varepsilon_i(Z) = Z_{ii}$. Choosing as before the canonical ordering $ij > 0 \leftrightarrow i < j$, the simple roots are $\alpha_i = i \quad i+1$, $i = 1,...,N-1 = l$. Thus $H'_{\alpha_i} = \operatorname{diag}(0,...,0,1,-1,0,...,0)$ (one in the ith position). Hence, by a simple calculation,

$$H'_{\lambda_j} = (1/N)\operatorname{diag}(N-j,...,N-j,-j,...,-j) \tag{3.35}$$

(with j times $N-j$ and $N-j$ times $-j$). From this we obtain the action of λ_j acting on Z, namely,

$$\operatorname{tr}(H'_{\lambda_j}, Z) = \sum_{i=1}^{j} Z_{ii}.$$

Thus, in the multiplicative language,

$$\lambda_j(t) = \prod_{i=1}^{j} t_{ii}, \quad t \in T.$$

Given a choice of representative $g_c(z)$ of $m = [g_c(z)]_B$ on some coordinate patch, it is convenient to define the standard section s_0 of $\det_k$ by

$$s_0([g_c(z)]_B) = e_1 \wedge \cdots \wedge e_k, \tag{3.36}$$

where e_i = (ith column of $g_c(z)$). In general, if $L_\chi = \Pi_{k=1}^{l} (\det_k)^{m_k}$, then define the standard section s_0 of L_χ to be the product of standard sections in each of the factors. From the product rule (3.32) for Hermitian structures one has

$$|s_0|_\chi^2 = \prod_{k=1}^{l} (\mathscr{H}_{\det_k}(e_1 \wedge \cdots \wedge e_k, e_1 \wedge \cdots \wedge e_k))^{m_k}. \tag{3.37}$$

As an application, we calculate the Kähler potential arising from the line bundle

$$L_{\lambda_1^{m_1}\lambda_2^{m_2}\lambda_3^{m_3}} = (\det_1)^{m_1}(\det_2)^{m_2}(\det_3)^{m_3}$$

over $SU(4)/T$. Choose the standard section s_0 determined by the representative

$$g_c(z) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ z_1 & 1 & 0 & 0 \\ z_2 & z_4 & 1 & 0 \\ z_3 & z_5 & z_6 & 1 \end{pmatrix}. \tag{3.38}$$

To simplify the calculations we may use the assertion after 7.5.3 in Ref. 11, which states that for an $n \times k$ matrix $\mathscr{M}$ with columns $e_1 \cdots e_k$,

$$\det((e_i, e_j)) = \sum_m |\det m|^2, \tag{3.39}$$

where m ranges over all $k \times k$ minors of $\mathscr{M}$. Thus, putting the results together, we arrive at the Kähler potential:

$$\begin{aligned} F = \ln[&(1 + |z_1|^2 + |z_2|^2 + |z_3|^2)^{m_1} \\ &\times (1 + |z_4|^2 + |z_5|^2 + |z_1 z_4 - z_2|^2 \\ &+ |z_1 z_5 - z_3|^2 + |z_2 z_5 - z_3 z_4|^2)^{m_2} \\ &\times (1 + |z_6|^2 + |z_4 z_6 - z_5|^2 \\ &+ |z_1(z_4 z_6 - z_5) - (z_2 z_6 - z_3)|^2)^{m_3}]. \end{aligned} \tag{3.40}$$

Now we turn to a calculation of the symplectic volume in the line bundle approach. Let $\dim M = 2n$ and let ω be expressed in some coordinate patch as $\omega = i g_{\alpha\beta}\, dz_\alpha\, dz_\beta$. Then the symplectic volume, $\omega^n/n!$, in the same patch is given by

$$\omega^n/n! = (\det g) i^n \prod_{\gamma=1}^{n} dz_\gamma \wedge d\bar{z}_\gamma. \tag{3.41}$$

Thus we seek an expression for $\det g$. To this end, recall first the classical result that the "Kähler potential" for the Ricci tensor Ric is $-\ln(\det g)$, i.e.,

$$\mathrm{Ric} = R_{\alpha\beta}\, dz_\alpha\, d\bar{z}_\beta, \tag{3.42}$$

where

$$R_{\alpha\beta} = \frac{\partial^2}{\partial z_\alpha\, \partial \bar{z}_\beta}(-\ln(\det g)). \tag{3.43}$$

Equivalently this may be expressed in terms of the Ricci form ρ, a two-form that plays the same role for Ric as the Kähler form ω does for g:

$$\rho = i\, \partial\bar{\partial}(-\ln(\det g)), \tag{3.44}$$

where

$$\rho = i R_{\alpha\beta}\, dz_\alpha \wedge d\bar{z}_\beta. \tag{3.45}$$

On the other hand, in Ref. 8, BFR showed the remarkable result that for the G-invariant Kähler metrics under consideration here, the Ricci tensor and Ricci form are independent of the particular metric chosen. Because of G invariance, one again only needs to specify the components at the identity e by a formula analogous to (3.1):

$$\begin{aligned} \rho_e(E_\alpha,E_{-\alpha}) &= i\alpha(Z^\rho), \\ \rho_e(E_\alpha,E_\beta) &= 0, \quad \alpha+\beta\neq 0, \end{aligned} \tag{3.46}$$

where Z^ρ is a particular element of C_0 given by

$$Z^\rho = \sum_{\alpha\in R_+} H_\alpha. \tag{3.47}$$

Thus Z^ρ corresponds to the weight ρ, which is the sum of the positive roots or, equivalently, twice the sum of the basic weights. We remark that this weight plays a special role in representation theory, e.g., in the Weyl character formula. Now transferring to the line bundle approach, we may express ρ in terms of a nonvanishing section s of L_ρ:

$$\rho = i\,\partial\bar\partial \ln|s|^2_\rho. \tag{3.48}$$

Combining this with (3.42), we arrive at the expression

$$\det g = \varepsilon/|s_0|^2_\rho, \tag{3.49}$$

where ε is a constant to be determined from (the scale of) $g_{\alpha\beta}$, and where we have chosen s to be the standard section s_0, which is nonvanishing. [If we were to take instead $s = fs_0$, where $f(z)$ is a nonvanishing holomorphic function, then the fact that $\det g$ is a function on all of M means that $f(z)$ extends to a holomorphic function on all of M, i.e., $f(z)$ is constant.] A natural choice for the metric is the Einstein metric $g = \mathrm{Ric}$, $\omega = \rho$; in this case, by comparing ω and ω^n at the identity, one obtains

$$\varepsilon = \prod_{\alpha\in R_+}\sum_{\beta\in R_+}\left(\frac{\langle\alpha,\beta\rangle}{\langle\alpha,\alpha\rangle}\right). \tag{3.50}$$

Making this choice in the example of $\mathrm{SU}(3)/T$, the symplectic volume is given by

$$\frac{\omega^3}{3!} = \frac{2\cdot 2\cdot 4}{(K_1K_2)^2}\, i^3 \prod_{\gamma=1}^{3} dz_\gamma\wedge d\bar z_\gamma, \tag{3.51}$$

in agreement with (3.17) with $\mu = \nu = 2$.

In the language of determinant bundles, if we define the "volume bundle" $\mathrm{Vol} = \Pi_k\,(\det_k)^{-2}$, then we may phrase the result (3.47) as follows: $\det g$ is the norm squared of the standard section of Vol.

To illustrate how this approach works for nongeneric flag manifolds we discuss the case of $\mathbb{C}P^n = \mathrm{SU}(n+1)/S(\mathrm{U}(1)\times\mathrm{U}(n))$, where $S(U(1)\times U(n))$ is the centralizer of

$$T_0 = \{\mathrm{diag}(e^{ina},e^{-ia},\ldots,e^{-ia})\,|\,a\in\mathbb{R}\}$$

and thus consists of unitary $(1+n)\times(1+n)$ block diagonal matrices. In the complexified description $M = \mathrm{SL}(n+1,\mathbb{C})/P$, where the parabolic subgroup P is the subgroup of $\mathrm{SL}(n+1,\mathbb{C})$ consisting of $p = [p_{ij}]$ with $p_{i1} = 0$, $i>1$. The Bruhat coordinatization is then

$$(z_1,\ldots,z_n)\mapsto\left[\begin{pmatrix} 1 & 0 & 0 & \cdots & 0 \\ z_1 & & & & \\ z_2 & & 1_n & & \\ \vdots & & & & \\ z_n & & & & \end{pmatrix}\right]_P = [g_c(z)]_P. \tag{3.52}$$

Again we construct a line bundle L_χ over M using a character χ of $C(T_0)$ extended to a character of P. There is only one choice of character in this case, namely, $\chi(t) = \lambda_1(t) = t_{11}$, $\forall t\in T$ [resp. $\chi(p) = p_{11}$, $\forall p\in P$], up to multiplicity. Interpreting $g_c(z)$ as defining the standard section s_0 as before, we obtain, for the Kähler potential on $L_{\lambda_1^m}$,

$$F = m\ln\left(1+\sum_{i=1}^n |z_i|^2\right). \tag{3.53}$$

To obtain the symplectic volume we must first find the "λ_1 content" of the special character $\rho = 2$ (sum of basic weights). More accurately we write $\rho = m_1\lambda_1 +$ (weights that do not affect t_{11}). A simple approach to finding the coefficient m_1 is to notice that

$$2\sum_i (H'_{\lambda_i})_{11} = m_1(H'_{\lambda_1})_{11}, \tag{3.54}$$

and, using formula (3.35) for H'_{λ_j}, we easily find $m_1 = n+1$. The factor $\det g$ multiplying the coordinate measure in (3.41) now becomes

$$\det g = \frac{\varepsilon}{|s_0|^2_{\lambda_1^{n+1}}} = \frac{\varepsilon}{(1+\Sigma_{i=1}^n|z_i|^2)^{n+1}}, \tag{3.55}$$

leading to the expression for the symplectic volume of $\mathbb{C}P^n$ with $\omega = i\,\partial\bar\partial\ln F$ given by (3.53):

$$\frac{\omega^n}{n!} = m^n\prod_{j=1}^n \frac{(i\,dz_j\wedge d\bar z_j)}{(1+\Sigma_{k=1}^n|z_k|^2)^{n+1}}. \tag{3.56}$$

Finally, we briefly recapitulate the main points of this section. The algebraic method of BFR leads to a family of G-invariant Kähler structures on M parametrized by the positive Weyl chamber C_0. The components of the Kähler form ω with respect to a G-invariant basis of one-forms are constant and are given by the algebraic expression (3.1). The G-invariant one-forms are found in terms of the Bruhat coordinates by a somewhat laborious process involving an Iwasawa decomposition. In the line bundle approach, one looks instead for the Kähler potential F of ω and finds that e^F is the norm squared, $|s|^2$, of a section of the line bundle L_χ that is twisted by means of the (dominant) character χ of G. One still has to do some work to find the Hermitian structure $\mathscr{H}_\chi(.,.)$ on L_χ needed to define $|s|^2$. By defining a product of line bundles, we were able to reduce the problem to finding Hermitian structures on the basic weight bundles L_{λ_i}. In the important generic case of $M = \mathrm{SU}(N)/T$, N arbitrary, this problem was solved using a geometric description of the basic weight bundles as generalized determinant bundles. Finally, to achieve an expression for the symplectic volume $\omega^n/n!$, the most efficient method was to calculate the function $\det g$, which is proportional to the norm squared of the standard section of the "volume bundle" Vol

$= \Pi_k\,(\det_k)^{-2} = L_{-\rho}$ [where $\rho = 2$ (sum of basic weights)].

IV. PROJECTION HAMILTONIANS ON FLAG MANIFOLDS

From the dynamical system point of view symplectic manifolds constitute a (possibly reduced) phase space of some dynamical system. The dynamics is determined by specifying a Hamiltonian function H on the manifold, which leads to a Hamiltonian vector field rot_H via

$$dH(X) = \omega(X,\mathrm{rot}_H), \quad \forall X\in TM \tag{4.1}$$

(rot is an abbreviation of rotation). In this section we will be studying certain preferred Hamiltonian functions on flag manifolds, which will be termed "projection Hamiltonians." The reason for this terminology is that, after suitable rescaling, these Hamiltonians are obtained geometrically by projecting the flag manifold, viewed as an adjoint orbit embedded in $\mathfrak{g}$, onto a line in the $\mathfrak{t}$ hyperplane. Specifically, let M be the flag manifold $M = \{giZ^\omega g^{-1}|g\in G\}$, $iZ^\omega\in\mathfrak{t}$ with the corresponding symplectic form ω [cf. (3.1)] and let $iZ^H\in\mathfrak{t}$ be such that Z^H lies in C_0, the positive Weyl chamber. The corresponding projection Hamiltonian is then given by

$$H(giZ^\omega g^{-1}) = \langle iZ^H, giZ^\omega g^{-1}\rangle. \tag{4.2}$$

In the simplest example of $M = \mathrm{SU}(2)/T$ the flag manifold is viewed as a two-sphere in $\mathrm{su}(2)\cong\mathbb{R}^3$ intersecting the positive $\sigma_3/2i$ axis in iZ^ω (which specifies the symplectic form on M), and the projection Hamiltonians are (up to an overall scale determined by iZ^H) simply projection onto the $\sigma_3/2i$ axis.

It is also instructive to consider an example with a two-dimensional torus such as $M = \mathrm{SU}(3)/T$ and to evaluate H in two stages: first, orthogonal projection (with respect to $\langle\cdot,\cdot\rangle$) onto $\mathfrak{t}$, and then application of the linear function $\langle iZ^H,\cdot\rangle$, regarded as an element of $\mathfrak{t}^*$. This gives the same result as (4.2). After the intermediate stage of projection a remarkable phenomenon may be observed: the image of M under orthogonal projection is the convex hull of the points in the Weyl orbit of iZ^ω (which are the only points where the flag manifold actually intersects $\mathfrak{t}$). As shown by Heckman (see Ref. 20, Sec. 2.5) and Atiyah (see Ref. 15, Theorem 3), the Bruhat cells $M_w = [B\omega_w]_B$, $w\in W$, of Sec. II also have nice projections onto $\mathfrak{t}$: the image of M_w is the convex hull of $\{w'(iZ^\omega)|w'\geqslant w\}$. For $\mathrm{SU}(3)/T$ the situation is sketched in Fig. 2, which is taken from Ref. 15. The Weyl group acts, as

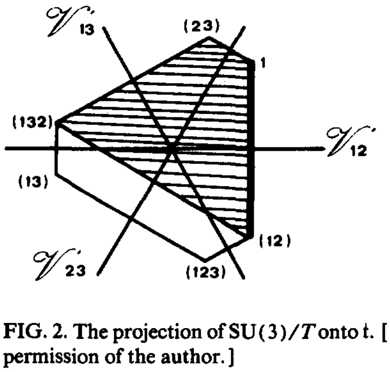

FIG. 2. The projection of $\mathrm{SU}(3)/T$ onto $\mathfrak{t}$. [Reproduced from Ref. 15 with permission of the author.]

in the Appendix, through the isomorphism $\mathfrak{t}\cong i\mathfrak{t} = \mathscr{V}$ and the lines $\mathscr{V}'_\alpha$ correspond to $\mathscr{V}_\alpha$ under this isomorphism. The vertices of the hexagon are labeled by Weyl group elements, where w is shorthand for $w(iZ^\omega)$. The shaded region is the projection of the complex two-cell $M_{(132)}$ and the double line is the projection of $M_{(12)}$.

The rotation flow (or Hamiltonian flow) of the projection Hamiltonians H, i.e., the flow of which rot_H is the tangent vector field, has an extremely simple description in terms of the Bruhat coordinates of M. We will show the following statements about this flow and the related gradient flow (expanding on the rather concise treatment in Ref. 15):

(a) the rotation flow of H corresponds to a rotation $z\mapsto e^{ict}z$ for each Bruhat coordinate z, where c is a constant depending on iZ^H and on the choice of coordinate z [this rotation flow may be described as the left action of $\exp(iZ^H t)$ on G_c/B, (b) the gradient vector field of H, defined by

$$dH(X) = g(X,\mathrm{grad}_H), \quad \forall X\in TM, \tag{4.3}$$

is tangent to the gradient flow described by left action on G_c/B with $\exp(-Z^H t)$; and (c) each Bruhat cell M_w contains precisely one stationary point of H and M_w is the unstable manifold of this stationary point under the gradient flow of H.

We show (a) by showing first that the action of $\exp(iZ^H t)$ on G_c/B is linear in each Bruhat cell. Let $M_w\subset G_c/B$ be coordinatized as in Sec. II, i.e.,

$$M_w = \{[b^w(z)\omega_{\bar w}]_B\}, \tag{4.4}$$

where $\bar w = w_m^{-1}w$ and

$$b^w(z) = 1 + \sum_{\alpha\in R^w_-} z^\alpha E_\alpha. \tag{4.5}$$

The left action of $\exp(iZ^H t)$ on an element of M_w may be replaced by the adjoint action of $\exp(iZ^H t)$ on the corresponding matrix $b^w(z)$, as any torus element may be permuted through $\omega_{\bar w}$ and dumped in B. Then the action on the coordinates z^α is $z^\alpha\mapsto\exp(\alpha(iZ^H)t)\,z^\alpha$, i.e., a linear action. Now we must still show that the left action of $\exp(iZ^H t)$ on G_c/B corresponds to the rotation flow of H. Left multiplication by $\exp(iZ^H t)$ in the G_c/B description of M corresponds to adjoint action of $\exp(iZ^H t)$ in the adjoint orbit description of M, because the Iwasawa decomposition of $\exp(iZ^H t)g_c = (\exp(iZ^H t)\underset{\sim}{g})b$. Denote the derivative vector field of this flow $\widetilde{\mathrm{rot}}$, i.e., at $m\in M = \{giZ^\omega g^{-1}|g\in G\}$,

$$\mathrm{rot}_m = \frac{d}{dt}\exp(iZ^H t)m\exp(-iZ^H t)\big|_{t=0}. \tag{4.6}$$

Thus, in the notation introduced after (3.1), $a(\widetilde{\mathrm{rot}}) = iZ^H$ at m. Now we calculate, from (4.1) and (4.2),

$$\begin{aligned} dH(X) &= d\,\langle giZ^\omega g^{-1},\ iZ^H\rangle(X)\\ &= \langle[giZ^\omega g^{-1},\ g\,dg^{-1}],\ iZ^H\rangle(X)\\ &= \langle giZ^\omega g^{-1},[g\,dg^{-1}(X),\ iZ^H]\rangle\\ &= \langle giZ^\omega g^{-1},[a(X),\ iZ^H]\rangle, \end{aligned} \tag{4.7}$$

where the final equality derives, by pullback, from the equation on the group manifold

$$gdg^{-1}(\tilde{X}) = X, \quad \forall X \in g, \tag{4.8}$$

where $\tilde{X}$ is the vector field

$$\tilde{X}_g = \frac{d}{dt}(\exp tX \cdot g)|_{t=0}. \tag{4.9}$$

On the other hand, we have, from the expression for the Kirillov form,

$$\begin{aligned} dH(X) &= \omega(X, \mathrm{rot}_H) \\ &= \langle giZ^\omega g^{-1}, [a(x), a(\mathrm{rot}_H)] \rangle, \end{aligned} \tag{4.10}$$

and thus $iZ^H = a(\mathrm{rot}_H)$ also, showing the equality of $\widetilde{\mathrm{rot}}$ and rot_H.

To show (b) we observe that the isomorphism $G/T \cong G_c/B$ defines an almost complex structure J on M, i.e., a smoothly varying, linear automorphism of $T_m M$ satisfying $J^2 = -1$ (J corresponds to multiplication by i in the tangent space). In the present case the complex structure is most easily described using the obvious left G_c action on G_c/B: if the vector $\tilde{X}$ at $[g_c]_B$ is given by

$$\tilde{X} = \frac{d}{dt} \exp(tX) \cdot [g_c]_B|_{t=0},$$

where $X \in g_c$, then

$$JX = \frac{d}{dt} \exp(itX) [g_c]_B|_{t=0}.$$

Now for flag manifolds we have

$$\mathrm{grad}_H = J \,\mathrm{rot}_H, \tag{4.11}$$

because of the relationship between symplectic form and metric

$$\omega(X,Y) = g(X,JY), \quad \forall X,Y \in TM. \tag{4.12}$$

Thus grad_H is tangent to the flow of $\exp(-Z^H t)$ acting on G_c/B. This flow is a linear flow when expressed in the Bruhat coordinates by exactly the same argument as in (a). It does not, however, have a natural description on the adjoint orbit version of M as the Iwasawa decomposition of $\exp(-Z^H t) g_c$ is no longer straightforward.

For (c) we just need to observe that any z^α coordinatizing M_w gets multiplied by the factor $\exp(-\alpha(Z^H t))$, which tends to zero as $t \to -\infty$ by the assumption in (4.2) that $-\alpha(Z^H)$ is positive for all negative roots α. Thus on M_w the gradient flow converges to the point $[\omega_{\bar{w}}]_B$ with all coordinates z^α zero. In other words, M_w is the unstable manifold of $[\omega_{\bar{w}}]_B$ under the gradient flow. Clearly on M_w the gradient flow is only trivial at $[\omega_{\bar{w}}]_B$ and thus grad_H and dH vanish at this point and nowhere else on M_w.

When one wishes to derive explicit expressions for the projection Hamiltonians H, in a particular case, one is faced with the problem that H is given in terms of the adjoint orbit description of M, whereas the Bruhat coordinates come from the complex description $M \cong G_c/B$. Thus one has to perform an Iwasawa decomposition in order to express $giZ^\omega g^{-1}$ in (4.2) in terms of the Bruhat coordinates. For the cases $M = \mathrm{SU}(2)/T$, $\mathrm{SU}(3)/T$, this was done in the previous section. However, it is clear that for larger groups the Iwasawa decomposition becomes an increasingly complicated task. For this reason we will proceed to derive in the remainder of this section an expression for H based on complex line bundles over $M \cong G_c/B$ (the "line bundle approach" of the previous section) which leads to a formula for H directly in terms of the Bruhat coordinates. Again we restrict ourselves to the case $M = \mathrm{SU}(N)/T$. We start by stating the main result: Let Z^H, Z^ω be given by

$$Z^H = \sum_{i=1}^{l} \beta_i H'_{\lambda_i}, \quad \beta_i \in \mathbb{R}_+, \tag{4.13}$$

$$Z^\omega = \sum_{j=1}^{l} m_j H_{\lambda_j}, \quad m_j \in \mathbb{Z}_+, \tag{4.14}$$

where H'_{λ_i}, H_{λ_j} correspond to λ_i, λ_j via

$$\lambda_i(Z) = \mathrm{tr}(H'_{\lambda_i} Z), \quad \lambda_j(Z) = \langle H_{\lambda_j}, Z \rangle, \quad \forall Z \in \mathscr{T}. \tag{4.15}$$

[Thus Z^H, Z^ω lie in the positive Weyl chamber and, in addition, Z^ω lies on the weight lattice, i.e., it corresponds to the multiplicative weight $\Pi_{j=1}^{l} (\lambda_j)^{m_j}$. We have chosen to express Z^H in terms of H'_{λ_i} as this gives a nicer normalization for phrasing the result.] Then the projection Hamiltonian given, as in (4.2), by $H = \langle iZ^H, giZ^\omega g^{-1} \rangle$ has the alternative expression

$$H = \sum_{i,j=1}^{l} \beta_i m_j H_{ij}, \tag{4.16}$$

where

$$H_{ij} = -i \frac{d}{dt} \frac{\mathscr{H}_{\lambda_j}(s_0, \exp(iH'_{\lambda_i} t) \cdot s_0)}{\mathscr{H}_{\lambda_j}(s_0, s_0)} \Bigg|_{t=0}. \tag{4.17}$$

Here $\mathscr{H}_{\lambda_j}(\cdot,\cdot)$ is the Hermitian structure on the jth basic line bundle $L_{\lambda_j} = \det_j$ from Sec. III, s_0 is the standard section of $\det_j$ obtained from the exterior product of the first j columns of $g_c(z)$, the coordinatized element of G_c representing $m = [g_c(z)]_B$, and $\cdot$ is the action of the torus on sections by left multiplication (discussed in more detail below).

We start by describing the action of the torus on $[g_c(z)]_B$. Here $g_c(z)$ is the lower triangular matrix with 1's on the diagonal and complex coordinates z^α in the αth position where α is a negative root, i.e., $\alpha = ij$ with $j < i$. When considering the action of the torus on $[g_c(z)]_B$ it is sufficient to look at the basic flows $\exp(itH'_{\lambda_i}) \cdot [g_c(z)]_B$. The action is simply left multiplication, and, as $T \subset B$, one may simultaneously multiply $g_c(z)$ on the right with $\exp(-itH'_{\lambda_i})$, thus preserving the standard form. Recall from (3.35) that

$$H'_{\lambda_i} = (1/N)\mathrm{diag}(N-i, \ldots, N-i, -i, \ldots, -i), \tag{4.18}$$

with i times $N-i$ and $N-i$ times $-i$. Thus the effect is to multiply the coordinates in the set $R_i = \{z^{kl} | k \geqslant i+1, l \leqslant i\}$ by a factor $\exp(-it)$, while leaving the other coordinates unchanged. The situation has been sketched in Fig. 3.

Tangent to the basic flow $\exp(itH'_{\lambda_i}) \cdot [g_c(z)]_B$ there is a vector field ξ_i,

$$\xi_i = \frac{d}{dt} \exp(itH'_{\lambda_i}) \cdot [g_c(z)]_B \Big|_{t=0}. \tag{4.19}$$

Writing $z = re^{i\varphi}$ the rotation flow $z \to e^{-it} z$ corresponds to the vector $-\partial_\varphi = -(i/2)(z\,\partial_z - \bar{z}\,\partial_{\bar{z}})$. Thus ξ_i may be written

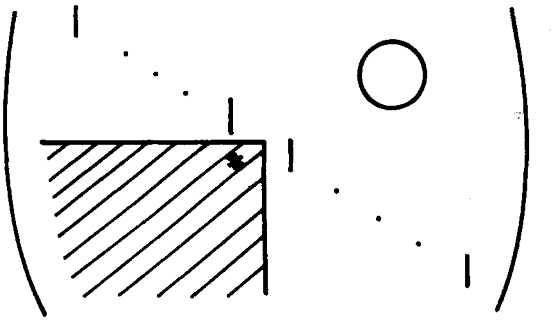

FIG. 3. The action of $\exp(itH'\lambda_i)$ on $[g_c(z)]_B$: the elements in the shaded region R_i are multiplied by $\exp(-it)$. $* = z^{i\,i+1}$.

$$\xi_i = -\frac{i}{2}\sum_{z^{kl}\in R_i}(z^{kl}\partial_{z^{kl}} - \bar z^{kl}\partial_{\bar z^{kl}}). \tag{4.20}$$

Any element of $\mathfrak{t}$ gives rise to a flow that may be decomposed into basic flows, and the corresponding vector field is a linear combination of the basic vector fields ξ_i. In particular, the vector field corresponding to Z^H [(4.13)], and thus, by (a), the rotation vector field of H, is given by

$$\xi = \mathrm{rot}_H = \sum_i \beta_i \xi_i. \tag{4.21}$$

We now turn to the action of the torus on sections s of L_{λ_j}. Regarding L_{λ_j} as the jth determination bundle $\det_j$ (see Sec. III) a section s is an assignment to each flag $\mathscr{F}$ [represented by $g_c(z)$] of the exterior product of a set of j column vectors $\{e_1,...,e_j\}$ in $\mathbb{C}^N$ such that the span of $e_1,...,e_j$ equals the span of the first j columns of $g_c(z)$. For example, we recall that the standard section s_0 corresponds to the choice: $e_i = i$th column of $g_c(z)$. There is a natural action of the torus on sections given by left action on the base point m (as discussed above) combined with left multiplication on the $N\times j$ matrix formed by the juxtaposition of $(e_1,...,e_j)$.

A Hermitian structure $\mathscr{H}_{\lambda_j}(\cdot,\cdot) = \mathscr{H}_{\det_j}$ on the fiber over any flag was defined in Sec. III. However, the definition extends naturally to any pair of j vectors $e_1\wedge\cdots\wedge e_j$, $e_1'\wedge\cdots\wedge e_j'$ (not necessarily spanning the same j plane):

$$\mathscr{H}_{\lambda_j}(e_1\wedge\cdots\wedge e_j, e_1'\wedge\cdots\wedge e_j') = \det[(e_i,e_j')]. \tag{4.22}$$

Using the assertion after 7.5.3 in Ref. 11 one may also write

$$\mathscr{H}_{\lambda_j}(e_1\wedge\cdots\wedge e_j, e_1'\wedge\cdots\wedge e_j') = \sum_{L=1}^{\binom{N}{j}} m_L\,\overline{m_L'}, \tag{4.23}$$

where m_L (resp. m_L') is the L th $j\times j$ minor of the $N\times j$ matrix formed by the juxtaposition of $(e_1\cdots e_j)$ [resp. $(e_1'\cdots e_j')$].

With these definitions the expression (4.17) for H_{ij} makes sense. It is convenient also to define an action of the torus from the right on sections: by right multiplication on $g_c(z)$ (which corresponds to a trivial action as the torus element may be reabsorbed into B; in other words, rotating each basis vector of a flag by a phase does not change the flag) and simultaneously by right multiplication on $(e_1\cdots e_j)$ with the principal $j\times j$ minor of the torus element. Now we claim that the formula (4.17) for H_{ij} may be rewritten

$$H_{ij} = -i\frac{d}{dt}\left.\frac{\mathscr{H}_{\lambda_j}(s_0,\exp(iH'_{\lambda_i}t)\cdot s_0\cdot\exp(-iH'_{\lambda_i}t))}{\mathscr{H}_{\lambda_j}(s_0,s_0)}\right|_{t=0} \tag{4.24}$$

(up to an irrelevant constant). The reason for this is that each minor determination m_L gets multiplied by the same factor $\lambda_j(\exp(-itH'_{\lambda_i}))$ through the right torus action, which thus gives an overall factor. On taking the t derivative at $t=0$ the difference with H_{ij} as in (4.17) is just a constant. [The difference between H in (4.16) and (4.17) and the H using (4.24) for H_{ij} instead is $-\Sigma_{i,j=1}^{l} m_j\beta_i\lambda_j(H'_{\lambda_i})$.] Thus, modulo this constant, one may replace the left action of the torus on s_0 in (4.17) by the adjoint action on $g_c(z)$ (which was described above) but restricted to the first j columns of $g_c(z)$.

It is now important to study the effect of the adjoint action of $\exp(itH'_{\lambda_i})$ on the determinants of the $j\times j$ minors $m_L(z)$, $L=1,...,\binom{N}{j}$. The question is whether each such determinant transforms homogeneously, i.e., with an overall multiple of $\exp(-it)$, despite the fact that $m_L(z)$ is not necessarily a homogeneous polynomial of the z's. For $j\leqslant i$, the answer is clearly affirmative as each row gets multiplied by 1 or $\exp(-it)$ depending on whether the row is in the shaded region R_i or not (Fig. 3). Thus $m_L(z)$ is multiplied by $\exp(-in_L t)$, where n_L is the number or rows of $m_L(z)$ in the shaded region. Now suppose $j>i$. Then a generic minor will have the staircase form sketched in Fig. 4. It is clear that the top right-hand element of the shaded region will always at least touch the staircase as sketched in Fig. 4; in particular the element (marked ● in Fig. 4) directly to the right and above the top right-hand element of the shaded region will always belong to the zero region. Defining the swap of a pair of elements (ij) and (kl) as the pair of elements (il) and (kj), it is clear that, if we swap any pair of elements below the staircase diagonal, either one of the swapped elements is zero, or the total number of elements in the shaded region before and after the swap is unchanged. The determinant m_L is a weighted sum of all products of j elements such that no two lie in the same row or column. Obviously any such product may be obtained from any other by a series of swaps. Thus the total number of elements in the shaded region is constant for all the nonzero products constituting m_L.

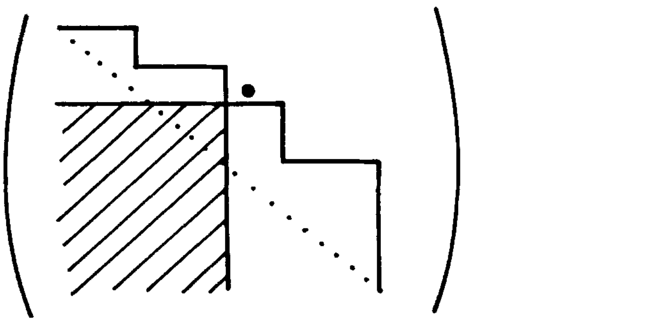

FIG. 4. A $j\times j$ minor of $g_c(z)$, where $j>i$. Above the staircase all elements are zero. The shaded elements lie in the region R_i in Fig. 3.

Thus in all cases m_L transforms homogeneously under the action of $\exp(itH'_{\lambda_i})$. The multiplicity of the factor $\exp(-it)$ in the transformation of m_L will be denoted $\text{mult}_i(L)$, i.e., $m_L \to \exp(-it\,\text{mult}_i(L))\, m_L$ under the action of $\exp(itH'_{\lambda_i})$. Then we may rewrite the expression (4.24) for H_{ij}, using the formula (4.23):

$$H_{ij} = \frac{\Sigma_{L=1}^{(N)} \text{mult}_i(L)|m_L(z)|^2}{\Sigma_{L=1}^{(N)} |m_L(z)|^2}. \tag{4.25}$$

Next consider the symplectic form ω given, according to (4.14), by the multiplicative weight $\Pi_{j=1}^{l}(\lambda_j)^{m_j}$. From the line bundle approach in Sec. III this means that one can write

$$\omega = i\,\partial\bar\partial \ln \prod_j K_j^{m_j}, \tag{4.26}$$

where

$$K_j = \mathscr{H}_{\lambda_j}(s_0,s_0) = \sum_L |m_L(z)|^2. \tag{4.27}$$

Thus $\omega = \Sigma_j m_j \omega_j$, where $\omega_j = i\,\partial\bar\partial \ln K_j$. Then combining with (4.16) and (4.21) it remains to show

$$dH_{ij} = -\iota_{\xi_i}\omega_j, \tag{4.28}$$

where ξ_i was given in (4.20) [$\iota_\xi \omega$ is the one-form whose evaluation on the vector X is given by $\omega(\xi,X)$]. First calculate ω_j:

$$\begin{aligned}\omega_j &= i\partial \frac{1}{K_j}\sum_L m_L(z)\, d\,\overline{m_L(z)} \\ &= i\sum_{M,L}\left(\frac{1}{K_j}\delta_{ML} - \frac{\overline{m_M(z)}\, m_L(z)}{K_j^2}\right) \\ &\quad \times dm_M(z)\wedge d\,\overline{m_L(z)}.\end{aligned} \tag{4.29}$$

Now we assert

$$\begin{aligned}\iota_{\xi_j} dm_M(z)\wedge d\,\overline{m_L(z)} &= -i(\text{mult}_i(M)m_M(z)\,d\,\overline{m_L(z)} \\ &\quad + \text{mult}_i(L)\,\overline{m_L(z)}\,dm_M(z)).\end{aligned} \tag{4.30}$$

Consider a specific coordinate z^α in the region R_i corresponding to ξ_i and the contraction $\iota_{z^\alpha \partial_{z^\alpha}} dm_M(z)$. Then if $m_M(z)$ possesses a term containing z^α the contraction yields that term times a factor 2 [from $\iota\,\partial_z\, dz = \iota_{\partial_x - i\partial_y}(dx + i\,dy) = 2$]. Because each term in $m_M(z)$ has the same number, $\text{mult}_i(M)$, of z's in the region R_i, one has

$$\iota_{\xi_i} dm_M(z) = 2\,\text{mult}_i(M)m_M(z), \tag{4.31}$$

and the assertion follows. Now a short calculation gives

$$\begin{aligned}\iota_{\xi_i}\omega_j &= \frac{1}{K_j}\left(\sum_L \text{mult}_i(L)d\,|m_L(z)|^2\right) - \sum_{M,L}\left\{\frac{1}{K_j^2}\,\overline{m_M(z)}\,m_L(z)(\text{mult}_i(M)m_M(z)\,d\,\overline{m_L(z)} + \text{mult}_i(L)\,\overline{m_L(z)}\,dm_M(z))\right\} \\ &= \frac{1}{K_j}\sum_L \text{mult}_i(L)d\,|m_L(z)|^2 - \frac{dK_j}{K_j^2}\sum_L \text{mult}_i(L)|m_L(z)|^2,\end{aligned} \tag{4.32}$$

where the last step follows after an interchange of the summation variables L and M. Clearly the final expression is the exterior derivative of H_{ij} as given in (4.25).

To illustrate the result (4.16) and (4.17), which has just been proved, we give some examples. For $M = SU(2)/T$, we have $\xi_1 = (-i/2)(z\,\partial_z - \bar z\,\partial_{\bar z})$ and

$$H_{11} = |z|^2/(1+|z|^2). \tag{4.33}$$

For $M = SU(3)/T$ and writing $g_c(z)$ in the usual way as

$$g_c(z) = \begin{pmatrix} 1 & 0 & 0 \\ -z_3 & 1 & 0 \\ -z_2 & z_1 & 1 \end{pmatrix}, \tag{4.34}$$

the basic vector fields are

$$\begin{aligned}\xi_1 &= -(i/2)(z_3\,\partial_{z_3} - \bar z_3\,\partial_{\bar z_3} + z_2\,\partial_{z_2} - \bar z_2\,\partial_{\bar z_2}), \\ \xi_2 &= -(i/2)(z_2\,\partial_{z_2} - \bar z_2\,\partial_{\bar z_2} + z_1\,\partial_{z_1} - \bar z_1\,\partial_{\bar z_1}),\end{aligned} \tag{4.35}$$

and the four basic Hamiltonians H_{ij} are

$$\begin{aligned} H_{11} &= \frac{|z_3|^2 + |z_2|^2}{1 + |z_3|^2 + |z_2|^2}, \\ H_{12} &= \frac{|z_2 - z_1 z_3|^2}{1 + |z_1|^2 + |z_2 - z_1 z_3|^2}, \\ H_{21} &= \frac{|z_2|^2}{1 + |z_3|^2 + |z_2|^2}, \\ H_{22} &= \frac{|z_1|^2 + |z_2 - z_1 z_3|^2}{1 + |z_1|^2 + |z_2 - z_1 z_3|^2}. \end{aligned} \tag{4.36}$$

These results agree with the Hamiltonians found by the more laborious direct route involving the Iwasawa decomposition.

Finally we look at the nongeneric flag manifolds $\mathbb{C}P^n = SU(n+1)/S(U(1)\times U(n))$. We refer to the discussion at the end of Sec. III concerning $C(T_0)$, P, and the Bruhat coordinatization. For simplicity, we choose the character defining ω to be $\chi = \lambda_1$, corresponding to the choice $m = 1$ in (3.53). Then H is given by [cf. (4.16) and (4.25)]

$$H = \sum_{i=1}^{n} \beta_i H_{i1}, \tag{4.37}$$

where

$$H_{i1} = \frac{\Sigma_{j>i}|z_j|^2}{1 + \Sigma_{k=1}^{n}|z_k|^2}. \tag{4.38}$$

This follows from the action of the torus on $[g_c(z)]_P$, analogously to the discussion of the generic case. In practice it is more convenient to change the basis of parameters and write instead

$$H = \frac{\Sigma_{i=1}^{n}\mu_i|z_i|^2}{1 + \Sigma_{i=1}^{n}|z_i|^2}. \tag{4.39}$$

The condition in (4.13), $\beta_i \in \mathbb{R}_+$, corresponds to $0 < \mu_1 < \mu_2 < \cdots < \mu_n$.

V. THE DUISTERMAAT–HECKMAN INTEGRATION FORMULA ON FLAG MANIFOLDS

In this section we will devote ourselves to a discussion of the Duistermaat–Heckman (DH) integration formula, referred to in the Introduction, as applied in the special case of flag manifolds. Let us first state a version of the DH result appropriate to our purposes: let M be a compact, symplectic manifold of dimension $2n$, with symplectic form ω, and H be a Hamiltonian function on M such that the corresponding Hamiltonian vector field $\xi = \mathrm{rot}_H$, given by [cf. (4.1)]

$$dH + \iota_\xi \omega = 0, \tag{5.1}$$

has only isolated zeros (or equivalently H itself has only isolated critical points). Here $\iota_\xi\omega$, the contraction of ξ with ω, is the one-form defined by

$$\iota_\xi \omega(X) = \omega(\xi,X), \quad \forall X\in TM. \tag{5.2}$$

Assume furthermore that the vector field ξ is almost periodic, i.e., that there exists some torus [product of U(1)'s] acting as a group of transformations on M such that ξ is the induced vector field corresponding to some element of the torus Lie algebra. Now we define the mixed form

$$\begin{aligned}\mu &= \exp(H+\omega)\\ &= \exp(H)\left(1+\frac{\omega}{1!}+\frac{\omega^2}{2!}+\cdots+\frac{\omega^n}{n!}\right),\end{aligned} \tag{5.3}$$

i.e., μ is the formal sum of forms of different degrees,

$$\mu = \sum_{p=0}^{n} \mu^{(2p)}, \tag{5.4}$$

where $\mu^{(2p)}$ is the $2p$-form component of μ. From (5.1) and using the fact that ω is closed, we have

$$(d + \iota 2_\xi)(H+\omega) = 0, \tag{5.5}$$

and thus also

$$(d+\iota_\xi)\mu = 0. \tag{5.6}$$

We say that the mixed form μ is equivariantly closed. Equivariant cohomology uses the operator $d+\iota_\xi$ instead of d (for more information on the use of equivariant cohomology in the present context, see Ref. 6). Now the DH formula gives an expression for the integral of the highest degree component of μ over M, namely,

$$\int_M \mu^{(2n)} = (2\pi)^n \sum_{\{m_I\}} \frac{\mu^{(0)}(m_I)}{\mathrm{Pf}(-J_\xi(m_I))} \tag{5.7}$$

or, in terms of H and ω,

$$\int_M \exp(H)\frac{\omega^n}{n!} = (2\pi)^n \sum_{\{m_I\}} \frac{\exp(H(m_I))}{\mathrm{Pf}(-J_\xi(m_I))}. \tag{5.8}$$

Here $\{m_I\}$ are the zeros of ξ and $\mathrm{Pf}(J_\xi(m_I))$ is a kind of winding number of the vector field ξ around its zero at m_I. More specifically, J_ξ is a linear map $T_{m_I}M \to T_{m_I}M$ given by

$$J_\xi(Y) = -\pounds_\xi Y = (Y\cdot\xi^k)\partial_k. \tag{5.9}$$

$\mathrm{Pf}(J_\xi)$, the Pfaffian of J_ξ, is the square root of the determinant of this linear map in the following sense. It is always possible to choose coordinates x_i, y_i, $i=1,\dots,n$, around m_I which vanish at m_I and such that

$$\xi = \beta_i(x_i\,\partial_{y_i} - y_i\,\partial_{x_i}) + \cdots, \tag{5.10}$$

where the β_i are constants and where the dots indicate higher-order terms in the x's and y's. Thus with respect to the (positively oriented) basis $\{\partial_{x_1},\partial_{y_1},\dots,\partial_{x_n},\partial_{y_n}\}$ of $T_{m_I}M$, J_ξ has the block diagonal form

$$J_\xi = \text{block diag}\left(\begin{pmatrix}0 & \beta_1\\ -\beta_1 & 0\end{pmatrix},\dots,\begin{pmatrix}0 & \beta_n\\ -\beta_n & 0\end{pmatrix}\right), \tag{5.11}$$

and we define the Pfaffian of J_ξ in this form to be

$$\mathrm{Pf}(J_\xi) = \prod_{i=1}^{n}\beta_i. \tag{5.12}$$

Of course, with this definition,

$$\mathrm{Pf}(-J_\xi) = \prod_{i=1}^{n}(-\beta_i).$$

That J_ξ is antisymmetric can be seen from the identity

$$\xi\cdot\omega(X,Y) = (\pounds_\xi\omega)(X,Y) + \omega(\pounds_\xi X,Y) + \omega(X,\pounds_\xi Y), \tag{5.13}$$

using the vanishing of $\pounds_\xi\omega = (d+\iota_\xi)^2\omega$, which follows from (5.1) and the vanishing of ξ at m_I. This justifies the antisymmetric form (5.11) of J_ξ.

As an example, take $M=S^2$ with $H=\cos\theta$ and $\omega = \frac{1}{2}\sin\theta\, d\theta\wedge d\varphi$ (the $\frac{1}{2}$ factor is for later convenience). Actually these coordinates are not strictly speaking correct as ω has coordinate singularities at $\theta = 0,\pi$, but we will ignore this for the moment. The left-hand side of formula (5.8) is the integral

$$\int_{S^2}\exp(H)\omega = \int_{\varphi=0}^{2\pi}\int_{\theta=0}^{\pi}\exp(\cos\theta)\,\frac{1}{2}\sin\theta\, d\theta\wedge d\varphi, \tag{5.14}$$

which may be evaluated directly yielding $\pi(e-e^{-1})$. In fact, as we will be concentrating on the south pole $\theta=\pi$ first, we prefer to take $\omega = -\frac{1}{2}\sin\theta\, d\theta\wedge d\varphi$ and write

$$\int_{S^2}\exp(H)\omega = \int_{\varphi=0}^{2\pi}\int_{\theta=\pi}^{0}\exp(\cos\theta)\times\left(-\frac{1}{2}\sin\theta\, d\theta\wedge d\varphi\right). \tag{5.15}$$

Because of the different direction of integrating θ, the answer is still $\pi(e-e^{-1})$. Now we introduce a change of coordinates to a complex coordinate $z = x+iy$, where

$$x = \cot(\theta/2)\cos\varphi, \tag{5.16}$$

$$y = \cot(\theta/2)\sin\varphi. \tag{5.17}$$

The south pole $\theta=\pi$ now corresponds to $z=0$. In these coordinates, H and ω [in (5.15)] take the form

$$\begin{aligned}H &= -1 + 2(x^2+y^2)/(1+x^2+y^2)\\ &= -1 + 2|z|^2/(1+|z|^2),\end{aligned} \tag{5.18}$$

$$\begin{aligned}\omega &= 2\,dx\wedge dy/(1+x^2+y^2)^2\\ &= i\,dz\wedge d\bar z/(1+|z|^2)^2.\end{aligned} \tag{5.19}$$

Apart from an irrelevant constant -1, H corresponds to the expression obtained in Sec. IV [Eq. (4.33) with $\beta_1 = 2$]. Also, one recognizes ω as the symplectic form obtained in Sec. III [Eqs. (3.10) and (3.30) with μ_1 (resp. m_1) equal to

1]. Notice that the change of coordinates has removed the singularity at the south pole. The Hamiltonian vector field ξ, which equals $2\,\partial_\varphi$ in (θ,φ) coordinates, is given by

$$\xi = 2(x\,\partial_y - y\,\partial_x) = i(z\,\partial_z - \bar z\,\partial_{\bar z}) \tag{5.20}$$

in terms of the coordinate z. Using (5.11) and (5.12), we obtain

$$\mathrm{Pf}(-J_\xi(z=0)) = -2. \tag{5.21}$$

Thus the south pole contributes a term

$$2\pi \exp(H(z=0))/\mathrm{Pf}(-J_\xi(z=0)) = -\pi e^{-1} \tag{5.22}$$

to the right-hand side of the DH formula (5.8). To get the contribution from the north pole one changes coordinates from z to $z' = (z)^{-1}$, which vanish at the north pole and in which H, ω take the form

$$H = -1 + 2/(1+|z'|^2), \tag{5.23}$$

$$\omega = i\,dz' \wedge d\bar z'/(1+|z'|^2)^2. \tag{5.24}$$

The Hamiltonian vector field ξ is now

$$\xi = -i(z'\,\partial_{z'} - \bar z'\,\partial_{\bar z'}), \tag{5.25}$$

and thus one has

$$\mathrm{Pf}(-J_\xi(z'=0)) = 2 \tag{5.26}$$

(after introducing a basis of vectors $\{\partial_{x'},\partial_{y'}\}$, where $z' = x' + iy'$). Hence the north pole contributes

$$2\pi \exp(H(z'=0))/\mathrm{Pf}(-J_\xi(z'=0)) = \pi e \tag{5.27}$$

to the right-hand side of the DH formula (5.8). Combining (5.22) and (5.27), one reobtains the correct answer to the integral (5.15). Finally we remark that one may replace the integral (5.15) over S^2 by an integral over a maximal Bruhat cell (e.g., the complex plane corresponding to the coordinate z). With H and ω given by (5.18) and (5.19), one reobtains the same result after changing to polar coordinates (r,θ), the two critical points now being $r=0$ and $r=\infty$.

The essence of the DH formula is "localization": the result of the integral may be expressed entirely in terms of data localized at the critical points of H. The mechanism behind this phenomenon becomes clear from the following consideration: let M_ε denote the manifold M punctured by removing small solid spheres of radius ε centered around the critical points m_I. (For the case $M = S^2$, M_ε is the doubly punctured sphere with θ ranging from ε to $\pi - \varepsilon$.) Then clearly one has

$$\int_M \exp(H)\,\frac{\omega^n}{n!} = \lim_{\varepsilon\to 0}\int_{M_\varepsilon} \exp(H)\,\frac{\omega^n}{n!}. \tag{5.28}$$

On M_ε we may perform the following manipulation: let $d\varphi$ be any one-form dual to the Hamiltonian vector field ξ, i.e.,

$$\iota_\xi\, d\varphi = 1. \tag{5.29}$$

Such a $d\varphi$ exists as we have excluded the zeros of ξ. Then, as the symplectic volume is nondegenerate, we have

$$\frac{\omega^n}{n!} = d\varphi \wedge \iota_\xi\left(\frac{\omega^n}{n!}\right) = d\varphi \wedge (\iota_\xi \omega) \wedge \frac{\omega^{n-1}}{(n-1)!}$$
$$= dH \wedge d\varphi \wedge \frac{\omega^{n-1}}{(n-1)!}. \tag{5.30}$$

This allows us to write the integrand as an exact form on M_ε:

$$\exp(H)\omega^n/n! = d(\exp(H)\,d\varphi \wedge \omega^{n-1}/(n-1)!). \tag{5.31}$$

Then, using Stokes' theorem we have, from (5.28) and (5.31),

$$\int_M \exp(H)\,\frac{\omega^n}{n!} = \lim_{\varepsilon\to 0}\int_{\partial M_\varepsilon} \exp(H)\,d\varphi\,\frac{\omega^{n-1}}{(n-1)!}. \tag{5.32}$$

The precise expression for the right-hand side of (5.8) follows after evaluating the integral and taking the limit in (5.32). It is, however, clear that the answer can only depend on data localized at the critical points m_I, as data from any other points can always be excluded by taking ε sufficiently small. The argument we have just given is very similar to Berline–Vergne Théorème 1.6.[1]

Before discussing the generalization to $\mathbb{C}P^n$ we would like to derive an alternative procedure for obtaining $\mathrm{Pf}(J_\xi)$ in terms of data derived from H and ω. Suppose we choose coordinates (x_i,y_i), $i=1,\dots,n$, around the critical point m_I, which vanish at m_I, as in (5.10). Then the three key objects H, ω, and ξ are expressed in these coordinates as

$$H = H(m_I) + \sum_i \beta_i m_i \frac{(x_i^2+y_i^2)}{2} + \cdots, \tag{5.33}$$

$$\omega = \sum_i m_i\, dx_i \wedge dy_i + \cdots, \tag{5.34}$$

$$\xi = \sum_i \beta_i (x_i\,\partial_{y_i} - y_i\,\partial_{x_i}) + \cdots, \tag{5.35}$$

where $\beta_i \in \mathbb{R}_+$, $m_i \in \mathbb{Z}_+$ are constants. Then we have the alternative expression for $\mathrm{Pf}(J_\xi(m_I))$,

$$\mathrm{Pf}(J_\xi(m_I)) = (\det(\mathrm{Hess}(H(m_I))))^{1/2}/|\omega|(m_I), \tag{5.36}$$

where $|\omega|(m_I)$, the determinant of ω at m_I, is defined as

$$|\omega|(m_I) = \prod_{i=1}^n m_i. \tag{5.37}$$

In the example of $\mathbb{C}P^1$ we can use (5.36) to read off the value of $\mathrm{Pf}(J_\xi)$ directly from (5.18) and (5.19) [resp. (5.23) and (5.24)].

We now turn to a discussion of $\mathbf{M} = \mathbb{C}P^n$, n arbitrary. As for $\mathbb{C}P^1$ we may replace the integration over $\mathbb{C}P^n$ by an integration over the largest Bruhat cell, in this case $\mathbb{C}^n$. Using the expressions previously found for H (4.39) and the symplectic volume (3.56) [for simplicity we choose $m=1$ in (3.56)] the integral to be evaluated is

$$\int_{\mathbb{C}^n} \exp\left(\frac{\Sigma_{i=1}^n \mu_i |z_i|^2}{1+\Sigma_{i=1}^n |z_i|^2}\right)\frac{\Pi_{i=1}^n\, i\,dz_i \wedge d\bar z_i}{(1+\Sigma_{i=1}^n |z_i|^2)^{n+1}}. \tag{5.38}$$

We use the method described in the previous paragraph to determine the right-hand side of the integration formula. For the critical point at the origin of $\mathbb{C}^n$, we have

$$(\det(\mathrm{Hess}(H)))^{1/2} = \prod_{i=1}^n (2\mu_i) \tag{5.39}$$

and

$$|\omega| = 2^n, \tag{5.40}$$

as $i\,dz \wedge d\bar z = 2\,dx \wedge dy$. Hence

$$\mathrm{Pf}(J_\xi) = \prod_{i=1}^n (\mu_i) \tag{5.41}$$

for this critical point, which we call m_0. There are a further n critical points m_i, $i=1,\ldots,n$ corresponding to the limits $|z_i|\to\infty$ with $|z_j|<\infty, j\neq i$. Thus there are $(n+1)$ critical points labeled by $I=0,i$, with $i=1,\ldots,n$. We obtain the factor $\mathrm{Pf}(J_\xi(m_i))$ by changing to Bruhat coordinates centered around the corresponding critical point. Explicitly we use the freedom to multiply on the right with elements of P to convert

$$[g_c(z)]_P=\left[\begin{pmatrix} 1 & * & \cdots & * \\ z_1 & & & \\ z_2 & & & \\ \vdots & \vdots & & \vdots \\ z_n & * & & * \end{pmatrix}\right]_P$$

to

$$[g_c(z')]_P=\left[\begin{pmatrix} z_i' & * & \cdots & * \\ z_1' & & & \\ z_2' & & & \\ \vdots & & & \\ 1 & & & \\ \vdots & \vdots & & \vdots \\ z_n' & * & \cdots & * \end{pmatrix}\right]_P=[g_c(z)p(z)]_P,$$

where the 1 is in the $(i+1)$th row. We do not display the remaining $n-1$ columns of the g_c's as they contain no information about the flag. Thus $p_{11}=(z_i)^{-1}$ and the coordinate transformation is

$$z_i'=1/z_i,\quad z_j'=z_j/z_i,\quad j\neq i. \tag{5.42}$$

The same result is, of course, obtained using the more conventional homogeneous coordinates for $\mathbb{C}P^n$.

Now, in the primed coordinates,

$$|\omega|(m_i)=2^n \tag{5.43}$$

still holds, as the Kähler potential F is $F=e_1^T\bar{e}_1$ [where e_1 is the first column of $g_c(z')$], which has the same functional form in both sets of coordinates. Also, H is given in the primed coordinates by

$$H=\frac{\mu_i+\Sigma_{j\neq i}\mu_j|z_j'|^2}{1+\Sigma_{j=1}^n|z_j'|^2}, \tag{5.44}$$

and thus

$$(\det(\mathrm{Hess}(H(m_i))))^{1/2}=2^n(-\mu_i)\prod_{j\neq i}(\mu_j-\mu_i), \tag{5.45}$$

from expanding H to second order. Thus, for the case $\mathbb{C}P^n$, the DH formula (5.8) for the integral (5.38) gives the result

$$(2\pi)^n\left(\frac{1}{\Pi_{j=1}^n(-\mu_j)}+\sum_{i=1}^n\left(\frac{\exp(\mu_i)}{\mu_i\Pi_{j\neq i}(\mu_i-\mu_j)}\right)\right). \tag{5.46}$$

We will return to an alternative derivation of this expression at the end of this section.

Now, however, we turn to the DH formula for a more complicated flag manifold of generic type, namely, $M=\mathrm{SU}(3)/T$. Recall from (4.20) that H is

$$\begin{aligned}H=&\beta_1m_1\frac{(|z_3|^2+|z_2|^2)}{(1+|z_3|^2+|z_2|^2)}\\&+\beta_1m_2\frac{|z_2-z_1z_3|^2}{(1+|z_1|^2+|z_2-z_1z_3|^2)}\\&+\beta_2m_1\frac{|z_2|^2}{(1+|z_3|^2+|z_2|^2)}\\&+\beta_2m_2\frac{(|z_1|^2+|z_2-z_1z_3|^2)}{(1+|z_1|^2+|z_2-z_1z_3|^2)}\end{aligned} \tag{5.47}$$

and from (3.17) [resp. 3.51)] that the symplectic volume is

$$\frac{\omega^3}{3!}=m_1\cdot m_2\cdot(m_1+m_2)\frac{1}{(K_1K_2)^2}\prod_{j=1}^3 i\,dz_j\wedge d\bar{z}_j, \tag{5.48}$$

where

$$K_1=1+|z_2|^2+|z_3|^2, \tag{5.49}$$

$$K_2=1+|z_1|^2+|z_2-z_1z_3|^2. \tag{5.50}$$

For this flag manifold there are six critical points [one in each Bruhat cell—cf. (c) at the beginning of Sec. IV], namely, $m_w=[\omega_w]_B$, $w\in W$. The index set I is thus equal to the Weyl group W. The critical point m_1 lies at the origin of the Bruhat coordinate patch $M_{(13)}$, which has coordinates (z_1,z_2,z_3) and we evaluate its contribution to the right-hand side of the DH formula using (5.36). From

$$\begin{aligned}&(\det(\mathrm{Hess}(H(m_1))))^{1/2}\\&\quad=2^3\beta_2m_2\cdot(\beta_1+\beta_2)(m_1+m_2)\cdot\beta_1m_1\end{aligned} \tag{5.51}$$

and

$$|\omega|(m_1)=2^3m_1\cdot m_2\cdot(m_1+m_2), \tag{5.52}$$

the contribution of this critical point is found to be

$$\exp(H(m_1))/\mathrm{Pf}(-J_\xi(m_1))=-1/\beta_2\cdot(\beta_1+\beta_2)\cdot\beta_1. \tag{5.53}$$

To get the contributions from the other critical points we adopt the following procedure: (a) find the change of coordinates relating the z coordinates in $M_{(13)}$ to z' coordinates in $\omega_wM_{(13)}$; (b) express H in the primed coordinates; and (c) obtain $H(m_w)$ and $\mathrm{Hess}(H(m_w))^{1/2}$. This is all that is required, because it turns out that

$$|\omega|(m_w)=|\omega|(m_1),\quad\forall w\in W. \tag{5.54}$$

The reason is that the Kähler potential in the primed coordinates has exactly the same functional form as in the unprimed coordinates. We will return to this point later.

To demonstrate this procedure, consider the coordinate system in $\omega_{(23)}M_{(13)}$, which is described by

$$(z_1',z_2',z_3')\mapsto\left[\begin{pmatrix}1&0&0\\-z_2'&z_1'&1\\z_3'&-1&0\end{pmatrix}\right]_B. \tag{5.55}$$

The change of coordinates may be found by converting $g_c(z)$ into the same form as (5.55) using the freedom to multiply by elements of B on the right:

$$\left[\begin{pmatrix}1&0&0\\-z_3&1&0\\-z_2&z_1&1\end{pmatrix}\right]_B$$

$$= \left[\begin{pmatrix} 1 & 0 & 0 \\ -z_3 & 1 & 0 \\ -z_2 & z_1 & 1 \end{pmatrix} \begin{pmatrix} 1 & 0 & 0 \\ 0 & -(z_1)^{-1} & 1 \\ 0 & 0 & z_1 \end{pmatrix} \right]_B$$

$$= \left[\begin{pmatrix} 1 & 0 & 0 \\ -z_3 & -(z_1)^{-1} & 1 \\ -z_2 & -1 & 0 \end{pmatrix} \right]_B . \qquad (5.56)$$

Thus we read off the coordinate transformation by comparing (5.55) and (5.56). The transformed Hamiltonian reads

$$H = \beta_1 m_1 \frac{|z_2'|^2 + |z_3'|^2}{1 + |z_3'|^2 + |z_2'|^2} + \beta_1 m_2 \frac{|z_2' - z_1' z_3'|^2}{1 + |z_1'|^2 + |z_2' - z_1' z_3'|^2} + \beta_2 m_1 \frac{|z_3'|^2}{1 + |z_3'|^2 + |z_2'|^2} + \beta_2 m_2 \frac{1 + |z_2' - z_1' z_3'|^2}{1 + |z_1'|^2 + |z_2' - z_1' z_3'|^2}, \qquad (5.57)$$

which we expand around the origin of the primed coordinate patch:

$$H = \beta_2 m_2 + |z_1'|^2 \cdot -\beta_2 m_2 + |z_2'|^2 \cdot \beta_1 (m_1 + m_2) + |z_3'|^2 \cdot m_1 (\beta_1 + \beta_2) + O(z^3). \qquad (5.58)$$

From this we read off

$$(\mathrm{Hess}(H(m_{(23)})))^{1/2} = 2^3 \cdot -\beta_2 m_2 \cdot (m_1 + m_2) \beta_1 \cdot m_1 (\beta_1 + \beta_2). \qquad (5.59)$$

There is, however, an alternative procedure that enables one to express H in the primed coordinates without finding the explicit coordinate transformation. This method is simply to replace each minor m_L in the expression for H (5.47) by the corresponding minor in the matrix (5.55). Thus $\mathbf{1} \to \mathbf{1}$; $-z_3 \to -z_2'$, $-z_2 \to z_3'$, $1 \cdot 1 - (-z_3) \cdot 0 \to 1 \cdot z_1' - (-z_2') \cdot 0$, etc. Under these transformations H [(5.57)] is obtained. The reason for this phenomenon is that the torus acts in a universal way on the minors which can be constructed from the first two columns of a 3×3 matrix, as is evident from

$$\exp(iZt) \begin{pmatrix} a_{11} & a_{12} & * \\ a_{21} & a_{22} & * \\ a_{31} & a_{32} & * \end{pmatrix} \exp(-iZt) = \begin{pmatrix} a_{11} & e^{i12(Z)t} a_{12} & * \\ e^{i21(Z)t} a_{21} & a_{22} & * \\ e^{i31(Z)t} a_{31} & e^{i32(Z)t} a_{32} & * \end{pmatrix} \qquad (5.60)$$

(cf. the general analysis of the torus action on m_L in Sec. IV). Thus we give the remaining coordinatizations and corresponding formulas for the Hamiltonians in Table II. Now it is also obvious why $|\omega|$ is the same for all the critical points: the basic Kähler potentials are $F_i = \ln(K_i)$, where K_i is the sum of the norms squared of the determinants of the $i \times i$ minors m_L, and by inspection K_1', K_2' have the same functional form as K_1, K_2. Putting everything together, the final result for the right-hand side of the DH formula for $M = \mathrm{SU}(3)/T$ is

$$\frac{(2\pi)^3}{\beta_1 \cdot (\beta_1 + \beta_2) \cdot \beta_2} \Bigg[\underset{(m_1)}{-1} + \underset{(m_{(23)})}{\exp(\beta_2 m_2)} + \underset{(m_{(12)})}{\exp(\beta_1 m_1)} - \underset{(m_{(123)})}{\exp(\beta_1 m_1 + \beta_1 m_2 + \beta_2 m_2)} - \underset{(m_{(132)})}{\exp(\beta_1 m_1 + \beta_2 m_1 + \beta_2 m_2)} + \underset{(m_{(13)})}{\exp((\beta_1 + \beta_2)(m_1 + m_2))} \Bigg]. \qquad (5.61)$$

We have indicated the origin of each contribution in parentheses beneath it. Given the complicated nature of H and $\omega^3/3!$ [(5.47) and (5.48)] it is unlikely that this integral could have been evaluated by any direct method without a great deal of effort.

We postpone any analysis of the general $\mathrm{SU}(N)/T$ case to a future publication. However, before leaving the reader, we would like to point out an amusing recursive property of the $\mathbb{C}P^n$ integrals, closely related to the decomposition of

TABLE II. The remaining coordinatizations and the corresponding formulas for H.

		$H = \beta_1 m_1(\cdots) + \beta_1 m_2(\cdots) + \beta_2 m_1(\cdots) + \beta_2 m_2(\cdots)$			
w	$\omega_w g_c(z')$	$\beta_1 m_1$	$\beta_1 m_2$	$\beta_2 m_1$	$\beta_2 m_2$
(12)	$\begin{pmatrix} -z_3' & 1 & 0 \\ -1 & 0 & 0 \\ -z_2' & z_1' & 1 \end{pmatrix}$	$\frac{1+\lvert z_2'\rvert^2}{K_1'}$	$\frac{\lvert z_1'\rvert^2}{K_2'}$	$\frac{\lvert z_2'\rvert^2}{K_1'}$	$\frac{(\lvert z_1'\rvert^2 + \lvert z_2' - z_1' z_3'\rvert^2)}{K_2'}$
(123)	$\begin{pmatrix} -z_2' & z_1' & 1 \\ -1 & 0 & 0 \\ z_3' & -1 & 0 \end{pmatrix}$	$\frac{1+\lvert z_3'\rvert^2}{K_1'}$	$\frac{1}{K_2'}$	$\frac{\lvert z_3'\rvert^2}{K_1'}$	$\frac{1+\lvert z_2' - z_1' z_3'\rvert^2}{K_2'}$
(132)	$\begin{pmatrix} z_3' & -1 & 0 \\ -z_2' & z_1' & 1 \\ -1 & 0 & 0 \end{pmatrix}$	$\frac{1+\lvert z_2'\rvert^2}{K_1'}$	$\frac{\lvert z_1'\rvert^2}{K_2'}$	$\frac{1}{K_1'}$	$\frac{1+\lvert z_1'\rvert^2}{K_2'}$
(13)	$\begin{pmatrix} -z_2' & z_1' & 1 \\ -z_3' & 1 & 0 \\ -1 & 0 & 0 \end{pmatrix}$	$\frac{1+\lvert z_3'\rvert^2}{K_1'}$	$\frac{1}{K_2'}$	$\frac{1}{K_1'}$	$\frac{1+\lvert z_1'\rvert^2}{K_2'}$

$\mathbb{C}P^n$ into cells. This also has the merit of providing a check on the result (5.46).

Our starting point in the $\mathbb{C}P^n$ case is the integral (5.38), which we transform by changing variables

$$i\, dz_i \wedge d\bar{z}_i \to 2\, dx_i \wedge dy_i \to 2r_i\, dr_i \wedge d\theta_i \to da_i \wedge d\theta_i,$$

where $a_i = r_i^2$, and by integrating out the n trivial θ_i integrals, giving

$$(2\pi)^n \int_{a_i=0}^{\infty} \cdots \int \exp(H(n)) \frac{(\Pi_{k=1}^n da_k)}{(K(n))^{n+1}}, \qquad (5.62)$$

where

$$H(i) = \frac{\Sigma_{j=1}^i \mu_j a_j}{K(i)}, \qquad (5.63)$$

$$K(i) = 1 + \sum_{j=1}^{i} a_j. \qquad (5.64)$$

Also define

$$c(i) = \mu_i - \sum_{j=1}^{i-1} (\mu_j - \mu_i) a_j. \qquad (5.65)$$

It is easily checked that

$$c(i)/K(i-1) = \mu_i - H(i-1), \qquad (5.66)$$

$$c(i)/K(i) = \mu_i - H(i). \qquad (5.67)$$

We carry out the integral (5.62) in steps: first integrating the variable a_n, taking the upper and lower limits $a_n = \infty$, $a_n = 0$, then integrating the variable a_{n-1}, etc. After taking j lower limits we denote the indefinite integral in the next variable a_{n-j} by $I_{j+1}(a_{n-j})$ (with the remaining variables a_i, $i < n-j$ regarded as constants). Thus

$$I_{j+1}(a_{n-j}) = -\int I_j(a_{n-j+1} = 0)\, da_{n-j}, \quad j = 1,\ldots,n, \qquad (5.68)$$

and $I_1(a_n)$ is the indefinite integral obtained from carrying out the a_n integration in (5.62). The jth upper limit contribution

$$\int_0^\infty \cdots \int_0^\infty I_j(\infty)\, da_{n-j}\, da_{n-(j+1)} \cdots da_1 \qquad (5.69)$$

will be denoted P_j.

To give a flavor of the nature of this integration problem we obtain $I_1[a_n]$ and P_1. Define

$$I(p) = \int \exp(H(n)) \frac{da_n}{(K(n))^p}. \qquad (5.70)$$

Then using

$$\frac{\partial H(n)}{\partial a_n} = \frac{c(n)}{K^2(n)}, \qquad (5.71)$$

which follows from (5.63)–(5.65) and partial integration, we obtain the recursion

$$I(p+1) = \frac{(p-1)}{c(n)} I(p) + \frac{1}{c(n)} \frac{\exp(H(n))}{(K(n))^{p-1}}, \quad p \geqslant 1, \qquad (5.72)$$

which is solved by

$$I(p+1) = (p-1)! \left(\sum_{j=0}^{p-1} \left(\frac{K(n)}{c(n)} \right)^{p-j} \frac{1}{j!} \right) \frac{\exp(H(n))}{(K(n))^p}. \qquad (5.73)$$

Setting $p = n$ in (5.67) and taking the upper limit $a_n = \infty$ yield

$$P_1 = (n-1)! \exp(\lambda_n) \int_0^\infty \cdots \int_0^\infty \frac{1}{(c(n))^n} da_{n-1} \cdots da_1, \qquad (5.74)$$

which is a straightforward integral in the remaining variables as the exponential in the integrand has been removed. Notice that $K(n) \to \infty$ as $a_n \to \infty$ so that only the $j = 0$ summand in (5.73) contributes. At the lower limit $a_n = 0$, we have a new integration problem

$$I_2(a_{n-1}) = -(2\pi)^n (n-1)! \times \int \left(\sum_{j=0}^{n-1} \frac{1}{(\mu_n - H(n-1))^{n-j}} \frac{1}{j!} \right) \times \frac{\exp(H(n-1))}{(K(n-1))^n} da_{n-1}. \qquad (5.75)$$

Rather than attempting to solve this integral we now introduce a unified approach to solving all the integrals at all levels. Define the class of (indefinite) integrals

$$I[f;p](H) = \frac{1}{c} \int f^{(p)}(H) \frac{dH}{K^{p-1}}, \quad 1 \leqslant p \leqslant n, \qquad (5.76)$$

where f is a function of H and $f^{(p)}$ denotes the pth derivative of f. Equation (5.76) is to be understood in the following sense: we assume that we have taken, say, $n - p$ lower limits and thus we are integrating the variable a_p, but then we change the integration variable to $H(p)$ regarding the other variables on which $H(p)$ depends as constants. Now, using

$$\frac{dK(p)}{dH(p)} = \frac{dK(p)}{da_p} \frac{da_p}{dH(p)} = \frac{(K(p))^2}{c(p)} \qquad (5.77)$$

and partial integration, a recursion may be set up, similar to the previous one in (5.72), leading to

$$I[f;p](H) = \sum_{j=0}^{p-1} f^{(j)} \left(\frac{1}{\mu_p - H} \right)^{p-j} \frac{1}{j!} \quad \frac{(p-1)!}{K^p}. \qquad (5.78)$$

If we now take the lower limit $a_p = 0$ and perform the change of variables in the next (a_{p-1}) integration,

$$\int da_{p-1} = \frac{1}{c(p-1)} \int (K(p-1))^2\, dH(p-1), \qquad (5.79)$$

we are led to the new problem

$$-\frac{1}{c} \int \left(\sum_{j=0}^{p-1} f^{(j)} \left(\frac{1}{\mu_p - H} \right)^{p-j} \frac{1}{j!} \right) (p-1)! \frac{dH}{K^{p-2}}, \qquad (5.80)$$

where the argument $(p-1)$ in H, K, and c is understood. As is easily checked, this is the integration problem

$$I\,[f(H)/(H - \mu_p); p-1]. \qquad (5.81)$$

This may in turn be inserted into (5.78) and the whole process repeats itself. Now the initial integral

$$I_1(a_n) = (2\pi)^n \int \exp(H) \frac{da_n}{K^{n+1}} \tag{5.82}$$

is, on changing variable to H, precisely $I[(2\pi)^n \exp(H);n]$. Thus we arrive at the solution for the integrals $I_j(a_{n-j+1})$:

$$I_j(a_{n-j+1}) = I[f_j;n-j+1](H), \tag{5.83}$$

where

$$f_j(H) = \frac{(2\pi)^n \exp(H)}{\Pi_{n-j+1<k\leqslant n}(H-\mu_k)}, \quad 1\leqslant j\leqslant n. \tag{5.84}$$

The right-hand side of (5.83) is a function of a_{n-j+1} through the a_{n-j+1} dependence of H,K and is given by the formula (5.78). At the final step we have

$$I_n(a_1) = I[f_n;1](H) = [f_n^{(0)}/(\mu_1 - H)]\cdot(1/K), \tag{5.85}$$

yielding at its lower limit $a_1 = 0$ the contribution from the critical point m_0 [in the terminology introduced after (5.41)], namely,

$$-I_n(0) = (2\pi)^n \left(\frac{1}{\Pi_{k=1}^n(-\mu_k)}\right). \tag{5.86}$$

Now we consider the upper limit contributions $P_j, j = 1,...,n$. From (5.69), (5.83), (5.78), and (5.67) these are given by

$$P_j = f_j(\mu_{n-j+1})(n-j)! \times\left[\int_0^\infty \cdots \int_0^\infty \left(\frac{1}{c(n-j+1)}\right)^{n-j+1} da_{n-j} \cdots da_1\right]. \tag{5.87}$$

The integral in square brackets is easily found to be

$$\frac{1/(n-j)!}{\mu_{n-j+1}\Pi_{k=1}^{n-j}(\mu_{n-j+1}-\mu_k)}. \tag{5.88}$$

Together with (5.84) we get the n contributions

$$P_j = (2\pi)^n \frac{\exp(\mu_{n-j+1})}{\mu_{n-j+1}\Pi_{k\neq n-j+1}(\mu_{n-j+1}-\mu_k)}, \quad j = 1,...,n. \tag{5.89}$$

Thus (5.86) and (5.89) combine to reproduce the direct DH result (5.46).

We may rephrase this same result in a more appealing way by defining a class of DH problems $I[f;M]$, where f is a function of one variable and M is a flag manifold, by

$$I[f;M] = \int_M f(H+\omega). \tag{5.90}$$

As in (5.3), the mixed form $f(H+\omega)$ is expanded in a Taylor series in powers of ω. If $\dim_c M = n$, then $\int_M \cdot$ picks out the nth power of ω in the expansion, i.e.,

$$\int_M f(H+\omega) = \int_M f^{(n)}(H)\frac{\omega^n}{n!}. \tag{5.91}$$

After changing to the maximal Bruhat cell $\mathbb{C}^n$ and taking the lower limit of the integration in the first variable a_n, the following lower-dimensional integral remains:

$$-(2\pi)^n \int_0^\infty \cdots \int_0^\infty \left(\sum_{k=0}^{n-1} f^{(k)} \frac{1}{(\mu_n - H)^{n-k}}\frac{1}{k!}\right) \times\frac{(n-1)!}{K^n} da_{n-1} \cdots da_1, \tag{5.92}$$

which may be rewritten as

$$(2\pi)(2\pi)^{n-1}\int_0^\infty \cdots \int_0^\infty \left(\frac{f(H)}{H-\mu_n}\right)^{(n-1)} \times\frac{1}{K^n} da_{n-1} \cdots da_1. \tag{5.93}$$

This is, however, precisely the form that the first integration would have taken had we started with the DH problem instead:

$$I[2\pi f(H)/(H-\mu_n); \mathbb{C}P^{n-1}]. \tag{5.94}$$

Thus we may phrase the previous recursion for $\mathbb{C}P^n$ in terms of DH problems as follows:

$$I[f_j;\mathbb{C}P^{n-j+1}] = P_j + I[2\pi f_{j+1};\mathbb{C}P^{n-j}], \tag{5.95}$$

with P_j as calculated previously in (5.89) and f_j as given in (5.84) but without the factor $(2\pi)^n$.

Note added in proof: There are some inexact statements in Sec. III. The vector fields $\widetilde{E}_\alpha$ and one-forms e^α are neither G-invariant nor nonvanishing on M. However the one-forms e^α [defined in (3.4)] are nonvanishing and linearly independent at each point of the Bruhat coordinate patch. Furthermore, under G-transformations $\Sigma e^\alpha E_\alpha$ transforms by an Ad(T) transformation at each point and, seeing as the expression (3.1) for the components of ω_e is Ad(T)-invariant, the form ω in (3.5) is G-invariant as required. I am grateful to J. M. Mourão for explaining these points to me.

ACKNOWLEDGMENTS

I am grateful to B. Schroer for suggesting this topic and for many stimulating conversations. I benefited from some excellent lectures by R. Schrader and A. Knauf on the DH formula. It is a pleasure to record my gratitude to S. Kobayashi for helpful comments at the start of this work. At various stages of this project I enjoyed useful conversations with E. P. van den Ban, J. J. Duistermaat, and W.-Y. Hsiang. I am grateful to the referee for many constructive comments and for drawing my attention to several references.

This work was supported in part by Deutsche Forschungsgemeinschaft Sachbeihilfe Schr 160/8-1 and in part by Junta Nacional de Investigação Científica e Technológica/Grupo Teórico de Altas Energias.

APPENDIX: LIE ALGEBRA AND STRUCTURE AND REPRESENTATION THEORY

In this appendix we give a brief survey of relevant parts of Lie algebra structure theory and representation theory. For further details the reader is referred to Belinfante,[21] Macdonald,[22] BFR,[8] Pressley and Segal (Ref. 11, Chap. 1), and Bott (Ref. 17, Sec. 4).

1. Representations and weights

Let G be a compact Lie group and $\mathfrak{g}$ its Lie algebra. A *representation* φ of G is a group homomorphism φ: $G\to\mathrm{GL}(V)$, where V is a complex vector space and GL(V) is the group of nonsingular linear transformations of V. The group homomorphism property means that φ satisfies $\varphi(g_1g_2) = \varphi(g_1)\varphi(g_2)$. An equivalent way of viewing representations is to regard the group G as acting directly on V via the action $\cdot: G \times V \to V$, where $g \cdot v = \varphi(g)v$, $\forall g \in G$, $v \in V$. The group homomorphism property of φ translates into the property

$$g_1 \cdot (g_2 \cdot v) = (g_1 g_2) \cdot v, \quad \forall g_1, g_2 \in G, \quad v \in V.$$

Then V is said to be a *G module* under the G action. A representation φ of G induces a representation of $\mathfrak{g}$, i.e., a map $\dot{\varphi}$ from $\mathfrak{g}$ to $\mathfrak{gl}(V)$ [the Lie algebra of $\mathrm{GL}(V)$] via the definition

$$\dot{\varphi}(X) = \frac{d}{dt} \varphi(\exp tX) \bigg|_{t=0}.$$

Again this may be translated into a module formulation. A representation φ of G is said to be *irreducible* if there does not exist a proper submodule V' of V (i.e., $V' \neq 0$, $V' \neq V$) such that $\varphi(G) V' \subset V'$. The dimension of the representation φ, denoted $\dim \varphi$, is defined by $\dim \varphi = \dim_c V$.

Given G, we fix once and for all a maximal torus T of G, i.e., a maximal Abelian subgroup. As a group, T is isomorphic to a product of $\mathrm{U}(1)$'s. Its Lie algebra $\mathfrak{t}$ is a maximal commuting subalgebra of $\mathfrak{g}$. The dimension of T or $\mathfrak{t}$ is called the *rank* of G, and is denoted by l. A representation φ of G is by restriction also a representation of T on V. This representation splits into $\dim \varphi$ one-dimensional representations χ_i: $T \to \mathrm{U}(1)$, i.e., it is possible to find a basis of V such that in this basis $\varphi(t) = \mathrm{diag}(\chi_1(t), \ldots, \chi_{\dim \varphi}(t))$, $\forall t \in T$. The homomorphisms χ_i are known as the *weights* of the representation φ. The set of all possible weights forms a multiplicative group, also known as the group of *characters* of T. [A character of a group is a homomorphism from the group to $\mathrm{U}(1)$.] Instead of these multiplicative weights, many authors use an equivalent formulation with the additive weights $\dot{\chi}_i$: $\mathfrak{t} \to i\mathbb{R}$. The two approaches are related by $\chi_i(\exp X) = \exp \dot{\chi}_i(X)$. Henceforth we will omit the dot as it will be clear from the context which formulation is being used.

2. The adjoint representation and roots

For any G there is one particular representation known as the *adjoint representation* which is special because it is constructed directly from the Lie algebra of G itself. Let $\mathfrak{g}_c$ be the complexification of $\mathfrak{g}$, i.e., $\mathfrak{g}_c = \mathfrak{g} + i\mathfrak{g}$. Thus $\mathfrak{g}_c$ is a complex vector space. The adjoint representation Ad: $G \to \mathrm{GL}(\mathfrak{g}_c)$ is defined by

$$\mathrm{Ad}(g)(Z) = gZg^{-1}, \quad \forall Z \in \mathfrak{g}_c \tag{A1}$$

The corresponding representation of $\mathfrak{g}$ is denoted ad and is given by

$$\mathrm{ad}(X)(Z) = [X, Z], \quad \forall Z \in \mathfrak{g}_c. \tag{A2}$$

Consider the weights of Ad. Clearly, T acts trivially on the complexification of $\mathfrak{t}$, denoted $\mathfrak{h}$; thus the trivial weight, given by $\chi(t) = \mathbf{1}$, occurs with multiplicity precisely l (as T is maximal). The remaining nontrivial weights of Ad form the set R of *roots* of G. The one-dimensional subspace of $\mathfrak{g}_c$ corresponding to the root $\alpha \in R$ is denoted $\mathfrak{g}^\alpha$. The decomposition of $\mathfrak{g}_c$ into irreducible submodules under T,

$$\mathfrak{g}_c = \mathfrak{h} + \sum_{\alpha \in R} \mathfrak{g}^\alpha, \tag{A3}$$

is known as the *root space decomposition*.

3. The Cartan–Killing form

The adjoint representation may also be viewed as a real representation acting on the real vector space $\mathfrak{g}$, instead of $\mathfrak{g}_c$. This gives rise to a natural symmetric bilinear form $\langle \cdot, \cdot \rangle$ on $\mathfrak{g}$, the *Cartan–Killing form*, defined by

$$\langle X, Y \rangle = \mathrm{tr}(\mathrm{ad}(X), \mathrm{ad}(Y)), \quad \forall X, Y \in \mathfrak{g} \tag{A4}$$

[here $\mathrm{ad}(X)$ and $\mathrm{ad}(Y)$ act on $\mathfrak{g}$ only]. Having defined $\langle \cdot, \cdot \rangle$ on $\mathfrak{g}$, it extends by complex bilinearity to a form on $\mathfrak{g}_c$, also. If the Cartan–Killing form is nondegenerate, the Lie algebra $\mathfrak{g}$ is said to be *semisimple*. An equivalent criterion for semisimplicity is for $\mathfrak{g}$ to have no nonzero Abelian ideals (if a subspace $\mathfrak{i}$ of $\mathfrak{g}$ satisfies $[\mathfrak{i}, \mathfrak{g}] \subset \mathfrak{i}$ it is said to be an *ideal*). If $\mathfrak{g}$ has no nonzero ideals whatsoever it is said to be *simple*. In this case the adjoint representation is irreducible. From now on we will restrict our attention to simple Lie algebras.

4. Weyl chambers and the Weyl group

Regarded as maps from $\mathfrak{t}$ to iR, the weights may be thought of as elements of $i\mathfrak{t}^*$, which is, like $\mathfrak{t}$ itself, an l-dimensional vector space over $\mathbb{R}$. It is frequently convenient to identify $i\mathfrak{t}^*$ with an l-dimensional real subspace of $\mathfrak{h}$ via the Cartan–Killing form: $\chi \in i\mathfrak{t}^*$ is identified with $H_\chi \in \mathfrak{h}$, where $\chi(iZ) = \langle H_\chi, iZ \rangle$, $\forall$ $iZ \in \mathfrak{t}$. As the Cartan–Killing form is real-valued on $\mathfrak{g}$ this means that $i\mathfrak{t}^*$ is identified with $i\mathfrak{t} \subset \mathfrak{h}$, and thus becomes a Euclidean space, as $\langle \cdot, \cdot \rangle$ is positive definite on $i\mathfrak{t}$. We will denote this Euclidean space as $\mathscr{V}$ (not to be confused with the module V).

Each root α is normal to a hyperplane

$$\mathscr{V}_\alpha \subset \mathscr{V}: \quad \mathscr{V}_\alpha = \{Z \in \mathscr{V} | \alpha(Z) = \langle H_\alpha, Z \rangle = 0\}.$$

There are half as many hyperplanes $\mathscr{V}_\alpha$ as there are roots, because of the fact that, if α is a root, the only other root proportional to α is $-\alpha$. By removing the hyperplanes $\mathscr{V}_\alpha$, $\mathscr{V}$ is separated into wedge-shaped regions known as *Weyl chambers*. The *Weyl group* is the group generated by the orthogonal (with respect to $\langle \cdot, \cdot \rangle$) reflections in the hyperplanes $\mathscr{V}_\alpha$. The Weyl group permutes the Weyl chambers freely and transitively and maps the root system R into itself.

5. Orderings and simple roots

Next we wish to define an ordering on the weights. A *lexicographic ordering* on $\mathscr{V}$ is determined by choosing an ordered basis $\{e_1, \ldots, e_l\}$ of $\mathscr{V}$ and defining $\wp > 0$ if the first nonvanishing component of $\wp$ with respect to this basis is positive. If one just wishes to order the roots an alternative procedure is to pick one Weyl chamber C_0, which will be called the *positive Weyl chamber*, and define $\alpha > 0 \leftrightarrow \alpha(C_0) > 0$. If α is a root, then $-\alpha$ is a root also. Thus, with respect to any ordering, there are equal numbers of positive and negative roots. The set of positive (negative) roots with respect to some ordering is denoted R_+ (R_-). Given any ordering it is possible to find l positive roots, denoted α_i, $i = 1, \ldots, l$, known as the *simple roots*, with the property that any root is an integer linear combination of these roots with coefficients either all positive or all negative. For the ordering determined by the positive Weyl chamber C_0 the simple roots are those roots α whose orthogonal hyperplanes $\mathscr{V}_\alpha$ bound C_0.

The simple roots are important in the classification of simple Lie algebras. In fact, in order to reconstruct the whole Lie algebra, it is sufficient to know the *Cartan matrix*, which is an $l\times l$ matrix with ijth element $2\langle\alpha_i,\alpha_j\rangle/\langle\alpha_i,\alpha_i\rangle$. Alternatively the essential information about angles between simple roots and their lengths may be coded into the *Dynkin diagrams*, which can be classified combinatorially. We refer to any standard text on Lie algebra theory for further details.

6. A basis for $\mathfrak{g}_c$, Borel, and parabolic subgroups

The root space decomposition (A3) suggests a preferred basis for $\mathfrak{g}_c$, namely, a basis for $\mathfrak{h}$ (e.g., $\{H_j=H_{\alpha_j}|j=1,...,l\}$, where $\{\alpha_j\}$ are the simple roots) together with a choice of basis vectors $\{E_\alpha|\alpha\in R\}$, where $\mathfrak{g}^\alpha=\mathbb{C}E_\alpha$. One is at liberty to choose $E_\alpha=-E^\dagger_{-\alpha}$ and, by an appropriate scaling of E_α, one can cast the bracket relations in $\mathfrak{g}_c$ into the following standard form:

$$[H,E_\alpha]=\alpha(H)E_\alpha,\quad \forall H\in\mathfrak{h},$$
$$[E_\alpha,E_\beta]=\begin{cases}N_{\alpha,\beta}E_{\alpha+\beta}, & \alpha+\beta\in R,\\ H_\alpha, & \alpha+\beta=0,\end{cases}\tag{A5}$$

with all other brackets zero. The $N_{\alpha,\beta}$ are nonzero and integer, and satisfy $N_{\alpha,\beta}=-N_{-\alpha,-\beta}$. With these choices we have

$$\langle E_\alpha,E_\beta\rangle=\begin{cases}1, & \alpha+\beta=0,\\ 0, & \text{otherwise.}\end{cases}\tag{A6}$$

The Lie algebra $\mathfrak{g}$ is the real span of the set

$$\{iH_j,i(E_\alpha+E_{-\alpha}),E_\alpha-E_{-\alpha};\quad j=1,...,l,\quad \alpha\in R\}.$$

As we are only considering compact groups G we will not discuss other real forms of $\mathfrak{g}_c$.

From (A5) and ordering properties it is easy to see that the algebra

$$\mathfrak{b}=\mathfrak{h}+\sum_{\alpha\in R_+}\mathfrak{g}^\alpha\tag{A7}$$

is a subalgebra of $\mathfrak{g}_c$. The corresponding subgroup B is known as a *Borel subgroup* of G_c. Generally, if R_p is a subset of roots containing at least R_+ and closed under addition of roots, the algebra

$$\mathfrak{p}=\mathfrak{h}+\sum_{\alpha\in R_p}\mathfrak{g}^\alpha\tag{A8}$$

is a subalgebra of $\mathfrak{g}_c$ and the corresponding subgroup P of G_c is known as a *parabolic subgroup*. Let B be a Borel subgroup of G_c and $\mathfrak{b}$ its Lie algebra. Then the decomposition $G_c=GB$ corresponding to the Lie algebra decomposition $\mathfrak{g}_c=\mathfrak{g}+\mathfrak{b}$ is known as the *Iwasawa decomposition*. Note that it is not a disjoint decomposition as $\mathfrak{g}\cap\mathfrak{b}=\mathfrak{t}$.

7. More on the weights of a representation

Consider the set of weights of a representation module V. For any irreducible module one can find a weight χ and the corresponding weight vector v [i.e., $t\cdot v=\chi(t)v$, $\forall t\in T$] with the property $E_\alpha v=0$, $\forall\alpha\in R_+$. Such a weight vector is called an *extreme vector* and the corresponding weight is called the *highest weight* of the module. (For any α the vector $E_\alpha v$ is a weight vector with weight $\chi+\alpha$ assuming it is nonzero; thus for an extreme vector all weights obtained in this fashion are smaller than χ.) From the highest weight the whole weight system may be obtained by applying the *lowering algebra*, which consists of polynomials of E_α with $\alpha\in R_-$. Alternatively one can start with a weight vector annihilated by $\{E_\alpha|\alpha\in R_-\}$ and the corresponding *lowest weight*, and then generate the complete weight system by applying the *raising algebra* consisting of polynomials in $\{E_\alpha|\alpha\in R_+\}$.

The criterion for a weight χ to be a highest (resp. lowest) weight of some representation module V is simply that it lies in or on the edge of the positive Weyl chamber C_0 (resp. $-C_0$). Such a weight is said to be *dominant* (*antidominant*). An alternative criterion for χ to be dominant (antidominant) is that χ be greater than or equal to (less than or equal to) the other weights in its Weyl group orbit. If χ is a dominant weight and is not the integer sum of two nonzero dominant weights, then it is said to be a *basic weight*. There are precisely l basic weights, λ_i, $i=1,...,l$, and any weight may be written as an integer linear combination of basic weights. The basic weights may be labeled in such a way that one has

$$2\langle\lambda_i,\alpha_j\rangle/\langle\alpha_j,\alpha_j\rangle=\delta_{ij},\tag{A9}$$

where α_j, $j=1,...,l$, are the simple roots. This formula may be used to obtain the basic weights in terms of the simple roots.

[1] J. J. Duistermaat and G. J. Heckman, Invent. Math. **69**, 259 (1982); **72**, 153 (1983); N. Berline and M. Vergne, Duke. Math. J. **50**, 539 (1983).
[2] V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978).
[3] F. Kirwan, Topology **26**, 37 (1987).
[4] E. Witten, J. Differ. Geom. **17**, 661 (1982); in *Symposium on Anomalies, Geometry and Topology* (World Scientific, Singapore, 1985), pp. 61–99.
[5] M. F. Atiyah, in *Proceedings of the Conference in Honour of L. Schwartz* (Asterisque, Paris, 1985).
[6] M. F. Atiyah and R. Bott, Topology **23**, 1 (1984).
[7] L. Alvarez-Gaumé, Commun. Math. Phys. **90**, 161 (1983); D. Friedan and H. Windey, Nucl. Phys. B **235**, 395 (1984); E. Getzler, Commun. Math. Phys. **92**, 163 (1983).
[8] M. Bordemann, M. Forger, and H. Römer, Commun. Math. Phys. **102**, 605 (1986).
[9] A. A. Rosly and A. S. Schwartz, Commun. Math. Phys. **105**, 645 (1986).
[10] Harish-Chandra, Am. J. Math. **79**, 87 (1957).
[11] A. N. Pressley and G. Segal, *Loop Groups* (Oxford U.P., London, 1987).
[12] D. Gepner and E. Witten, Nucl. Phys. B **278**, 493 (1986).
[13] R. F. Picken, J. Phys. A : Math. Gen. **22**, 2285 (1989).
[14] A. A. Kirillov, *Elements of the Theory of Representations* (Springer, Berlin, 1978).
[15] M. F. Atiyah, Bull. Lond. Math. Soc. **14**, 1 (1982).
[16] A. Borel and J. Tits, Publ. Math. IHES **41**, 253 (1972).
[17] R. Bott, Ann. Math. **66**, 203 (1957).
[18] S. Kobayashi, Tohoku Math. J. **8**, 29 (1956).
[19] S.-S. Chern, *Complex Manifolds Without Potential Theory* (Springer, New York, 1979), 2nd ed., p. 78.
[20] G. J. Heckman, Ph.D. thesis, Leiden, 1980.
[21] J. G. F. Belinfante, *A Survey of Lie Groups and Lie Algebras with Applications and Computational Methods* (SIAM, Philadelphia, 1972).
[22] I. G. Macdonald, in *Representation Theory of Lie Groups, London Mathematical Society Lecture Note Series* (London Math. Soc., London, 1977).

Stochastic action of dynamical systems on curved manifolds. The geodesic interpolation

Ettore Aldrovandi
SISSA/ISAS-Trieste, Strada Costiera 11, I-34100 Trieste, Italy
Daniela Dohrn
Dipartimento di Fisica, Università di Salerno, I-84100 Salerno, Italy
Francesco Guerra
Dipartimento di Fisica, Università di Roma "La Sapienza," I-00185 Roma, Italy

(Received 1 February 1989; accepted for publication 27 September 1989)

Dynamical systems on curved manifolds, with a Lagrangian at most quadratic in the velocity, are considered. The classical action functional, on some finite time interval, is well defined for any smooth trial trajectory in configuration space. The action functional is at the basis of Lagrangian variational principles, from which all dynamical properties of the system can be derived. Here the problem of extending the action functional from the smooth deterministic trajectories of classical mechanics to the very irregular random trajectories of diffusions in configuration space is considered. In this way the action becomes a functional of the trial diffusion processes and can be put at the basis of stochastic variational principles. Since the problem is beset with ultraviolet divergences, the general strategy of renormalization theory is followed, by regularizing the trial diffusion processes through piecewise smooth geodesic lines for a generic given connection on the manifold. After cutoff removal and infinite counterterm subtraction, the quadratic part of the action shows a residual dependence on the generic regularizing connection field. Therefore, in the frame of this geodesic interpolation strategy, it is shown that a change in the connection field is equivalent to a well-defined renormalization of the scalar potential. These results apply to the problem of quantization of generic dynamical systems on curved manifolds, in particular to the definition of Feynman path integrals on curved configuration spaces.

I. INTRODUCTION

In this paper we consider a generic dynamical system with configuration space given by a smooth n-dimensional curved manifold, with Lagrangian at most quadratic in the velocity. In the classical case, it is very simple to calculate the action for a generic smooth trial trajectory, through direct time integration. Then, the physical effectively possible trajectories are selected through the stationary action variational principle, among all conceivable trial trajectories.

On the basis of previous work, related to Nelson's stochastic mechanics,[1] it looks very natural to consider the problem of extending the dynamical action from a functional of smooth trajectories to a functional of a generic given regular diffusion stochastic process on the manifold, in some average sense. Then, it becomes possible to exploit this stochastic action functional in the frame of stochastic variational principles, as done for example in Refs. 1 and 2. It is also known[1,2] that stochastic variational principles can be considered as a possible approach to quantization of dynamical systems in the frame of control theory.

From a structural point of view, the problem of the definition of the stochastic action is very similar to the problem of definition of Feynman path integrals on curved configuration manifolds[3–5] with one important general difference. In fact, Feynman path integrals give a very compact expression for the time development of the quantum wave function starting from some initial condition.[6] On the other hand, the action functional must be exploited in the general frame of stochastic variational principles[1,2] in order to be able to select, through stationarity, the stochastic processes associated to quantum states in the frame of stochastic quantization, provided by Nelson's stochastic mechanics.

Therefore we can say that the problem of the definition of the stochastic action is not only interesting in itself, because it provides an instructive interplay between aspects related to stochasticity and differential geometry, but it is also relevant for the quantization of dynamical systems on curved manifolds.

Two types of difficulties appear in the attempt to define the action as functional of stochastic processes, in the average sense.

The first difficulty is very similar to ultraviolet divergences in quantum field theory. In fact, the random trajectories of diffusions are nowhere differentiable, so that a direct definition of the action is senseless, since time derivatives are essentially involved. It is therefore necessary to introduce a regularization, smoothing out the time dependence. By following a very natural strategy in the theory of stochastic processes,[7–9] we perform a kind of geodesic regularization, through the substitution of the random trajectories with their piecewise smooth interpolation in small time intervals, with geodesic lines, with respect to a given regularizing connection field. After cutoff removal, i.e., the regularizing time interval going to zero, some divergent terms appear. However, it was a very important discovery of Nelson's[1] that these divergent terms are irrelevant, in the frame of the variational principles, because they disappear in the expression of the *variation* of the action. In any case, in general, it is very easy to deal with them, in the spirit of renormalization theory, by substitution of divergent constants with finite, but arbitrary, renormalized constants.

The second difficulty is very subtle and originates from geometric features, enhanced by the stochastic nature of the trial processes. In fact, since we are dealing with a generic configuration manifold, the regularization procedure has a local character, given by the regularizing connection field.

Since the trial processes have very irregular trajectories with Brownian-like properties, they make, so to speak, a very detailed exploration of the local geometric properties of the manifold. As a result, the stochastic action, even after elimination of the divergent infinite terms, is sensitive to the regularization procedure. In fact, we show that the quadratic part (in the velocity) of the action has a residual dependence on the chosen connection field, playing a central role in the regularization procedure. However, the effects of a change in the connection field can be completely absorbed into a suitable change of the scalar potential in the Lagrangian, of the second order in the diffusion constant.

Our results put into a general geometric perspective the old problem of whether the quantum Schrödinger equation on a curved manifold does contain or not a term involving the curvature invariant.[3–5,10] In fact, in the most general case, the quantization procedure, starting from a classical system, seems to be able to define the total action, up to a term in the scalar potential of second order with respect to the diffusion constant, and therefore also to the Planck constant.

In other words, at least in the frame of the regularization based on geodesic interpolation or equivalent means, the quantum theory, associated to a given classical theory is subject to a scalar potential defined but for a possible additive contribution, of second order with respect to Planck's constant, and therefore negligible in the semiclassical limit.

In a forthcoming paper,[11] as an additional proof of the subtleties of the interplay between geometric and stochastic aspects, we show that the exploitation of the developing map[5,11,12] (Cartan map) leads to a quite different qualitative structure, where the curvature term in the quantum Schrödinger equation does not arise at all, in agreement with the conclusions of Refs. 5 and 10.

The organization of the paper is as follows. In Sec. II, we recall all basic aspects of the kinematics of controlled diffusions on a manifold. Geodesic interpolation, with respect to a generic given connection field Γ, is also introduced and fully characterized up to order $O(\Delta t^{3/2})$, beyond the reach of ordinary stochastic differential calculus usually limited to $O(\Delta t)$. Section III has a technical nature. We show how to calculate higher-order corrections to correlation functions for the increments Δq, starting from the basic controlling fields and exploiting Markov property. In Sec. IV we introduce canonical Lagrangians of the second order with respect to the velocity, involving a generic kinetic tensor field $g_{ij}(x,t)$, a vector field $A_i(x,t)$, and a scalar potential $V(x,t)$. We define a regularized stochastic action functional by exploiting the well-known geodesic interpolation for diffusions. We show how to remove the regularization for the simple cases of the scalar and vector potentials. Section V contains the main results of this paper. We give the full expression of the contribution to the stochastic action functional coming from the kinetic term.

It is important to remark that in our treatment we consider a generic controlling covariance field η^{ij} and a generic kinetic field g_{ij}, not being necessarily correlated at this stage. In fact, η and g play a completely different role and must be kept as distinct objects, in the procedure of definition of the action functional (but see also Ref. 13).

Moreover, the regularizing connection field Γ is also completely generic, not being related to η or g. This is necessary, in order to have a complete independence of the procedure from the trial stochastic processes and a resulting action linear in the general fields g, A, and V.

In Sec. VI we study the dependence of the stochastic action on geodesic gauge transformations arising from a change in the regularizing connection field Γ.

Finally, Sec. VII is devoted to conclusions and outlook for future developments.

II. KINEMATICS OF CONTROLLED DIFFUSION ON MANIFOLDS

In this section we give a short review of all properties related to kinematics of controlled diffusions on manifolds, which will be exploited in the following. Let M be a generic smooth, n-dimensional manifold, taken as the configuration space of the system. We call TM_x the tangent space at the point $x \in M$ and TM the tangent bundle.

Let us introduce smooth, C^∞ bounded, controlling fields, $b_{(+)}$ and η, given in each local chart by the components

$$b^i_{(+)}(x,t), \quad \eta^{ij}(x,t),$$
$$x \in M, \quad t \in [t_0,t_1], \quad i,j = 1,2,\dots,n. \tag{2.1}$$

We consider controlled diffusions as Markov processes on the manifold

$$[t_0,t_1] \ni t \to q(t) \in M, \tag{2.2}$$

whose time development is fully specified by the local conditions

$$E(\Delta q^i(t) \,|\, q(t) = x) = b^i_{(+)}(x,t)\Delta t + O(\Delta t^2), \tag{2.3}$$

$$E(\Delta q^i(t)\Delta q^j(t) \,|\, q(t) = x) = 2\nu\eta^{ij}(x,t)\Delta t + O(\Delta t^2), \tag{2.4}$$

where

$$\Delta q(t) \equiv q(t+\Delta t) - q(t), \quad \Delta t > 0.$$

Here $E(\cdot \,|\, q(t) = x)$ denote conditional expectations with respect to all trajectories starting from the generic point $x \in M$ at time t_0. As a shorthand notation, we will write them sometimes in the form $E_{x,t}$. In general, we will denote E_t the conditional expectation with respect to the σ algebra generated by $q(t)$ and by E the expectations. Obviously we have $E(E_t(\cdot)) = E(\cdot)$. The diffusion coefficient ν in (2.4) is a given positive constant. We notice also that the controlling covariance matrix η must be positive semidefinite at each point.

On the basis of the standard theory of parabolic differential equations on manifolds, it is very well known, under suitable regularity conditions, that the controlling equations (2.3) and (2.4) define completely the process provided its initial density is given. For example,[9] we can assume

$$\begin{aligned} M \quad & n - \text{dimensional } C^\infty - \text{manifold}, \\ b_{(+)}, \eta \quad & C^\infty - \text{fields on } M, \\ \eta \quad & \text{positive definite on } M. \end{aligned} \tag{2.5}$$

It is convenient to introduce an arbitary fixed smooth invariant measure $d\mu(x)$ on the manifold and define the relative invariant density field of the process, $\rho(x,t)$, so that for the expectations we have

$$E(F(q(t),t)) = \int_M F(x,t)\, \rho(x,t) d\mu(x) , \tag{2.6}$$

for any smooth function F with values $F(x,t)$.

In the following, without effective loss of generality, we consider regular processes, corresponding to the regularity assumptions given in (2.5) for the controlling fields $b_{(+)},\eta$, with the additional assumption that $\rho(\cdot,t_0)$ is C^∞ bounded and everywhere positive on M, so that the same properties are preserved for all $t\in[t_0,t_1]$.

These very strong smoothness assumptions allow a drastic simplification of all mathematical treatment. However, in the applications, for example to the stochastic variational principles, it is necessary to consider, in some cases, also processes which are not regular according to the definition given before. This extension can be easily done, by exploiting limiting procedures of the type described in Ref. 14, where it is shown how to construct processes with possibly very singular drift fields and zeros in the density, exploiting remarkable stability estimates given by Carlen in Ref. 15.

For the moment, we consider some properties of the stochastic flow given by (2.3) and (2.4), not involving features related to the density.

In the following, we deal very frequently with stochastic increments, associated to $\Delta t > 0$. In order to simplify notations and evaluations, it is convenient to introduce stochastic symbols of the type

$$O(\Delta t^s), \quad s = 0,1/2,1,3/2,..., \quad \Delta t > 0 . \tag{2.7}$$

We assume the following conditions to hold for random variables $a(\Delta t)$, $b(\Delta t)$,

$$\begin{aligned} & a(\Delta t) = O(\Delta t^s) \\ & \quad \Rightarrow E_{x,t}(a(\Delta t)) \\ & \quad = \begin{cases} O(\Delta t^s), & \text{if } s \text{ is integer}, \\ O(\Delta t^{s+(1/2)}), & \text{if } s \text{ is half-integer}, \end{cases} \end{aligned} \tag{2.8}$$

$$\begin{aligned} & a(\Delta t) = O(\Delta t^s) , \\ & b(\Delta t) = O(\Delta t^{s'}) \\ & \quad \Rightarrow a(\Delta t) b(\Delta t) = O(\Delta t^{s+s'}) . \end{aligned} \tag{2.9}$$

Notice that the stochastic symbols $O(\Delta t^s)$ give always rise to ordinary symbols $O(\Delta t^k)$ under conditional expectation.

With these definitions, we can write the basic property of diffusions in the form

$$\Delta q^i(t) = O(\Delta t^{1/2}) \tag{2.10}$$

and check its agreement with the controlling equations (2.3) and (2.4) and all their consequences.

Let us now recall a very important property of stochastic increments. In the classical case, we know that Δq^i transform as the components of a vector in $TM_{q(t)}$, but for error terms of order $O(\Delta t^2)$. The situation is deeply different in the stochastic case. Here, as a consequence of (2.10), Δq fails to be a vector for very relevant first-order terms in Δt. As a result we find that $b_{(+)}$ is not a vector. In fact, we find

$$b'^{i}_{(+)}(x',t) = \partial_j X^i b^j_{(+)}(x,t) + \nu\, \partial^2_{jk} X^i \eta^{jk}(x,t) . \tag{2.11}$$

On the other hand, we can easily check, starting from the product of the expansion (2.9) taken for two different components, that η does indeed transform as a second-order tensor

$$\eta'^{ij}(x',t) = \partial_k X^i \partial_l X^j \eta^{kl}(x,t) . \tag{2.12}$$

In order to have a covariant scheme, it is convenient, by following the methods in Refs. 16, 10, and 1, to introduce a generic smooth symmetric connection field Γ, with components $\Gamma^i_{jk}(x) = \Gamma^i_{kj}(x)$ in each local chart transforming as

$$\Gamma \to \Gamma', \quad \Gamma^k_{ij}\, \partial_k X^q = \partial_i X^l \partial_j X^m \Gamma'^q_{lm} + \partial^2_{ij} X^q . \tag{2.13}$$

Consider now the points $q(t) = x$ and $q(t+\Delta t)$, $\Delta t > 0$. For the sake of simplicity, assume that $q(t+\Delta t)$ is so near to x that there is only one geodesic line y for Γ, connecting $q(t)$ to $q(t+\Delta t)$. With respect to an affine coordinate s we have

$$\begin{aligned} & [0,1] \ni s \to y(s) \in M , \\ & y(0) = q(t) = x , \\ & y(1) = q(t+\Delta t) , \\ & \ddot{y}^i(s) + \Gamma^i_{jk}(y(s))\, \dot{y}^j(s)\, \dot{y}^k(s) = 0 . \end{aligned} \tag{2.14}$$

Let us define the geodesic initial tangent

$$\widetilde{\Delta q} = \dot{y}(0) \in TM_x . \tag{2.15}$$

Obviously we have

$$\widetilde{\Delta q} = O(\Delta t^{1/2}) . \tag{2.16}$$

Through a very simple calculation we can prove the following.

Theorem 1 (Geodesic interpolation up to order $O(\Delta t^{3/2})$): Under the stated regularity conditions we have

$$\Delta q^i = \widetilde{\Delta q^i} - \tfrac{1}{2} \Gamma^i_{jk} \widetilde{\Delta q^j} \widetilde{\Delta q^k} + \tfrac{1}{6} S^i_{jkl} \widetilde{\Delta q^j} \widetilde{\Delta q^k} \widetilde{\Delta q^l} + O(\Delta t^2) , \tag{2.17}$$

$$\widetilde{\Delta q^i} = \Delta q^i + \tfrac{1}{2} \Gamma^i_{jk} \Delta q^j \Delta q^k + \tfrac{1}{6} \widetilde{S}^i_{jkl} \Delta q^j \Delta q^k \Delta q^l + O(\Delta t^2) , \tag{2.18}$$

$$S^i_{jkl} = \mathscr{S}_{(jkl)} (2\Gamma^i_{sj} \Gamma^s_{kl} - \partial_j \Gamma^i_{kl}) , \tag{2.19}$$

$$\widetilde{S}^i_{jkl} = \mathscr{S}_{(jkl)} (\Gamma^i_{sj} \Gamma^s_{kl} + \partial_j \Gamma^i_{kl}) . \tag{2.20}$$

In (2.19), (2.20), $\mathscr{S}_{(\dots)}$ is the symmetrizer with respect to the $(\cdots)$ indices, i.e.,

$$\mathscr{S}_{(jk)}(a_{jk}) = \tfrac{1}{2}(a_{jk} + a_{kj}) , \text{ etc.} \tag{2.21}$$

Proof: Starting from the definitions in (2.14), we notice

$$\Delta q(t) = q(t+\Delta t) - q(t) = y(1) - y(0) = \int_0^1 \dot{y}(s) ds . \tag{2.22}$$

Therefore $\dot{y}(s) = O(\Delta t^{1/2})$. Moreover, from the geodesic equation we see also $\ddot{y}(s) = O(\Delta t)$. By differentiation we have also

$$y^{\mathrm{III}\,i}(s) + \Gamma^i_{jk}\, \dot{y}^j \ddot{y}^k + \partial_l \Gamma^i_{jk}\, \dot{y}^l \dot{y}^j \dot{y}^k = 0 , \tag{2.23}$$

which gives $y^{\mathrm{III}\,i}(s) = O(\Delta t^{3/2})$. Analogously, we would find $y^{\mathrm{IV}}(s) = O(\Delta t^2)$, etc. Therefore we can write for the interpolating geodesic

$$y(s) = y(0) + s\dot{y}(0) + \tfrac{1}{2}s^2\ddot{y}(0) + \tfrac{1}{6}s^3 y^{\mathrm{III}}(0) + O(\Delta t^2), \tag{2.24}$$

where, by definitions (2.14) and (2.15) and Eq. (2.23),

$$\begin{aligned} &q(t) = x, \quad \dot{y}(0) = \widetilde{\Delta q}, \quad \ddot{y}^i(0) = -\Gamma^i_{jk}\widetilde{\Delta q^j}\widetilde{\Delta q^k}, \\ &y^{\mathrm{III}\,i}(0) = (-\partial_j\Gamma^i_{kl} + 2\Gamma^i_{sj}\Gamma^s_{kl})\widetilde{\Delta q^j}\widetilde{\Delta q^k}\widetilde{\Delta q^l}. \end{aligned} \tag{2.25}$$

Here Γ and $\partial\Gamma$ are evaluated at $x = q(t)$. But we have also

$$\begin{aligned} \Delta q &= y(1) - y(0) \\ &= \dot{y}(0) + \tfrac{1}{2}\ddot{y}(0) + \tfrac{1}{6}y^{\mathrm{III}}(0) + O(\Delta t^2), \end{aligned} \tag{2.26}$$

and (2.17) follows. Solving (2.17) with respect to $\widetilde{\Delta q}$, by iteration, taking into account only the relevant terms, we have also (2.18). Q.E.D.

In this theorem, we have pushed the evaluation up to order $\Delta t^{3/2}$, because we need these results in the following. Here we exploit only the $O(\Delta t)$ part. In fact, let us introduce the vector field $v_{(+)}(x,t)$, so that

$$E(\widetilde{\Delta q^i}|q(t) = x) = v^i_{(+)}(x,t)\Delta t + O(\Delta t^2). \tag{2.27}$$

From (2.18) we find

$$v^i_{(+)}(x,t) = b^i_{(+)}(x,t) + \nu\Gamma^i_{jk}(x)\eta^{jk}(x,t). \tag{2.28}$$

One can easily check, starting from the transformation properties (2.13) of Γ^i_{jk} and (2.11), that $v^i_{(+)}$ do indeed transform as the components of a vector, as it is in any case obvious from the definition (2.15). We see that the stochastic flow can be defined either through the couple $(b_{(+)},\eta)$ or through $(v_{(+)},\eta,\Gamma)$. Due to (2.28), a change in the connection field Γ can be completely absorbed into a change of $v_{(+)}$, giving rise to the same $b_{(+)}$, provided

$$v^i_{(+)} - \nu\Gamma^i_{jk}\eta^{jk} = v'^i_{(+)} - \nu\Gamma'^i_{jk}\eta^{jk} \quad (\text{same } \eta). \tag{2.29}$$

We call geodesic gauge transformations changes of the type

$$\Gamma \to \Gamma', \quad v_{(+)} \to v'_{(+)}, \tag{2.30}$$

such that (2.29) is satisfied. They keep invariant the controlled stochastic flow.

Mean forward derivatives of scalar fields $\varphi(x,t)$ are defined in the usual way[1,17]

$$\begin{aligned} &(D_{(+)}\varphi)(x,t) = \lim_{\Delta t \to 0^+}(\Delta t)^{-1}E(\Delta\varphi(t)|q(t) = x), \\ &\Delta\varphi(t) \equiv \varphi(q(t+\Delta t), t+\Delta t) - \varphi(q(t),t). \end{aligned} \tag{2.31}$$

Therefore one can also write

$$E_t(\Delta\varphi(t)) = (D_{(+)}\varphi)(q(t),t)\Delta t + O(\Delta t^2). \tag{2.32}$$

A simple calculation[1,17] gives

$$(D_{(+)}\varphi)(x,t) = (\partial_t\varphi + b^i_{(+)}\partial_i\varphi + \nu\eta^{ij}\partial^2_{ij}\varphi)(x,t). \tag{2.33}$$

It is also convenient to introduce covariant derivatives ∇ with respect to the connection Γ such that

$$\begin{aligned} &\nabla_i\varphi = \partial_i\varphi, \\ &\nabla_i A^j = \partial_i A^j + \Gamma^j_{si}A^s, \\ &\nabla_i B_j = \partial_i B_j - \Gamma^s_{ji}B_s, \quad \text{etc.} \end{aligned} \tag{2.34}$$

Then we can define the invariant operator[7,8,16,10]

$$D^{\mathrm{I}}_{(+)} = \partial_t + v^i_{(+)}\nabla_i + \nu\eta^{ij}\nabla_i\nabla_j \tag{2.35}$$

and check that $D^I_{(+)}\varphi = D_{(+)}\varphi$ on scalar fields. Moreover, one can easily verify the invariance of (2.35), when acting on scalar fields, under geodesic gauge transformations (2.30).

However, it is convenient to remark that the operators (2.33) and (2.35) have different effects when acting on components of generic tensor fields (other than scalars).

According to standard procedures[1,17] it is also convenient to introduce backward controlling fields $b_{(-)}$, $v_{(-)}$.

For the sake of simplicity, we assume the invariant measure $d\mu(x)$ in (2.6) to be compatible with the given connection Γ, so that we can freely perform integrations by parts on ∇, in the case of smooth fast decreasing fields, as for example,

$$\int_M \psi^i(x)(\nabla_i\varphi)(x)d\mu(x) = -\int_M(\nabla_i\psi^i)(x)\varphi(x)d\mu(x). \tag{2.36}$$

The compatibility condition between Γ and $d\mu$, by assuming a local form of the type $d\mu(x) = \mu(x)dx$, where dx is the Lebesgue measure in a local chart, can be written as

$$\Gamma^s_{si} = \partial_i \log\mu. \tag{2.37}$$

It is surely satisfied if, for example, Γ is the symmetric canonical Riemann connection associated to a metric field h_{ij} and $\mu = \sqrt{\det(h_{(\cdot\cdot)})}$. Under geodesic gauge transformations given by (2.29), (2.30) the local densities $\mu(x)$ should transform so that

$$\Gamma \to \Gamma' = \Gamma + \delta\Gamma, \quad \mu \to \mu', \quad \delta\Gamma^s_{si} = \partial_i\log(\mu'/\mu). \tag{2.38}$$

The transformation law for the relative densities of the process, as defined in (2.6), must be

$$\rho \to \rho' = \rho(\mu/\mu'). \tag{2.39}$$

We introduce the backward control field $b^i_{(-)}(x,t)$, so that for the backward increments we have

$$\begin{aligned} &\Delta q_{(-)}(t) \equiv q(t) - q(t-\Delta t), \quad \Delta t > 0, \\ &E(\Delta q^i_{(-)}(t)|q(t) = x) = b^i_{(-)}(x,t)\Delta t + O(\Delta t^2), \\ &E(\Delta q^i_{(-)}(t)\Delta q^j_{(-)}(t)|q(t) = x) \\ &\quad = 2\nu\eta^{ij}(x,t)\Delta t + O(\Delta t^2). \end{aligned} \tag{2.40}$$

We also introduce generic backward increments for scalar fields,

$$\Delta_{(-)}\varphi = \varphi(q(t),t) - \varphi(q(t-\Delta t), t-\Delta t), \quad \Delta t > 0, \tag{2.41}$$

and define mean backward derivatives $D_{(-)}$ as in (2.31) with $\Delta\varphi$ substituted by $\Delta_{(-)}\varphi$. Then one has

$$\begin{aligned} (D_{(-)}\varphi)(x,t) = &(\partial_t\varphi)(x,t) + b^i_{(-)}(x,t)\partial_i\varphi(x,t) \\ &- \nu\eta^{ij}(x,t)\partial^2_{ij}\varphi(x,t). \end{aligned} \tag{2.42}$$

As in the flat case[1,17] we can apply the Nelson formula

$$\frac{d}{dt}E(F(q(t),t)G(q(t),t))$$
$$= E((D_{(+)}F)(q(t),t)G(q(t),t)) + E(F(q(t),t)(D_{(-)}G)(q(t),t)), \quad (2.43)$$

and obtain (see also Ref. 18)

$$b^i_{(-)}(x,t) = b^i_{(+)}(x,t) - 2\nu\partial_j(\rho\mu\eta^{ij})/\rho\mu, \quad (2.44)$$

together with the forward Fokker–Planck equation

$$\mu\,\partial_t\rho = -\partial_i(\mu\rho b^i_{(+)}) + \nu\,\partial_i\partial_j(\mu\rho\eta^{ij}). \quad (2.45)$$

In order to write (2.44) in a covariant form, we exploit (2.34), (2.37) and arrive at

$$b^i_{(-)}(x,t) - \nu\Gamma^i_{jk}(x)\eta^{jk}(x,t) = b^i_{(+)} + \nu\Gamma^i_{jk}\eta^{jk} - 2\nu\rho^{-1}\nabla_j(\rho\eta^{ij}). \quad (2.46)$$

Recalling the definition (2.28) of $v_{(+)}$, we can introduce

$$v^i_{(-)}(x,t) = b^i_{(-)} - \nu\Gamma^i_{jk}\eta^{jk} \quad (2.47)$$

and write the following relation between the two fields:

$$v^i_{(-)} = v^i_{(+)} - 2\nu\rho^{-1}\nabla_j(\rho\eta^{ij}). \quad (2.48)$$

This should be compared with analogous expressions exploited in Refs. 1, 16, and 10, but it should be remarked that here the connection Γ giving rise to ∇ is completely arbitrary and not related to η.

Due to geodesic gauge invariance of $\mu\rho$, coming from (2.39), we see that $b_{(-)}$ is also invariant. We have also

$$b \equiv \tfrac{1}{2}(b_{(+)} + b_{(-)}) = \tfrac{1}{2}(v_{(+)} + v_{(-)}) \equiv v. \quad (2.49)$$

Therefore v is also invariant, while $v_{(+)}$ [recall (2.29)] and $v_{(-)}$ transform as

$$\begin{aligned} v^i_{(+)} &\to v'^i_{(+)} = v^i_{(+)} + \nu\delta\Gamma^i_{jk}\eta^{jk}, \\ v^i_{(-)} &\to v'^i_{(-)} = v^i_{(-)} - \nu\delta\Gamma^i_{jk}\eta^{jk}, \\ \Gamma &\to \Gamma' = \Gamma + \delta\Gamma. \end{aligned} \quad (2.50)$$

The physical interpretation of $v_{(-)}$ is very simple. In fact, introducing the backward geodesic starting at $q(t-\Delta t)$ for $s=0$ and arriving in $q(t) = x$ for $s=1$, one can immediately see that $v_{(-)}$ coincides with the s derivative of the geodesic at the *final* point $s=1$.

We can also introduce the osmotic velocity

$$u^i = \tfrac{1}{2}(v^i_{(+)} + v^i_{(-)}) = \nu\rho^{-1}\nabla_j(\rho\eta^{ij}). \quad (2.51)$$

It is a vector field under coordinate transformations, but under geodesic gauge transformations (2.29), (2.30) it behaves as

$$u^i \to u'^i = u^i + \nu\delta\Gamma^i_{jk}\eta^{jk}. \quad (2.52)$$

Finally (2.45) can be written in the explicitly covariant forms

$$\begin{aligned} \partial_t\rho &= -\nabla_i(\rho v^i_{(\pm)}) \pm \nu\nabla_i\nabla_j(\rho\eta^{ij}), \\ \partial_t\rho &= -\nabla_i(\rho v^i). \end{aligned} \quad (2.53)$$

Moreover the backward derivative (2.42) is written in the equivalent form

$$(D_{(-)}\varphi)(x,t) = \partial_t\varphi + v^i_{(-)}\nabla_i\varphi - \nu\eta^{ij}\nabla_i\nabla_j\varphi. \quad (2.54)$$

In many applications, the following transport formula is very useful:

$$E(F(q(t_1),t_1)|q(t_0) = x_0) = F(x_0,t_0) + \int_{t_0}^{t_1} E((D_{(+)}F)(q(t),t)|q(t_0) = x_0)dt. \quad (2.55)$$

It is easily interpreted as the integrated form of (2.32). An analogous formula holds also for $D_{(-)}$, with final conditioning.

III. HIGHER-ORDER CORRECTIONS TO CORRELATIONS

In the following we will need the explicit expressions of the higher-order terms, up to $O(\Delta t^2)$, arising in (2.4) and in the analogous correlations involving three and four components of the increments Δq.

These are easily found, by exploiting (2.3), (2.4), and Markov property, through a rescaling of the time interval Δt, in the frame of techniques very similar to the renormalization group in quantum field theory and statistical mechanics. Therefore, let us establish the following.

Theorem 2 (higher-order corrections to correlations): Under the stated regularity assumptions given in the previous section, we have in each local chart

$$E(\widetilde{\Delta q^i}(t)\widetilde{\Delta q^j}(t)|q(t) = x) = 2\nu\eta^{ij}(x,t)\Delta t + (b^i_{(+)}b^j_{(+)} + \nu(\eta^{ik}\partial_k b^j_{(+)} + \eta^{jk}\partial_k b^i_{(+)} + D_{(+)}\eta^{ij}))(x,t)\Delta t^2 + O(\Delta t^3), \quad (3.1)$$

$$E(\widetilde{\Delta q^i}(t)\widetilde{\Delta q^j}(t)\widetilde{\Delta q^k}(t)|q(t) = x) = 2\nu(\eta^{ij}b^k_{(+)} + \eta^{jk}b^i_{(+)} + \eta^{ki}b^j_{(+)})(x,t)\Delta t^2 + 2\nu^2(\eta^{il}\partial_l\eta^{jk} + \eta^{jl}\partial_l\eta^{ki} + \eta^{kl}\partial_l\eta^{ji})(x,t)\Delta t^2 + O(\Delta t^3), \quad (3.2)$$

$$E(\widetilde{\Delta q^i}(t)\widetilde{\Delta q^j}(t)\widetilde{\Delta q^k}(t)\widetilde{\Delta q^l}(t)|q(t) = x) = 2\nu^2(\eta^{ij}\eta^{kl} + \eta^{ik}\eta^{jl} + \eta^{il}\eta^{jk})(x,t)\Delta t^2 + O(\Delta t^3). \quad (3.3)$$

Proof: We give the explicit proof of (3.1), following our general method. Then, (3.2) and (3.3) are proven through a very simple adaptation.

Let us write the $O(\Delta t^2)$ term on the right-hand side of (2.13) in the form

$$E_{x,t}(\Delta q^i\Delta q^j) = 2\nu\eta^{ij}(x,t)\Delta t + a^{ij}(x,t)\Delta t^2 + O(\Delta t^3), \quad (3.4)$$

where a^{ij} is unknown for the moment.

Consider, for the sake of simplicity, the midpoint splitting of the time interval Δt and define

$$\begin{aligned} t' &= t + (\Delta t/2), \quad \Delta' q = q(t') - q(t), \\ \Delta'' q &= q(t+\Delta t) - q(t'), \\ \Delta q &= \Delta' q + \Delta'' q. \end{aligned} \quad (3.5)$$

Then the left-hand side of (3.4) becomes the sum of four terms involving $\Delta' q$ and $\Delta'' q$ instead of Δq.

$$E_{x,t}(\Delta q^i \Delta q^j) = E_{x,t}(\Delta' q^i \Delta' q^j) + E_{x,t}(\Delta' q^i \Delta'' q^j) + E_{x,t}(\Delta'' q^i \Delta' q^j) + E_{x,t}(\Delta'' q^i \Delta'' q^j) . \tag{3.6}$$

For the first term, in analogy with (3.4) written for $\Delta t/2$, we have

$$E_{x,t}(\Delta' q^i \Delta' q^j) = 2\nu\eta^{ij}(x,t)(\Delta t/2) + a^{ij}(x,t)(\Delta t/2)^2 + O(\Delta t^3) . \tag{3.7}$$

For the second term, we can write

$$\begin{aligned} E_{x,t}(\Delta' q^i \Delta'' q^j) &= E_{x,t}(\Delta' q^i E_{t'} \Delta'' q^j) \\ &= E_{x,t}(\Delta' q^i b^j_{(+)}(q(t'),t'))(\Delta t/2) \\ &\quad + O(\Delta t^3) , \end{aligned} \tag{3.8}$$

where first we have exploited Markov property, denoting with $E_{t'}$ the conditional expectation with respect to the σ algebra generated by $q(t')$, and then we have exploited (2.3) to evaluate

$$E_{t'} \Delta'' q^j = b^j_{(+)}(q(t'),t')(\Delta t/2) + O(\Delta t^2) . \tag{3.9}$$

Moreover, we have

$$b^j_{(+)}(q(t'),t') = b^j_{(+)}(q(t),t) + (\partial_k b^j_{(+)})(q(t'),t')\Delta' q^k + O(\Delta t) . \tag{3.10}$$

Therefore, collecting all terms in (3.8), we conclude that

$$\begin{aligned} E_{x,t}(\Delta' q^i \Delta'' q^j) &= (b^i_{(+)} b^j_{(+)})(x,t)(\Delta t/2)^2 \\ &\quad + 2\nu(\eta^{ik}\partial_k b^j_{(+)})(x,t)(\Delta t/2)^2 \\ &\quad + O(\Delta t^3) . \end{aligned} \tag{3.11}$$

The third term is identical to (3.11), with i,j interchanged.

For the fourth term, we have

$$\begin{aligned} E_{x,t}(\Delta'' q^i \Delta'' q^j) &= E_{x,t}(E_{t'} \Delta'' q^i \Delta'' q^j) \\ &= E_{x,t}(2\nu\eta^{ij}(q(t'),t')(\Delta t/2) \\ &\quad + a^{ij}(q(t'),t')(\Delta t/2)^2) + O(\Delta t^3) . \end{aligned} \tag{3.12}$$

Now we exploit the analog of (2.32), written for η and a instead of φ, and conclude

$$\begin{aligned} E_{x,t}(\Delta'' q^i \Delta'' q^j) &= 2\nu\eta^{ij}(x,t)(\Delta t/2) \\ &\quad + 2\nu(D_{(+)}\eta^{ij})(x,t)(\Delta t/2)^2 \\ &\quad + a^{ij}(x,t)(\Delta t/2)^2 + O(\Delta t^3) . \end{aligned} \tag{3.13}$$

Finally, we substitute the four terms, so found, in (3.4) and solve for a^{ij}, arriving at (3.1). Similar calculations lead to (3.2) and (3.3). Q.E.D.

We notice also that next higher-order terms could be easily calculated with the same method.

Of course, Eqs. (3.1)–(3.3) hold in any local chart and their covariance properties are very complicated, due to the peculiar transformation properties of Δq^i pointed out in Sec. II. In order to have an explicitly covariant formulation, we can introduce the correlations of Δq, defined in (2.18). Exploiting (2.18) and (2.34), through a very simple direct calculation we arrive at the following.

Theorem 3 (higher-order corrections for the geodesic tangent): Under the stated regularity conditions, we have

$$\begin{aligned} E(\widetilde{\Delta q}^i(t)\widetilde{\Delta q}^j(t)\,|\,q(t)=x) &= 2\nu\eta^{ij}(x,t)\Delta t + (v^i_{(+)}v^j_{(+)} + \nu(\eta^{ik}\nabla_k v^j_{(+)} + \eta^{jk}\nabla_k v^i_{(+)} + D^I_{(+)}\eta^{ij}) \\ &\quad + \tfrac{1}{3}\nu^2(R^{\ j}_{kls}\eta^{si} + R^{\ i}_{kls}\eta^{sj})\eta^{kl})(x,t)\Delta t^2 + O(\Delta t^3) , \end{aligned} \tag{3.14}$$

$$\begin{aligned} E(\widetilde{\Delta q}^i(t)\widetilde{\Delta q}^j(t)\widetilde{\Delta q}^k(t)\,|\,q(t)=x) &= 2\nu(\eta^{ij}v^k_{(+)} + \eta^{jk}v^i_{(+)} + \eta^{ki}v^j_{(+)})(x,t)\Delta t^2 \\ &\quad + 2\nu^2(\eta^{il}\nabla_l\eta^{jk} + \eta^{jl}\nabla_l\eta^{ki} + \eta^{kl}\nabla_l\eta^{ji})(x,t)\Delta t^2 + O(\Delta t^3) , \end{aligned} \tag{3.15}$$

$$E(\widetilde{\Delta q}^i(t)\widetilde{\Delta q}^j(t)\widetilde{\Delta q}^k(t)\widetilde{\Delta q}^l(t)\,|\,q(t)=x) = 2\nu^2(\eta^{ij}\eta^{kl} + \eta^{ik}\eta^{jl} + \eta^{il}\eta^{jk})(x,t)\Delta t^2 + O(\Delta t^3) , \tag{3.16}$$

where R is the curvature tensor associated to Γ,

$$R^{\ i}_{jkl} = \partial_k\Gamma^i_{jl} - \partial_l\Gamma^i_{jk} + \Gamma^i_{sk}\Gamma^s_{jl} - \Gamma^i_{sl}\Gamma^s_{jk} . \tag{3.17}$$

IV. GEODESIC INTERPOLATION AND THE REGULARIZED STOCHASTIC ACTION

On the configurational manifold M, we consider dynamical systems with the following form of canonical Lagrangian at each time t:

$$\mathscr{L}(x,v;t) = \tfrac{1}{2}mg_{ij}(x,t)v^iv^j + A_i(x,t)v^i - V(x,t) ,$$
$$x\in M, \quad v\in TM_x, \quad t\in[t_0,t_1] . \tag{4.1}$$

Here V, A, and g are given smooth external fields, possibly time dependent. Their transformation properties are as a scalar, a covariant vector, i.e., $A\in T^*M_x$, and a covariant symmetric tensor, respectively. The constant $m>0$ has been introduced for dimensional reasons.

In the classical case, for a generic smooth trajectory, $t\to q(t)$, the action functional for the time interval $t\in[t_0,t_1]$ is defined by

$$A(t_0,t_1;q) = \int_{t_0}^{t_1} \mathscr{L}(q(t),\dot q(t);t)dt . \tag{4.2}$$

We would like to extend, in some average sense, the definition (4.2) of the action, to the case where the smooth deterministic classical trajectory is substituted by a generic given controlled diffusion, of the type introduced in Sec. II.

Of course, the main difficulty is given by the fact that the time derivative $\dot q(t)$ is not well defined. In the spirit of renormalization procedures in quantum field theory and according to general methods exploited in the theory of stochastic processes,[1,5,7–9] we firstly perform a regularization and then remove it, with the purpose of isolating the infinite terms.

The results are essentially equivalent for a large class of regularization procedures. The following one, called geodesic interpolation, looks very natural, because of its geometric invariant character (see also Ref. 1).

Let us recall that in our treatment the controlling covariance matrix η, the kinetic matrix g, and the auxiliary connection Γ are completely generic, and not related *a priori* among themselves.

Let us split the time interval $[t_0,t_1]$ into K equal pieces of length Δt, so that $K\Delta t = t_1 - t_0$. Let us call t the initial points of each interval,

$$t = t_0, \quad t_0 + \Delta t, \quad t_0 + 2\Delta t, \ldots, \quad t_1 - \Delta t. \tag{4.3}$$

Call t' the dummy integration variable in the generic time interval $[t,t+\Delta t]$, and s the corresponding affine parameter, so that $t' = t + s\Delta t$.

Consider a generic smooth diffusion q. Let $q(t)$ be the points on M reached at the discrete times t, defined before. Call $\hat{q}(t')$ the interpolating geodesic between $q(t)$ and $q(t+\Delta t)$, with reference to a given symmetric connection Γ, as explained in Sec. II, so that

$$\hat{q}(t') = y(s), \quad t' = t + s\Delta t, \quad 0 \leqslant s \leqslant 1. \tag{4.4}$$

Then, the regularized stochastic action is defined through the average

$$A_{\Delta t}(t_0,t_1;q,\Gamma) = \sum_{t=t_0}^{t_1-\Delta t} \int_t^{t+\Delta t} E(\mathscr{L}(\hat{q}(t'),\hat{\dot{q}}(t');t'))dt'. \tag{4.5}$$

Notice the explicit Γ dependence.

The stochastic action will be reached in the limit $\Delta t \to 0$, $K \to \infty$.

Due to linearity, there will be three contributions to the limit, $A^{(S)}$, $A^{(V)}$, and $A^{(T)}$, coming from the scalar potential V, the vector potential A, and the tensorial kinetic metric g in (1) (tensor potential), respectively.

The scalar potential does not give any trouble. In fact, no regularization would be even necessary in this case. The result is

$$\begin{aligned} A^{(S)} &= -\int_{t_0}^{t_1} E(V(q(t),t))dt \\ &= -\int_{t_0}^{t_1} dt \int V(x,t)\rho(x,t)d\mu(x). \end{aligned} \tag{4.6}$$

Let us define the trivial Lagrangian fields

$$\mathscr{L}^{(S)}_{(+)} = \mathscr{L}^{(S)}_{(-)} = \mathscr{L}^{(S)} = -V(x,t), \tag{4.7}$$

so that

$$A^{(S)} = \int_{t_0}^{t_1} E(\mathscr{L}^{(S)}_{\#}(q(t),t))dt. \tag{4.8}$$

The evaluation of the vector potential contribution is also very easy (see for example Ref. 19 for a very similar stochastic line integration theory). First of all, one can easily prove the following.

Theorem 4 (conditional expectations of vector potential contributions): For any $t \in [t_0,t_1]$ and sufficiently small $\Delta t > 0$ we have

$$\begin{aligned} &\int_t^{t+\Delta t} E(A_i(\hat{q}(t'),t')\hat{\dot{q}}(t')\,|\,q(t)=x)dt' \\ &\quad = \mathscr{L}^{(V)}_{(+)}(x,t)\Delta t + O(\Delta t^2), \end{aligned} \tag{4.9}$$

with

$$\mathscr{L}^{(V)}_{(+)}(x,t) = (A_i v^i_{(+)} + \nu n^{ij}\nabla_i A_j)(x,t). \tag{4.10}$$

Proof: We give a detailed proof of this very simple result, because the much more complex proof of Theorem 5 in the next section is based essentially on extensions of the same technique.

It is convenient to introduce the affine coordinate s, as in (4.4), so that

$$\begin{aligned} &\hat{q}(t') = y(s), \quad t' = t + s\Delta t, \\ &\hat{\dot{q}}(t') = \dot{y}(s)/\Delta t, \quad dt' = \Delta t\, ds. \end{aligned} \tag{4.11}$$

Therefore, the integral in (4.9) reduces to

$$\int_0^1 E_{x,t}\hat{A}_i(s)\,\dot{y}^i(s)ds, \tag{4.12}$$

where

$$\hat{A}_i(s) \equiv A_i(y(s),t+s\Delta t). \tag{4.13}$$

In order to get (4.9), (4.10), we must evaluate (4.12) correctly up to order Δt. Recalling (2.24), (2.25), we have

$$\begin{aligned} \dot{y}(s) &= \dot{y}(0) + s\ddot{y}(0) + O(\Delta t^{3/2}), \\ \hat{A}_i(s) &= A_i(x,t) + \partial_j A_i(y^j(s) - y^j(0)) + O(\Delta t) \\ &= A_i(x,t) + s\partial_j A_i\,\dot{y}^j(0) + O(\Delta t). \end{aligned} \tag{4.14}$$

Therefore

$$\begin{aligned} (\hat{A}_i\,\dot{y}^i)(s) &= A_i(x,t)(\dot{y}^i(0) + s\ddot{y}^i(0)) \\ &\quad + s\partial_j A_i\,\dot{y}^i(0)\,\dot{y}^j(0) + O(\Delta t^{3/2}). \end{aligned} \tag{4.15}$$

Taking the conditional expectation $E_{x,t}$, performing the integration on s and taking into account (2.15), (3.14), we find immediately (4.9), (4.10). Q.E.D.

By exploiting the results of this theorem and taking into account the definition (4.5) for the action and the properties of the conditional expectation, we immediately find, in the limit $\Delta t \to 0$, the following contribution to the action:

$$A^{(V)} = \int_{t_0}^{t_1} E(\mathscr{L}^{(V)}_{(+)}(q(t),t))dt. \tag{4.16}$$

One can also introduce

$$\begin{aligned} &\mathscr{L}^{(V)}_{(-)}(x,t) = (A_i v^i_{(-)} - \nu\eta^{ij}\nabla_i A_j)(x,t), \\ &\mathscr{L}^{(V)}(x,t) = (A_i v^i)(x,t), \\ &\mathscr{L}^{(V)} = \tfrac{1}{2}(\mathscr{L}^{(V)}_{(+)} + \mathscr{L}^{(V)}_{(-)}), \end{aligned} \tag{4.17}$$

and notice that $A^{(V)}$ can also be calculated by substitution of $\mathscr{L}^{(V)}_{(+)}$ in (4.16) with $\mathscr{L}^{(V)}_{(-)}$ or $\mathscr{L}^{(V)}$, as shown by a simple integration by parts on the space variables.

It is also very important to remark that $\mathscr{L}^{(V)}_{\#}$ are invariant under geodesic gauge transformations given by (2.30). For example, the variation of $v_{(+)}$, given by (2.29), is exactly absorbed by the variation of ∇_i, following from (2.34).

V. THE STOCHASTIC KINETIC ACTION

The evaluation of the kinetic tensor contribution to the action is based on the following theorem, one of the main results of this paper.

Theorem 5 (conditional expectations of kinetic tensor contributions): For any $t \in [t_0,t_1]$ and small $\Delta t > 0$ we have

$$\int_t^{t+\Delta t} E(\tfrac{1}{2}mg_{ij}(\hat{q}(t'),t')\hat{\dot{q}}^i(t')\hat{\dot{q}}^j(t')|q(t)=x)dt'$$
$$= m\nu\big(g_{ij}\eta^{ij}+\tfrac{1}{2}D_{(+)}(g_{ij}\eta^{ij})\Delta t\big)(x,t)$$
$$+\mathscr{L}^T_{(+)}(x,t)\Delta t+O(\Delta t^2)\,, \tag{5.1}$$
$$\mathscr{L}^T_{(+)}(x,t)=\tfrac{1}{2}mg_{ij}v^i_{(+)}v^j_{(+)}$$
$$+m\nu\eta^{ik}\nabla_k(g_{ij}v^j_{(+)})-V_s\,,$$
$$\begin{aligned}-V_s&=\tfrac{1}{3}m\nu^2g_{ij}R^i{}_{klm}\eta^{kl}\eta^{jm}\\&+\tfrac{1}{2}m\nu^2\nabla_k g_{ij}(2\eta^{il}\nabla_l\eta^{kj}-\eta^{kl}\nabla_l\eta^{ij})\\&+\tfrac{1}{6}m\nu^2\nabla_k\nabla_l\,g_{ij}(4\eta^{ki}\eta^{lj}-\eta^{kl}\eta^{ij})\,.\end{aligned}\tag{5.2}$$

Proof: The proof is based on a careful evaluation of the contributions to all necessary orders.

In particular, terms of order $(\Delta t)^{3/2}$ in the displacement Δq enter in an essential way. Firstly we perform the substitution (4.11) in the integrand and get

$$(\Delta t)^{-1}\int_t^{t+\Delta t}E_{x,t}(\tfrac{1}{2}mg_{ij}(y(s),t+s\Delta t)\,\dot{y}^i(s)\dot{y}^j(s))ds\,. \tag{5.3}$$

Define

$$A(s)\equiv g_{ij}(y(s),t+s\Delta t)\,\dot{y}^i(s)\dot{y}^j(s)=O(\Delta t)\,. \tag{5.4}$$

We see that $A(s)$ must be calculated correctly up to order Δt^2 in order to have (5.1) correct up to Δt. Notice

$$\dot{A}(s)=\partial^2_{kl}\,g_{ij}\,\dot{y}^k\dot{y}^i\dot{y}^j+2g_{ij}\,\ddot{y}^i\dot{y}^j+\partial_t g_{ij}\,\dot{y}^i\dot{y}^j\Delta t$$
$$=O(\Delta t^{3/2})\,, \tag{5.5}$$
$$\ddot{A}(s)=\partial^2_{kl}\,g_{ij}\,\dot{y}^k\dot{y}^l\dot{y}^i\dot{y}^j+3[\partial_k g_{ij}]\,\ddot{y}^k\dot{y}^i\dot{y}^j$$
$$+2\partial_k g_{ij}\,\dot{y}^k\ddot{y}^i\dot{y}^j+2g_{ij}\,y^{\mathrm{III}i}\dot{y}^j+2g_{ij}\,\ddot{y}^i\ddot{y}^j+O(\Delta t^{5/2})$$
$$=O(\Delta t^2)\,, \tag{5.6}$$
$$A^{\mathrm{III}}(s)=O(\Delta t^{5/2})\,. \tag{5.7}$$

In (5.6) we have denoted by $[\partial_k g_{ij}]$ the symmetric part of the argument. Therefore, we can write

$$A(s)=A(0)+s\dot{A}(0)+\tfrac{1}{2}s^2\ddot{A}(0)+O(\Delta t^{5/2})\,. \tag{5.8}$$

It is convenient to work in normal coordinates for Γ at the initial point x, so that $\Gamma(x)=0$. Then, from (2.25) we have

$$\ddot{y}(0)=0,\quad y^{\mathrm{III}\,i}(0)=-\partial_j\Gamma^i_{kl}\Delta q^j\Delta q^k\Delta q^l\,. \tag{5.9}$$

In (5.8) we must substitute

$$\begin{aligned}A(0)&=g_{ij}(x,t)\Delta q^i\Delta q^j\,,\\\dot{A}(0)&=(\partial_k g_{ij})(x,t)\Delta q^i\Delta q^j\Delta q^k\\&\quad+\Delta t(\partial_t g_{ij})(x,t)\Delta q^i\Delta q^j\,,\\\ddot{A}(0)&=[(\partial^2_{kl}g_{ij})(x,t)-2g_{si}\partial_j\Gamma^s_{kl}]\\&\quad\times\Delta q^i\Delta q^j\Delta q^k\Delta q^l+O(\Delta t^{3/2})\,.\end{aligned}\tag{5.10}$$

Collecting all terms and performing the s integration in (5.3), we find that the integral in (5.1) is expressed as

$$\begin{aligned}\Delta t^{-1}[&(g_{ij}+\tfrac{1}{2}\Delta t\partial_t g_{ij})E_{x,t}\Delta q^i\Delta q^j\\&+\tfrac{1}{2}\nabla_k g_{ij}E_{x,t}\Delta q^i\Delta q^j\Delta q^k\\&+\tfrac{1}{6}\nabla_k\nabla_l g_{ij}E_{x,t}\Delta q^i\Delta q^j\Delta q^k\Delta q^l]\,.\end{aligned}\tag{5.11}$$

Finally, we can exploit the expressions (3.14)–(3.16) and, after a simple rearrangement of terms, we derive (5.1), (5.2). This ends the proof of Theorem 5.

At first sight, it could seem strange that the integral in (5.1), performed over the time interval $[t,\ t+\Delta t]$, starts with a term $O(\Delta t^0)$ and not $O(\Delta t)$. Obviously, this is due to the highly singular nature of the time derivatives $\hat{\dot{q}}$ in the limit $\Delta t\to 0$.

Starting from Theorem 5 we see immediately that the kinetic part of the regularized action, according to the definition (4.5), can be written in the form

$$\begin{aligned}A^{(T)}_{\Delta t}&=\frac{m\nu}{\Delta t}\int_{t_0}^{t_1}E((g\eta)(q(t),t))dt\\&\quad+\int_{t_0}^{t_1}E(\mathscr{L}^{(T)}_{(+)}(q(t),t))dt+O(\Delta t)\\&\equiv\frac{m\nu}{\Delta t}\int_{t_0}^{t_1}dt\int(g\eta)(x,t)\,\rho(x,t)d\mu(x)\\&\quad+\int_{t_0}^{t_1}dt\int\mathscr{L}^{(T)}_{(+)}(x,t)\rho(x,t)d\mu(x)+O(\Delta t),\end{aligned}\tag{5.12}$$

where

$$(g\eta)(x,t)=g_{ij}(x,t)\eta^{ij}(x,t)\,. \tag{5.13}$$

The proof of (5.12) is obvious for what is related to $\mathscr{L}^{(T)}_{(+)}$. For the first term, let us notice that for a generic smooth scalar field ϕ we have

$$\begin{aligned}\int_{t_0}^{t_1}&E(\phi(q(t),t))dt\\&=\sum_t E(\phi(q(t),t))\Delta t\\&\quad+\sum_t E((D_{(+)}\phi)(q(t),t))\frac{\Delta t^2}{2}+O(\Delta t^2)\,,\end{aligned}\tag{5.14}$$

where the dummy parameter t takes continuous values in $[t_0,t_1]$ on the lhs, and discrete values, as explained in (4.3), on the rhs. In order to prove (5.14), let us write the lhs in the form

$$\sum_t\int_t^{t+\Delta t}E(\phi(q(t'),t'))dt'\,. \tag{5.15}$$

Then we can exploit (2.32) to write

$$E(\phi(q(t'),t'))=E(\phi(q(t),t))+E((D_{(+)}\phi)(q(t),t))$$
$$\times(t-t')+O(\Delta t^2)\,. \tag{5.16}$$

By integrating on dt' and summing on the discrete t, we immediately recover (5.14).

VI. THE STOCHASTIC ACTION AND GEODESIC GAUGE TRANSFORMATIONS

Now we can collect all results, obtained so far, in the following.

Theorem 6 (the regularized stochastic action): Let us define, exploiting (4.7), (4.14), (4.21), and (5.2),

$$\begin{aligned}\mathscr{L}_{(\pm)}(x,t)&=\mathscr{L}^{(S)}_{(\pm)}+\mathscr{L}^{(V)}_{(\pm)}+\mathscr{L}^{(T)}_{(\pm)}\,,\\\mathscr{L}(x,t)&=\mathscr{L}^{(S)}+\mathscr{L}^{(V)}+\mathscr{L}^{(T)}\,,\end{aligned}\tag{6.1}$$

where

$$\mathscr{L}^{(T)}_{(-)}(x,t)=\tfrac{1}{2}mg_{ij}v^i_{(-)}v^j_{(-)} - m\nu\eta^{ik}\nabla_k(g_{ij}v^j_{(-)})-V_s\,, \qquad (6.2)$$

$$\mathscr{L}^T(x,t)=\tfrac{1}{2}mg_{ij}v^i_{(+)}v^j_{(-)}-V_s\,.$$

Then the regularized stochastic action, defined through the geodesic interpolation in (4.5), is given by

$$A_{\Delta t}(t_0,t_1;q;\Gamma)=\frac{m\nu}{\Delta t}\int_{t_0}^{t_1}E((g\eta)(q(t),t))dt + \int_{t_0}^{t_1}E(\mathscr{L}_{\#}(q(t),t))dt+O(\Delta t)\,. \qquad (6.3)$$

The stochastic action should be recovered in the limit $\Delta t\to 0$. However we see that a singularity develops in the first term. In the spirit of renormalization theory, we can substitute a finite undetermined constant $1/\tau$ to the infinite constant given by $\lim_{\Delta t\to 0}(\Delta t)^{-1}$. Therefore we assume, as effective action for our dynamical system on the manifold, $A(t_0,t_1;q,\Gamma)$ given by (6.3), with $1/\tau$ in place of $1/\Delta t$ and the term $O(\Delta t)$ suppressed as a result of the limit $\Delta t\to 0$.

In any case, following a basic remark of Nelson,[1] we notice that the $1/\tau$ term does not play any role in the variational principle where the controlling covariance η is kept fixed and only $b_{(+)}$ or $v_{(+)}$ are subject to variations. Therefore we have the following.

Definition: The renormalized stochastic action is given by

$$A(t_0,t_1;q;\Gamma)=\int_{t_0}^{t_1}E(\mathscr{L}_{\#}(q(t),t))dt\,. \qquad (6.4)$$

It is very interesting to investigate how the expression of the action, so obtained, does really depend on the regularizing connection Γ. We have already noticed that the scalar and vector parts are independent of Γ, as shown in Sec. IV. However, we find that this is no longer true for the kinetic part. In fact, through a long but straightforward calculation, one can obtain the following.

Theorem 7 (geodesic gauge transformation of the stochastic kinetic Lagrangian)**:** If the regularizing connection field changes according to $\Gamma\to\Gamma'$, then the effective potential V_S, appearing in (5.2), is modified as follows:

$$V_S\to V'_S=V_S+\tfrac{1}{3}m\nu^2\tfrac{1}{2}(\tilde\Gamma+\tilde\Gamma')_{k,ij}(\eta^{ij}\eta^{lm} + \eta^{il}\eta^{jm}+\eta^{im}\eta^{jl})(\Gamma'-\Gamma)^k_{lm}\,, \qquad (6.5)$$

where

$$\tilde\Gamma_{k,ij}=\tfrac{1}{2}(\nabla_j g_{ik}+\nabla_i g_{jk}-\nabla_k g_{ij})\,, \qquad (6.6)$$

and $\tilde\Gamma'$ has the same expression, but the covariant derivative ∇ with respect to Γ must be replaced by ∇' with respect to Γ'.

For the proof of the theorem we must take into account all changes in (5.2) arising from $\Gamma\to\Gamma'$. In particular, the change in $v_{(+)}$ is handled through (2.29), while for ∇ and R we exploit (2.34) and (3.17). By collecting all terms coming from $\Gamma\to\Gamma'$, we arrive at (6.5), (6.6).

If we assume that the geodesic interpolation gives the correct definition for the stochastic action, then Theorem 7 has far reaching consequences. In fact, if we are willing to insist on global geodesic gauge invariance of the total action, we must introduce a change in the potential so to balance the effects of (6.5):

$$V\to V' \text{ such that } V_S+V=V'_S+V'\,. \qquad (6.7)$$

The physical interpretation of this change is very simple. It only tells us that the effective potential in the stochastic theory is determined up to terms involving the square of the diffusion constant, starting from the potential of the classical theory.

On the other hand, the stochastic theory can be exploited as a way of providing a kind of stochastic quantization for classical dynamical systems on manifolds, according to the general strategy expounded in Refs. 1, 2, and 13. Then we must conclude that the classical theory determines the quantum potential up to second order terms in the Planck constant (recall that $\nu=\hbar/2m$ in Nelson's stochastic quantization).[1,17]

It is also very interesting to discuss some particular cases where the expression of the stochastic action simplifies. For example, assume that g is time independent, and choose for Γ the Riemann connection canonically associated to g,

$$\Gamma^i_{jk}=\tfrac{1}{2}g^{il}(\partial_j g_{lk}+\partial_k g_{lj}-\partial_l g_{jk}),\quad g^{ij}g_{jk}=\delta^k_i\,. \qquad (6.8)$$

In this case, the last two terms in the definition (5.2) of V_S disappear, because $\nabla_g=0$, while the first term involves the curvature tensor R_{iklm} associated to g, saturated with η. If, in addition, the controlling η is assumed in the form $\eta^{ij}=g^{ij}$, then we see that V_S reduces to the well-known form of Pauli–De Witt,[4,5] involving the curvature invariant R.

Therefore, in general, the expression (5.2) for V_S gives a very detailed understanding of the origin of the various contributions related to the, conceptually different, kinetic term g, control correlation η, and the curvature tensor for the auxiliary connection Γ.

VII. CONCLUSIONS AND OUTLOOK

We have introduced a general stategy, well known in the theory of stochastic processes, and based on piecewise geodesic approximations, in order to extend the definition of the action of dynamical systems on curved manifolds (see also Refs. 1 and 20).

The resulting average action, instead of being a functional of trial smooth trajectories, becomes a functional of controlled stochastic Markov processes on the manifold.

Infinite quantities have been substituted by finite undetermined constants, according to the spirit of renormalization theory.

The action functional, so obtained, can be considered as the starting point for stochastic variational principles, as introduced in Refs. 1, 2, and 13.

A very important property of the action functional is given by the noninvariance of its kinetic part, with respect to changes of the regularizing connection field. This is a new additional example of the strict interplay between stochastic and differential geometric aspects of dynamics on curved manifolds. Previous examples where given about the need of a geodesic correction to stochastic parallel displacement of vectors on a manifold, as explained, for example, in Refs. 1 and 10.

The general definition of the stochastic action, obtained in this paper, can be exploited in the frame of stochastic variational principles along two main lines. The first line deals with the extension of the results contained in Ref. 13, where a purely heuristic definition of the action was assumed. The second line is related to the extension to nonflat manifolds of the stochastic variational principles of Lagrangian type, introduced in Refs. 21 and 22. We plan to report on these topics in future publications.

ACKNOWLEDGMENT

This work was partially supported by Ministero della Pubblica Istruzione.

[1]E. Nelson, *Quantum Fluctuations* (Princeton U.P., Princeton, NJ, 1985), and previous references quoted there.
[2]F. Guerra and L. M. Morato, Phys. Rev. D **27**, 1774 (1983).
[3]B. S. DeWitt, Rev. Mod. Phys. **29**, 337 (1957).
[4]W. Pauli, *Pauli Lectures on Physics* (MIT Press, Cambridge, MA, 1973).
[5]C. DeWitt-Morette, K. D. Elworthy, B. L. Nelson, and G. S. Sammelman, Ann. Inst. Henri Poincaré **32**, 327 (1980).
[6]R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948).
[7]K. Itô, "The Brownian motion and tensor fields on Riemannian manifold," Proc. Int. Congress Math. (Stockholm), p. 536 (1963).
[8]E. Dynkin, Soviet Math. Dokl. **9**, 532 (1968).
[9]N. Ikeda and S. Watanabe, *Stochastic Differential Equations and Diffusion Processes* (North-Holland, Amsterdam, 1981).
[10]D. Dohrn and F. Guerra, Lett. Nuovo Cimento **22**, 121 (1978).
[11]E. Aldrovandi, D. Dohrn, and F. Guerra, "Stochastic action of dynamical systems on curved manifolds. The isokinetic developing map on trajectories," *Proceedings of the 1988 Ascona Conference Stochastic Processes—Geometry and Physics*, edited by S. Albeverio and D. Merlini (to be published).
[12]M. Spivak, *A Comprehensive Introduction to Differential Geometry* (Publish or Perish, Boston, 1979).
[13]D. Dohrn and F. Guerra, Phys. Rev. D **31**, 2531 (1985).
[14]F. Guerra, "Carlen processes: a new class of diffusions with singular drifts," in *Quantum Probability and Applications II*, edited by L. Accardi and W. von Waldenfels, *Lecture Notes in Mathematics*, Vol. 1136 (Springer–Verlag, Berlin, 1985).
[15]E. Carlen, Commun. Math. Phys. **94**, 293 (1984).
[16]T. G. Dankel, Jr., Arch. Rational Mech. Anal. **37**, 192 (1970).
[17]F. Guerra, Phys. Rep. **77**, 121 (1981).
[18]L. M. Morato, J. Math. Phys. **23**, 1020 (1982).
[19]N. Ikeda and S. Manabe, Publ. RIMS, Kyoto Univ. **15**, 827 (1979).
[20]J. D. Lafferty, Trans. Am. Math. Soc. **633**, 699 (1988).
[21]L. M. Morato, Phys. Rev. D **31**, 1982 (1985).
[22]F. Guerra, Ann. Inst. Henri Poincaré **49**, 315 (1988).

Space-times admitting special affine conformal vectors

A. A. Coley
Department of Mathematics, Statistics and Computing Science, Dalhousie University, Halifax, Nova Scotia B3H 3J5, Canada

B. O. J. Tupper
Department of Mathematics and Statistics, University of New Brunswick, Fredericton, New Brunswick E3B 5A3, Canada

(Received 21 August 1989; accepted for publication 18 October 1989)

Space-times admitting a special affine conformal vector (SACV) are shown to be precisely the space-times that admit a special conformal Killing vector. All possible SACV space-times are listed together with the corresponding SACV's and covariantly constant tensors.

I. INTRODUCTION

In general relativity the existence of certain symmetries in the space-time manifold is often assumed in the pursuit of exact solutions of Einstein's field equations, and such symmetries and their corresponding vector fields have been studied by many authors (see Ref. 1 for references). The Lie derivative of the metric tensor, g_{ab}, in the direction of the vector field ξ^a can be written in the general form

$$\mathop{\mathscr{L}}_{\xi} g_{ab} = 2\psi g_{ab} + K_{ab}, \tag{1}$$

where ψ is a scalar function of the coordinates and K_{ab} is a symmetric tensor. Among the various special symmetries generated by ξ^a are:

(i) Killing vector (KV): $\psi = 0$, $K_{ab} = 0$,
(ii) Homothetic vector (HV): $\psi = \text{const} \neq 0$, $K_{ab} = 0$,
(iii) Proper conformal Killing vector (CKV): $\psi_{,a} \neq 0$, $K_{ab} = 0$,
(iv) Special conformal Killing vector (SCKV): $\psi_{;ab} = 0$, $\psi_{,a} \neq 0$, $K_{ab} = 0$,
(v) Affine vector (AV): $\psi = 0$, $K_{ab} \neq 0$, $K_{ab;c} = 0$,
(vi) Proper affine conformal vector (ACV): $\psi_{,a} \neq 0$, $K_{ab} \neq 0$, $K_{ab;c} = 0$,
(vii) Special affine conformal vector (SACV): $\psi_{;ab} = 0$, $\psi_{,a} \neq 0$, $K_{ab} \neq 0$, $K_{ab;c} = 0$,

where, in (v), (vi), and (vii), K_{ab} is not proportional to g_{ab}. An ACV or a CKV [which can be regarded as a special case of (vi) in which K_{ab} is proportional to g_{ab}] generates a conformal collineation characterized by

$$\mathop{\mathscr{L}}_{\xi} \begin{Bmatrix} a \\ bc \end{Bmatrix} = \delta^a_b \psi_{,c} + \delta^a_c \psi_{,b} - g_{bc}\psi^{a}. \tag{2}$$

An SACV or an SCKV generates a Ricci collineation characterized by

$$\mathop{\mathscr{L}}_{\xi} R_{ab} = 0. \tag{3}$$

Recently, we have investigated SCKV's[1] and have found all space-times admitting SCKV's. The existence of the covariantly constant vector $\psi_{,a}$ and the SCKV equations given by (1) and (iv) imply that there are very few SCKV space-times, none of which can represent a perfect fluid, a non-null electromagnetic field, or a vacuum space-time other than a *pp*-wave space-time. However, such space-times can represent viscous fluids and a set of anisotropic fluids satisfying particularly restrictive equations of state. A thorough study of AV's has been made by Hall and da Costa,[2] while ACV's have been discussed by Duggal[3], and SACV's have been studied by Duggal and Sharma[4] and Mason and Maartens.[5]

In this paper we discuss SACV's and show that the only space-times admitting SACV's are precisely those which admit SCKV's. We display the form of the SACV for each space-time and give the associated covariantly constant tensor K_{ab}.

II. SACV SPACE-TIMES

We first note a result due to Hall and da Costa,[2,6] namely if a simply connected space-time admits a global, nowhere zero, covariantly constant, second-order symmetric tensor, K_{ab}, which is not a constant multiple of the metric tensor, then one of the following three possibilities occur:

(a) There exists locally a timelike or spacelike nowhere zero covariantly constant vector field $\eta_a \equiv \eta_{,a}$ such that $K_{ab} = \eta_{,a}\eta_{,b}$ and the space-time decomposes into a $1+3$ space-time, in the notation of Ref. 2.

(b) There exists locally a null nowhere zero covariantly constant vector field $\eta_a \equiv \eta_{,a}$ such that $K_{ab} = \eta_{,a}\eta_{,b}$ and the space-time is the generalized *pp*-wave space-time[1,7] which, in general, is not decomposable.

(c) The space-time is locally decomposable into a $2+2$ space-time and no covariantly constant vector exists unless the space-time decomposes further.

We note that, if a $2+2$ space-time does admit a covariantly constant vector, then it must locally decompose further into a $1+1+2$ space-time. This follows immediately from holonomy considerations[6] or is shown easily by writing the general $2+2$ space-time metric in the form

$$ds^2 = e^{2\mu}(-dt^2 + dx^2) + e^{2\nu}(dy^2 + dz^2), \tag{4}$$

where $\mu = \mu(t,x)$ and $\nu = \nu(y,z)$, and solving the equations, $\eta_{a;b} = 0$. It is found that $\eta_a = 0$ unless one (or both) of the following conditions hold:

$$\mu_{tt} - \mu_{xx} = 0, \quad \nu_{yy} + \nu_{zz} = 0. \tag{5}$$

The first of these implies that the first two-space is flat (i.e., decomposes into a $1+1$ space), while the second condition implies that the second two-space is flat. Since the SACV

 0022-2488/90/030649-04$03.00

space-times admit the covariantly constant vector $\psi_{,a}$, it follows that we need consider only cases (a) and (b), the $1+1+2$ decomposition being treated as a special case of the $1+3$ decomposition.

In cases (a) and (b) Eq. (1) becomes

$$\xi_{a;b} + \xi_{b;a} = 2\psi g_{ab} + \eta_{,a}\eta_{,b}, \tag{6}$$

which can be rewritten in the form

$$(\xi_a - \tfrac{1}{2}\eta\eta_{,a})_{;b} + (\xi_b - \tfrac{1}{2}\eta\eta_{,b})_{;a} = 2\psi g_{ab}, \tag{7}$$

so that $\zeta_a \equiv \xi_a - \frac{1}{2}\eta\eta_{,a}$ is a CKV (which will be an SCKV if $\psi_{;ab} = 0$ or an HV if $\psi_{,a} = 0$), and $\tau_a \equiv \xi_a - \zeta_a = \frac{1}{2}\eta\eta_{,a}$ is an AV. Thus, in cases (a) and (b), the ACV is necessarily the sum of a CKV and an AV, so that space-times admitting ACV's must also admit a CKV and an AV.

Applying this result to SACV's, we see that space-times admitting SACV's are precisely those which admit SCKV's and which were found in Ref. 1. On the other hand, given an SCKV space-time satisfying

$$\zeta_{a;b} + \zeta_{b;a} = 2\psi g_{ab} \tag{8}$$

with $\psi_{;ab} = 0$, we see that the vector $\chi_a = \frac{1}{2}\psi\psi_{,a}$ is an AV since

$$\mathop{\mathscr{L}}_{\chi} g_{ab} \equiv \chi_{a;b} + \chi_{b;a} = \psi_{,a}\psi_{,b} \tag{9}$$

so that $\xi_a = \zeta_a + \chi_a$ satisfies

$$\mathop{\mathscr{L}}_{\xi} g_{ab} = 2\psi g_{ab} + \psi_{,a}\psi_{,b} \tag{10}$$

and ξ_a is an SACV. Thus we have the following theorem.[8]

Theorem: *A simply connected space-time will admit an SACV if and only if it admits an SCKV.* The theorem asserts the complete equivalence between SACV and SCKV space-times.

From Ref. 1, when $\psi_{,a}$ is timelike we can choose local coordinates such that $\psi_{,a} = (-1,0,0,0)$ and the space-time metric is

$$ds^2 = -dt^2 + dx^2 + x^2(dy^2 + f^2(y,z)dz^2). \tag{11}$$

The SCKV is $\zeta^a = (-\frac{1}{2}t^2 - \frac{1}{2}x^2, -tx, 0, 0)$ and, since $\psi_{,a}$ is, up to constant scalings, the only covariantly constant vector, we have $K_{ab} = \psi_{,a}\psi_{,b}$ so that $K_{00} = 1$ is the only nonzero component. It follows that the AV, τ^a, and the SACV, ξ^a, have components

$$\tau^a = (\tfrac{1}{2}t,0,0,0), \tag{12}$$

$$\xi^a = (-\tfrac{1}{2}t^2 + \tfrac{1}{2}t - \tfrac{1}{2}x^2, -tx,0,0). \tag{13}$$

When $\psi_{,a}$ is spacelike, the space-time metric takes one of the following possible forms:

$$ds^2 = dx^2 - dt^2 + t^2(dy^2 + g^2(y,z)dz^2), \tag{14}$$

$$ds^2 = dx^2 + dy^2 + y^2(-dt^2 + h^2(t,z)dz^2). \tag{15}$$

In each case $\psi_{,a} = (0,1,0,0)$, which is the only covariantly constant vector, and $K_{ab} = \psi_{,a}\psi_{,b}$ has $K_{11} = 1$ as its only nonzero component, so that the AV is

$$\tau^a = (0,\tfrac{1}{2}x,0,0). \tag{16}$$

For the metric (14) the SCKV and SACV are, respectively,

$$\zeta^a = (xt,\tfrac{1}{2}x^2 + \tfrac{1}{2}t^2,0,0), \tag{17}$$

$$\xi^a = (xt,\tfrac{1}{2}x^2 + \tfrac{1}{2}x + \tfrac{1}{2}t^2,0,0), \tag{18}$$

while for the metric (15), the corresponding quantities are

$$\zeta^a = (0,\tfrac{1}{2}x^2 - \tfrac{1}{2}y^2,xy,0), \tag{19}$$

$$\xi^a = (0,\tfrac{1}{2}x^2 + \tfrac{1}{2}x - \tfrac{1}{2}y^2,xy,0). \tag{20}$$

Note that the space-time metric (15) does not satisfy the dominant energy condition and so has no reasonable physical interpretation.

When $\psi_{,a}$ is null, the space-time must be the generalized *pp* wave space-time with metric

$$ds^2 = P^{-2}(dx^2 + dy^2) - 2du(dv - mdx + Hdu), \tag{21}$$

where H, m, and P are arbitrary functions of u, x, and y only. Not all space-times of the form (21) admit an SCKV, but in the case when the Ricci scalar, R, is not zero, if an SCKV exists it is of the form

$$\begin{aligned}\zeta^a = [&-(u^2+\alpha u+\beta), \alpha v - D(u,x,y)\\ &+ (2H + m^2P^2)(u^2+\alpha u+\beta)\\ &+ mP^2B(u,x,y), mP^2(u^2+\alpha u+\beta)\\ &+ P^2B(u,x,y), P^2C(u,x,y)],\end{aligned} \tag{22}$$

where α, β are arbitrary constants and B, C, and D satisfy a set of six first-order differential equations [Eqs. (4.23) of Ref. 1] which serves to delineate those members of the set of space-times (21) which admit an SCKV.

When $R = 0$, the space-time metric can be written in the form

$$ds^2 = dx^2 + dy^2 - 2du(dv + H\,du), \tag{23}$$

and such a space-time will admit the SCKV

$$\begin{aligned}\zeta^a = [&-(u^2+\alpha u+\beta), \alpha v - \tfrac{1}{2}x^2 - \tfrac{1}{2}y^2 + J_u x\\ &+ K_u y + L(u),\\ &-ux + \gamma y + J(u), -uy - \gamma x + K(u)],\end{aligned} \tag{24}$$

provided that the metric function H satisfies the equation

$$\begin{aligned}&H_u(u^2+\alpha u+\beta) + H_x(ux - \gamma y - J)\\ &+ H_y(uy + \gamma x - K)\\ &+ 2H(u+\alpha) - J_{uu}x - K_{uu}y + L_u = 0,\end{aligned} \tag{25}$$

where α, β, and γ are arbitrary constants and J, K, and L are arbitrary functions of u only.

In each of the above cases, the null vector $\psi_{,a} = (-1,0,0,0)$ is the only covariantly constant vector (see the Appendix), and $K_{ab} = \psi_{,a}\psi_{,b}$ has only $K_{00} = 1$ as a nonzero component. The AV and SACV are, respectively,

$$\tau^a = (0, -\tfrac{1}{2}u,0,0), \tag{26}$$

$$\xi^a = \zeta^a + \tau^a, \tag{27}$$

with ζ^a given by either (22) or (24).

The equivalence of SACV and SCKV space-times and the form of K_{ab} can be demonstrated also by using a coordinate dependent approach similar to that used in Ref. 1 (see MacLean[9]). In addition, the various results obtained in Ref. 1 are applicable in the SACV case. In particular (Ref. 1, Theorem 7), the energy-momentum tensors for the space-times (11) and (14) are of Segré type {(1,1) (1 1)}, while the space-times (21) and (23) are either of this type or of type {2(1 1)}, where, in either case, the bracketed pair of space-like vectors have zero eigenvalue (see also Hall[10]). This implies that the existence of an SCKV or, equivalently, an SACV, eliminates all perfect fluid space-times and all non-null electrovac fields. The only vacuum space-times are the *pp*-wave solutions, and the only null electrovac fields are the conformally flat *pp*-wave type. On the other hand, the space-times can be interpreted as viscous fluid solutions or, if $R \neq 0$, as anisotropic fluid solutions subject to the restriction

$$\mu = -p_{\parallel} = \tfrac{1}{2}R, \quad p_{\perp} = 0, \tag{28}$$

where μ is the energy density, and $p_{\parallel}$ and $p_{\perp}$ denote the parallel and perpendicular pressures, respectively. Equation (28) indicates that the various subcases considered by Mason and Maartens[5] are not, in fact, possible.

III. CONCLUSION

We have found all space-times that admit an SACV [i.e., space-times (11), (14), (15), and the appropriate members of (21) and (23)]; these are identical to the set of space-times admitting an SCKV and, consequently, have only a limited number of possible physical interpretations.

Duggal and Sharma[4] have considered space-times admitting SACV's (which they refer to as "special conformal collineations"), particularly those representing anisotropic fluids, subject to the condition

$$K_{ab} = \gamma R_{ab}, \tag{29}$$

where γ is a scalar function. This condition implies that the space-time is Ricci recurrent. However, on calculating the Ricci tensor in the cases of the space-times (11), (14), (15), and (21), it is easy to see that $K_{ab} = \psi_{,a}\psi_{,b}$ is never proportional to R_{ab}. In fact, (29) cannot be satisfied even if we take K_{ab} to be of the form $K_{ab} = \psi_{,a}\psi_{,b} + Cg_{ab}$, for some constant C. In the case of the metric (23), the only nonzero component of K_{ab} is $K_{00} = 1$ and the only nonzero component of R_{ab} is R_{00}, so Eq. (29) does hold. Since $R = 0$, the metric (23) admits no anisotropic fluid solutions—only viscous fluid, null electrovac, and pure radiation solutions—and the energy-momentum tensor is of the form $T_{ab} = \lambda\psi_{,a}\psi_{,b}$. Thus many of the results concerning anisotropic and isotropic fluids presented in Ref. 4 are illusory since there exist virtually no SACV space-times satisfying the physical interpretations and mathematical conditions considered in that article.

ACKNOWLEDGMENTS

We are grateful to Graham Hall for preprints of his work and for discussions and communications which enabled us to better understand the geometrical aspects of the problem.

This research was supported in part by the Natural Sciences and Engineering Research Council of Canada through operating grants to each author.

APPENDIX

The statement that each of the SACV space-times admits only one covariantly constant vector, namely $\psi_{,a}$, is easily proved in the cases of the space-times (11), (14), (15), and (23) by solving the equations $m_{a;b} = 0$ and showing that $m_a = \psi_{,a}$ is the only solution. However, in the case of the space-time (21) the proof is rather less simple; here we present an outline of the proof.

We consider the metric (21) and specify that $R \neq 0$ [otherwise we have the metric (23)]. The covariantly constant vector $\psi_{,a} \equiv k_a$ is of the form $k^a = (0,1,0,0)$. Suppose there exists another vector m_a satisfying $m_{a;b} = 0$. From the integrability conditions we find that

$$R_{ab}m^b = 0, \tag{A1}$$

and since the only nonzero components of R_{ab} are R_{00}, R_{02}, R_{03}, and $R_{22} = R_{33} = \tfrac{1}{2}P^{-2}R$, these conditions are

$$\begin{aligned} R_{00}m^0 + R_{02}m^2 + R_{03}m^3 &= 0,\\ R_{20}m^0 + R_{22}m^2 \qquad\qquad &= 0,\\ R_{30}m^0 \qquad\qquad + R_{33}m^3 &= 0. \end{aligned} \tag{A2}$$

Now m^0, m^2, and m^3 cannot all be zero, otherwise m^a will be a constant multiple of k^a, so that the determinant of the system (A2) must vanish, i.e.,

$$RR_{00} = 2P^2(R_{02}^2 + R_{03}^2). \tag{A3}$$

But, from Ref. 1, this is precisely the condition for the metric (21) to admit a T_{ab} of Segré type {(1,1) (1 1)}, which implies that there are two null eigenvectors, k^a and l^a, such that $k^a l_a = 1$, each corresponding to the same eigenvalue, and T_{ab} is given by[1]

$$T_{ab} = -\tfrac{1}{2}R(k_a l_b + k_b l_a), \tag{A4}$$

so that $T_{ab}k^b = -\tfrac{1}{2}Rk_a$ and $T_{ab}l^b = -\tfrac{1}{2}Rl_a$. Now from (A1) we see that m_a also satisfies $T_{ab}m^b = -\tfrac{1}{2}Rm_a$, so that m_a lies in the two-space spanned by k_a and l_a and, without loss of generality, we may take $m_a \equiv l_a$. Defining a timelike unit vector, u_a, and a spacelike unit vector, n_a, by $u_a = (1/\sqrt{2})(k_a + l_a)$, $n_a = (1/\sqrt{2})(k_a - l_a)$, so that $u_a n^a = 0$, we see that u_a and n_a are each covariantly constant and coordinates can be chosen so that the space-time can be written in the form

$$ds^2 = -dt^2 + dv^2 + p_{AB}(x^C)dx^A dx^B,$$

where $(A,B) = (2,3)$, and a further coordinate transformation leads to the form

$$ds^2 = -du^2 - 2\,du\,dv + P^{-2}(x,y)(dx^2 + dy^2), \tag{A5}$$

which is the metric (21) with $m = 0$, $H = 1/2$, $P_u = 0$.

We now have to determine whether or not this metric admits an SACV or SCKV. Applying Eqs. (4.23) of Ref. 1 to the metric (A5), we find that these equations can be satisfied only if

$$P_x^2 + P_y^2 - PP_{xx} - PP_{yy} = 0,$$

which is precisely the condition $R = 0$, thus contradicting our initial premise. Hence, none of the space-times admitting SACV's or SCKV's admit any covariantly constant vector other than $\psi_{,a}$.

[1]A. A. Coley and B. O. J. Tupper, J. Math. Phys. **30**, 2616 (1989).
[2]G. S. Hall and J. da Costa, J. Math. Phys. **29**, 2465 (1988).
[3]K. L. Duggal, J. Math. Phys. **30**, 1316 (1989).
[4]K. L. Duggal and R. Sharma, J. Math. Phys. **27**, 2511 (1986).
[5]D. P. Mason and R. Maartens, J. Math. Phys. **28**, 2182 (1987).
[6]G. S. Hall, Gen. Relativ. Gravit., to appear.
[7]D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980).
[8]G. S. Hall (private communication).
[9]M. MacLean, in *Proceedings of the Third Canadian Conference on General Relativity and Relativistic Astrophysics*, edited by A. Coley, F. Cooperstock, and B. Tupper (World Scientific, Singapore, 1989).
[10]G. S. Hall, Gen. Relativ. Gravit., to appear.

The group theoretic analysis of hyperheavenly equations

M. Przanowski
Institute of Physics, Technical University of Łódź, Wólczańska 219, 93-005 Łódź, Poland

(Received 16 June 1989; accepted for publication 11 October 1989)

Contact symmetries of generalized hyperheavenly and hyperheavenly equations are investigated. It is shown that the groups of contact transformations admitted by these equations are the first prolongations of appropriate point transformation groups.

I. INTRODUCTION

This is the second part of our work devoted to the group theoretic analysis of nonlinear partial differential equations playing a distinguished role in the gravitational instanton theory and/or in the complex relativity.

The previous paper[1] has dealt with the main equations of gravitational instanton theory and their complex extensions. Now we intend to study the hyperheavenly equations that are evidently the most fundamental differential equations of the complex relativity.

As it has been demonstrated by Plebański and Robinson[2] (see also Refs. 3–5) for all one-sided algebraically degenerated, Ricci-flat complex space-times, Einstein's equations can be reduced to a single nonlinear partial differential equation of the second order on one holomorphic function. This equation is called the hyperheavenly equation. Although several solutions of the hyperheavenly equations are known[5–7] we are very far from understanding the procedure leading to the general solutions of these equations. Our belief is that the group theoretic analysis can enlighten this problem.

The role of group theoretical methods in the theory of differential equations is well known[8–11] (the method of invariant variables, the generating of new solutions by group transformations, the linearization of nonlinear equations, etc.). Therefore, it seems to be reasonable to accomplish the systematic analysis of hyperheavenly equations from the group theoretical viewpoint. This is just the purpose of the present paper. We find the general generating functions of the groups of contact transformations leaving the hyperheavenly equations invariant. It appears that these groups are the first prolongations of appropriate point transformation groups.

In this paper, as in the previous one,[1] we employ the formalism of Lie–Bäcklund transformations.[8,10,11] According to this formalism, we consider the invariance equation for the partial differential equation

$$F(x^i,u,u_{i_1},u_{i_1 i_2},\ldots)=0, \tag{1.1}$$

in the form of

$$X_c F|_{\mathscr{F}^\infty}=0, \tag{1.2}$$

where X_c is the canonical Lie–Bäcklund operator:

$$X_c=\mu\frac{\partial}{\partial u}+\sum_{s\geq 1} D_{i_1}\cdots D_{i_s}(\mu)\frac{\partial}{\partial u_{i_1\cdots i_s}},$$

$$\mu=\mu(x^i,u,u_i),$$

$$D_i:=\frac{\partial}{\partial x^i}+u_i\frac{\partial}{\partial u}+\sum_{s\geq 1} u_{ii_1\cdots i_s}\frac{\partial}{\partial u_{i_1\cdots i_s}}; \tag{1.3}$$

$\mathscr{F}^\infty$ is the infinite prolongation of Eq. (1.1) and $|_{\mathscr{F}^\infty}$ means the "restriction to $\mathscr{F}^\infty$"; $x^i, u, u_i, u_{i_1 i_2,\ldots}, i, i_1, i_2,\ldots=1,2,3,4,$ are the coordinates on the complex jet bundle, $J^\infty(C^4,C)$.

Then the infinitesimal operator of the group of contact transformations admitted by Eq. (1.1),

$$X=\xi^i\frac{\partial}{\partial x^i}+\eta\frac{\partial}{\partial u}+\zeta_i\frac{\partial}{\partial u_i}, \tag{1.4}$$

is defined by the operator X_c as follows:

$$\xi^i=-\frac{\partial\mu}{\partial u_i},\quad \eta=\mu-u_i\frac{\partial\mu}{\partial u_i},\quad \zeta_i=\frac{\partial\mu}{\partial x^i}+u_i\frac{\partial\mu}{\partial u}. \tag{1.5}$$

The function $\mu=\mu(x^i,u,u_i)$ is called a generating function of the contact transformation group.

One can easily find that the contact transformation group appears to be the first prolongation of some group of point transformations iff

$$\frac{\partial^2\mu}{\partial u_{i_1}\,\partial u_{i_2}}=0. \tag{1.6}$$

Our paper is organized as follows: In Sec. II we examine the so-called generalized hyperheavenly equation that contains as special cases the hyperheavenly equations. It is shown that the maximal group of contact transformations admitted by the generalized hyperheavenly equation is the first prolongation of the group of point transformations. Section III is devoted to the group theoretic analysis of the hyperheavenly equations. We consider "nonexpanding" and "expanding" spaces and for all cases the general solution of the invariance equation (1.2) is given. The computations are very long and hard, and, as a rule, we omit them here. Concluding remarks close the paper.

II. THE GENERALIZED HYPERHEAVENLY EQUATION

In this section, we study symmetries of the following equation:

$$F: J^2(C^4,C)\to C,$$

$$F=u_{11}u_{22}-u_{12}^2+f(x^i)\cdot(u_{14}+u_{23})+g(x^i,u,u_i,u_{11},u_{12},u_{22})=0,\quad i=1,\ldots,4, \tag{2.1}$$

where $f=f(x^i)\neq 0$ and $g=g(x^i,u,u_i,u_{11},u_{12},u_{22})$ are holomorphic functions of their arguments and, moreover, g is linear with respect to the variables u_{11}, u_{12}, u_{22}. Since Eq. (2.1) contains as its special cases the hyperheavenly equations we call it the *generalized hyperheavenly equation.*

In order to find the contact symmetries of Eq. (2.1) one

 0022-2488/90/030653-06$03.00

must solve the invariance equation (1.2), which now takes the form

$$\Big\{\Big(u_{22}+\frac{\partial g}{\partial u_{11}}\Big)\cdot D_1D_1(\mu)+\Big(u_{11}+\frac{\partial g}{\partial u_{22}}\Big)\cdot D_2D_2(\mu)$$
$$+\Big(-2u_{12}+\frac{\partial g}{\partial u_{12}}\Big)\cdot D_1D_2(\mu)+f\cdot[D_1D_4(\mu)$$
$$+D_2D_3(\mu)]+\mu\frac{\partial g}{\partial u}+\frac{\partial g}{\partial u_i}\cdot D_i(\mu)\Big\}_{\mathcal{F}^\infty}=0. \quad (2.2)$$

Extracting from (2.2) the elements containing u_{33} or u_{44} one has

$$\Big[fu_{33}\cdot\Big(\frac{\partial^2\mu}{\partial x^2\,\partial u_3}+u_2\frac{\partial^2\mu}{\partial u\,\partial u_3}+u_{2i}\frac{\partial^2\mu}{\partial u_i\,\partial u_3}\Big)\Big]_{\mathcal{F}^\infty}=0, \quad (2.3)$$

$$\Big[fu_{44}\cdot\Big(\frac{\partial^2\mu}{\partial x^1\,\partial u_4}+u_1\frac{\partial^2\mu}{\partial u\,\partial u_4}+u_{1i}\frac{\partial^2\mu}{\partial u_i\,\partial u_4}\Big)\Big]_{\mathcal{F}^\infty}=0. \quad (2.4)$$

Hence,

$$\frac{\partial^2\mu}{\partial u_i\,\partial u_3}=0=\frac{\partial^2\mu}{\partial u_i\,\partial u_4},$$
$$\frac{\partial^2\mu}{\partial x^2\,\partial u_3}+u_2\frac{\partial^2\mu}{\partial u\,\partial u_3}=0=\frac{\partial^2\mu}{\partial x^1\,\partial u_4}+u_1\frac{\partial^2\mu}{\partial u\,\partial u_4}. \quad (2.5)$$

From (2.5) it follows that

$$\mu=\rho(x^i,u,u_1,u_2)+u_3\nu(x^1,x^3,x^4)$$
$$+u_4\omega(x^2,x^3,x^4), \quad (2.6)$$

where $\rho=\rho(x^i,u,u_1,u_2)$, $\nu=\nu(x^1,x^3,x^4)$, and $\omega=\omega(x^2,x^3,x^4)$ are some holomorphic functions of their arguments.

Then, employing (2.6) and extracting from (2.2) the elements containing u_{13}, u_{24}, or u_{34} we find

$$\Big\{u_{13}\cdot\Big[2\cdot\Big(u_{22}+\frac{\partial g}{\partial u_{11}}\Big)\frac{\partial\nu}{\partial x^1}+f\cdot\Big(\frac{\partial\nu}{\partial x^4}+\frac{\partial^2\rho}{\partial x^2\,\partial u_1}$$
$$+u_2\frac{\partial^2\rho}{\partial u\,\partial u_1}+u_{12}\frac{\partial^2\rho}{\partial u_1^2}+u_{22}\frac{\partial^2\rho}{\partial u_1\,\partial u_2}\Big)\Big]\Big\}_{\mathcal{F}^\infty}=0, \quad (2.7)$$

$$\Big\{u_{24}\cdot\Big[2\cdot\Big(u_{11}+\frac{\partial g}{\partial u_{22}}\Big)\frac{\partial\omega}{\partial x^2}+f\cdot\Big(\frac{\partial\omega}{\partial x^3}+\frac{\partial^2\rho}{\partial x^1\,\partial u_2}$$
$$+u_1\frac{\partial^2\rho}{\partial u\,\partial u_2}+u_{12}\frac{\partial^2\rho}{\partial u_2^2}+u_{11}\frac{\partial^2\rho}{\partial u_1\,\partial u_2}\Big)\Big]\Big\}_{\mathcal{F}^\infty}=0, \quad (2.8)$$

$$\Big\{fu_{34}\cdot\Big(\frac{\partial\nu}{\partial x^1}+\frac{\partial\omega}{\partial x^2}\Big)\Big\}_{\mathcal{F}^\infty}=0. \quad (2.9)$$

From (2.7)–(2.9) one infers that

$$\frac{\partial^2\rho}{\partial u_1^2}=\frac{\partial^2\rho}{\partial u_2^2}=\frac{\partial^2\rho}{\partial u_1\,\partial u_2}=0,\quad \frac{\partial\nu}{\partial x^1}=\frac{\partial\omega}{\partial x^2}=0,$$
$$\frac{\partial^2\rho}{\partial u\,\partial u_1}=\frac{\partial^2\rho}{\partial u\,\partial u_2}=0,\quad \frac{\partial\nu}{\partial x^4}+\frac{\partial^2\rho}{\partial x^2\,\partial u_1}=0, \quad (2.10)$$
$$\frac{\partial\omega}{\partial x^3}+\frac{\partial^2\rho}{\partial x^1\,\partial u_2}=0.$$

Consequently, μ is of the form

$$\mu=u_1\alpha(x^i)+u_2\beta(x^i)+u_3\gamma(x^3,x^4)$$
$$+u_4\delta(x^3,x^4)+\varepsilon(x^i,u), \quad (2.11)$$

where $\alpha=\alpha(x^i)$, $\beta=\beta(x^i)$, $\gamma=\gamma(x^3,x^4)$, $\delta=\delta(x^3,x^4)$, and $\varepsilon=\varepsilon(x^i,u)$ are some holomorphic functions of their arguments, and the following equations are satisfied:

$$\frac{\partial\gamma}{\partial x^4}+\frac{\partial\alpha}{\partial x^2}=0,\quad \frac{\partial\delta}{\partial x^3}+\frac{\partial\beta}{\partial x^1}=0. \quad (2.12)$$

Gathering the elements of (2.2) containing $u_{11}\cdot u_{22}$ or u_{12}^2 or u_{23} one finds that

$$\varepsilon(x^i,u)=u\sigma(x^i)+\tau(x^i); \quad (2.13)$$

moreover,

$$\frac{\partial\alpha}{\partial x^1}+2\frac{\partial\beta}{\partial x^4}+f^{-1}\cdot\Big(\frac{\partial f}{\partial x^1}\alpha+\frac{\partial f}{\partial x^2}\beta+\frac{\partial f}{\partial x^3}\gamma+\frac{\partial f}{\partial x^4}\delta\Big)$$
$$-\frac{\partial\delta}{\partial x^4}+\sigma=0, \quad (2.14)$$
$$\frac{\partial\alpha}{\partial x^1}-\frac{\partial\beta}{\partial x^2}-\frac{\partial\gamma}{\partial x^3}+\frac{\partial\delta}{\partial x^4}=0.$$

The conditions (2.12) yield

$$\alpha=\alpha_1(x^3,x^4)\cdot x^1+\alpha_2(x^3,x^4)\cdot x^2+\alpha_3(x^3,x^4),$$
$$\beta=\beta_1(x^3,x^4)\cdot x^1+\beta_2(x^3,x^4)\cdot x^2+\beta_3(x^3,x^4), \quad (2.15)$$

and

$$\alpha_2=-\frac{\partial\gamma}{\partial x^4},\quad \beta_1=-\frac{\partial\delta}{\partial x^3}. \quad (2.16)$$

Then, from (2.14) with (2.15) one gets

$$\alpha_1+2\beta_2-\frac{\partial\delta}{\partial x^4}+f^{-1}\cdot\Big(\frac{\partial f}{\partial x^1}\alpha+\frac{\partial f}{\partial x^2}\beta+\frac{\partial f}{\partial x^3}\gamma+\frac{\partial f}{\partial x^4}\delta\Big)$$
$$+\sigma=0, \quad (2.17)$$
$$\alpha_1-\beta_2=\frac{\partial\gamma}{\partial x^3}-\frac{\partial\delta}{\partial x^4}.$$

Finally, gathering the remaining elements of (2.2) we obtain the following condition:

$$\Big\{u_{11}\cdot\Big[\frac{\partial g}{\partial u_{11}}\cdot(\sigma+2\alpha_1)+\frac{\partial g}{\partial u_{12}}\alpha_2+u\frac{\partial^2\sigma}{\partial(x^2)^2}+\frac{\partial^2\tau}{\partial(x^2)^2}+2u_2\frac{\partial\sigma}{\partial x^2}+f\frac{\partial\alpha}{\partial x^4}\Big]+u_{12}\cdot\Big[2\frac{\partial g}{\partial u_{11}}\beta_1+\frac{\partial g}{\partial u_{12}}\cdot(\sigma+\alpha_1+\beta_2)+2\frac{\partial g}{\partial u_{22}}\alpha_2$$
$$-2\cdot\Big(u\frac{\partial^2\sigma}{\partial x^1\,\partial x^2}+\frac{\partial^2\tau}{\partial x^1\,\partial x^2}\Big)-2u_1\frac{\partial\sigma}{\partial x^2}-2u_2\frac{\partial\sigma}{\partial x^1}+f\cdot\Big(\frac{\partial\alpha}{\partial x^3}+\frac{\partial\beta}{\partial x^4}\Big)\Big]+u_{22}\cdot\Big[\frac{\partial g}{\partial u_{12}}\beta_1+\frac{\partial g}{\partial u_{22}}\cdot(\sigma+2\beta_2)+u\frac{\partial^2\sigma}{\partial(x^1)^2}+\frac{\partial^2\tau}{\partial(x^1)^2}$$
$$+2u_1\cdot\frac{\partial\sigma}{\partial x^1}+f\frac{\partial\beta}{\partial x^3}\Big]+\frac{\partial g}{\partial u_{11}}\cdot\Big(u\frac{\partial^2\sigma}{\partial(x^1)^2}+\frac{\partial^2\tau}{\partial(x^1)^2}+2u_1\frac{\partial\sigma}{\partial x^1}\Big)+\frac{\partial g}{\partial u_{12}}\cdot\Big(u\frac{\partial^2\sigma}{\partial x^1\,\partial x^2}+\frac{\partial^2\tau}{\partial x^1\,\partial x^2}+u_1\frac{\partial\sigma}{\partial x^2}+u_2\frac{\partial\sigma}{\partial x^1}\Big)+\frac{\partial g}{\partial u_{22}}$$
$$\cdot\Big(u\frac{\partial^2\sigma}{\partial(x^2)^2}+\frac{\partial^2\tau}{\partial(x^2)^2}+2u_2\frac{\partial\sigma}{\partial x^2}\Big)+\frac{\partial g}{\partial u_i}\cdot\Big(\frac{\partial\mu}{\partial x^i}+u_i\sigma\Big)+\frac{\partial g}{\partial u}\cdot(u\sigma+\tau)-\frac{\partial g}{\partial x^i}\frac{\partial\mu}{\partial u_i}-2g\cdot(\sigma+\alpha_1+\beta_2)+f\cdot\Big[u_1\cdot\Big(\frac{\partial\alpha_1}{\partial x^4}+\frac{\partial\alpha_2}{\partial x^3}+\frac{\partial\sigma}{\partial x^4}\Big)$$
$$+u_2\cdot\Big(\frac{\partial\beta_1}{\partial x^4}+\frac{\partial\beta_2}{\partial x^3}+\frac{\partial\sigma}{\partial x^3}\Big)+u_3\frac{\partial\sigma}{\partial x^2}+u_4\frac{\partial\sigma}{\partial x^1}+u\cdot\Big(\frac{\partial^2\sigma}{\partial x^1\,\partial x^4}+\frac{\partial^2\sigma}{\partial x^2\,\partial x^3}\Big)+\frac{\partial^2\tau}{\partial x^1\,\partial x^4}+\frac{\partial^2\tau}{\partial x^2\,\partial x^3}\Big]\Big\}_{\mathcal{F}^\infty}=0. \quad (2.18)$$

Now, as μ defined by (2.11) satisfies the condition (1.6) we arrive at the following theorem.

Theorem 2.1: Every group of contact transformations admitted by the generalized hyperheavenly equation (2.1) is the first prolongation of some group of point transformations admitted by this equation. The general infinitesimal operator of the maximal point transformation group leaving Eq. (2.1) invariant is of the form

$$X_p = \xi^i \frac{\partial}{\partial x^i} + \eta \frac{\partial}{\partial u},$$
$$\xi^1 = -\alpha_1(x^3,x^4)\cdot x^1 - \alpha_2(x^3,x^4)\cdot x^2 - \alpha_3(x^3,x^4),$$
$$\xi^2 = -\beta_1(x^3,x^4)\cdot x^1 - \beta_2(x^3,x^4)\cdot x^2 - \beta_3(x^3,x^4),$$
$$\xi^3 = -\gamma(x^3,x^4), \quad \xi^4 = -\delta(x^3,x^4),$$
$$\eta = u\sigma(x^i) + \tau(x^i), \tag{2.19}$$

where the functions $\alpha_1, \alpha_2, \alpha_3, \beta_1, \beta_2, \beta_3, \gamma, \delta, \sigma, \tau$, constitute the general solution of Eqs. (2.16)–(2.18). ■

III. SYMMETRIES OF THE HYPERHEAVENLY EQUATIONS

Plebański and Robinson[2] (see also Refs. 3–5) have shown that ten Einstein equations for the metric of one-sided algebraically degenerated, Ricci-flat complex space-time can be reduced to a single nonlinear partial differential equation of the second order on one holomorphic function. The crucial point in this reduction procedure is the existence of a null string foliation. The leaves of the foliation are null strings, i.e., totally null, geodesic two-dimensional complex surfaces. One can choose the local coordinates x^i, $i = 1,...,4$, so that the equations $x^3 = \text{const}$, $x^4 = \text{const}$ define the null strings of the foliation and x^1, x^2 are coordinates along these null strings. Then we define the expansion form of the null string foliation to be a one-form

$$\Theta = \tfrac{1}{2}\,[(x^3)^{;i}{}_{;i}\, dx^4 - (x^4)^{;i}{}_{;i}\, dx^3], \tag{3.1}$$

where ";" stands for the covariant derivative.

As it has been demonstrated in Ref. 2, it is necessary to consider two essentially distinct cases: (i) the nonexpanding null string foliation, i.e., $\theta = 0$; (ii) the expanding null string foliation, i.e., $\theta \neq 0$. The equations obtained by Plebański and Robinson are called the hyperheavenly equations.

In the present section, we give the group theoretic analysis of these equations. It appears that the hyperheavenly equations are speical cases of our generalized hyperheavenly equation. Consequently, all results of Sec. II remain valid for the hyperheavenly equations. But now we can do much better. Namely, we can considerably simplify Eqs. (2.16)–(2.18), and in most cases we are able to find the general solutions of these equations. The cases $\theta = 0$ or $\theta \neq 0$ are considered separately. Notation is the same as in Sec. II.

A. Nonexpanding null string foliation, $\Theta = 0$

1. The Petrov–Penrose–Plebański type [III]⊗[any]

The hyperheavenly equation now takes the form of (2.1) with

$$f = 1, \quad g = \tfrac{1}{2}\cdot(x^1x^4 + x^2x^3)^2 + 3\cdot(u_1x^3 - u_2x^4) - 2u_{11}x^1x^3 + 2u_{12}\cdot(x^1x^4 - x^2x^3) + 2u_{22}x^2x^4. \tag{3.2}$$

[Compare Refs. 4 and 5 for

$$p^{\dot A} = \begin{pmatrix} -x^1 \\ -x^2 \end{pmatrix}, \quad p_{\dot A} = \begin{pmatrix} -x^2 \\ x^1 \end{pmatrix},$$
$$q^{\dot A} = \begin{pmatrix} -x^3 \\ x^4 \end{pmatrix}, \quad q_{\dot A} = \begin{pmatrix} x^4 \\ x^3 \end{pmatrix}, \tag{3.3}$$

and $N^{\dot A} = 0$, $F^{\dot A} = 3q^{\dot A}$.]

Then, by long and tedious manipulations one finds that Eqs. (2.16)–(2.18) lead to the following results:

$$\sigma(x^i) = \sigma_0 = \text{const},$$
$$\alpha_1 = \frac{\partial^2\varphi}{\partial x^3\,\partial x^4} - \frac{1}{3}\sigma_0, \quad \alpha_2 = -\frac{\partial^2\varphi}{\partial (x^4)^2},$$
$$\beta_1 = \frac{\partial^2\varphi}{\partial (x^3)^2}, \quad \beta_2 = -\frac{\partial^2\varphi}{\partial x^3\,\partial x^4} - \frac{1}{3}\sigma_0,$$
$$\alpha_3 = \frac{1}{6}\frac{\partial\psi}{\partial x^4} + \frac{1}{2}\psi\cdot x^3, \quad \beta_3 = \frac{1}{6}\frac{\partial\psi}{\partial x^3} - \frac{1}{2}\psi\cdot x^4,$$
$$\gamma = \frac{\partial\varphi}{\partial x^4}, \quad \delta = -\frac{\partial\varphi}{\partial x^3},$$
$$\tau = -\frac{1}{6}\cdot\left(x^1\frac{\partial}{\partial x^3} - x^2\frac{\partial}{\partial x^4}\right)^3\varphi + \frac{1}{2}\cdot\left(2x^4\beta_3 - \frac{\partial\beta_3}{\partial x^3}\right)\cdot(x^1)^2 - \frac{1}{2}\cdot\left(2x^3\alpha_3 + \frac{\partial\alpha_3}{\partial x^4}\right)\cdot(x^2)^2 + \frac{1}{2}\cdot\left(2x^3\beta_3 - 2x^4\alpha_3 + \frac{\partial\alpha_3}{\partial x^3} + \frac{\partial\beta_3}{\partial x^4}\right)x^1x^2 + \chi_1x^1 + \chi_2x^2 + \varkappa, \tag{3.4}$$

where $\varkappa = \varkappa(x^3,x^4)$ is an arbitrary holomorphic function of x^3,x^4, and $\varphi = \varphi(x^3,x^4)$, $\psi = \psi(x^3,x^4)$, $\chi_1 = \chi_1(x^3,x^4)$, $\chi_2 = \chi_2(x^3,x^4)$ constitute the general solution of the following equations:

$$x^3\frac{\partial\varphi}{\partial x^3} + x^4\frac{\partial\varphi}{\partial x^4} - 2\varphi = 0, \tag{3.5a}$$
$$x^3\frac{\partial\psi}{\partial x^3} + x^4\frac{\partial\psi}{\partial x^4} + 4\psi = 0, \tag{3.5b}$$
$$\frac{\partial\chi_2}{\partial x^3} + \frac{\partial\chi_1}{\partial x^4} + 3x^3\chi_1 - 3x^4\chi_2 = 0. \tag{3.5c}$$

The general solution of Eqs. (3.5a) and (3.5b) takes the form

$$\varphi = (x^3)^2\cdot\tilde\varphi(x^3/x^4), \quad \psi = (x^3)^{-4}\tilde\psi(x^3/x^4), \tag{3.6}$$

where $\tilde\varphi(x^3/x^4)$, $\tilde\psi(x^3/x^4)$ are arbitrary holomorphic functions of the variable x^3/x^4.

Thus we arrive at the conclusion that the general generating function μ is now defined by the formulas (3.4), (3.6), and (3.5c).

2. The type [N]⊗[any]

In this case,

$$f = 1, \quad g = N(x^3,x^4)\cdot(x^1x^4 + x^2x^3), \tag{3.7}$$

where $N = N(x^3,x^4)$ is a nowhere vanishing holomorphic function of x^3,x^4. (Compare Refs. 4 and 5 for $N^{\dot A} = N\cdot q^{\dot A}$, $F^{\dot A} = 0$). With (3.7) assumed, Eqs. (2.16)–(2.18) yield

$$\sigma(x^i)=\sigma_0=\text{const},$$
$$\alpha_1=\frac{\partial\gamma}{\partial x^3}-\sigma_0+\frac{2}{3}a_0,\quad \alpha_2=-\frac{\partial\gamma}{\partial x^4},$$
$$\beta_1=-\frac{\partial\delta}{\partial x^3},\quad \beta_2=\frac{\partial\delta}{\partial x^4}-\sigma_0+\frac{2}{3}a_0,\quad a_0=\text{const},$$
$$\begin{aligned}\tau=&\frac{1}{6}\cdot\left[\frac{\partial^2\delta}{\partial(x^3)^2}\cdot(x^1)^3+\frac{\partial^2\gamma}{\partial(x^4)^2}\cdot(x^2)^3\right]\\&-\frac{1}{2}\cdot\left[\frac{\partial^2\delta}{\partial x^3\,\partial x^4}\cdot(x^1)^2x^2+\frac{\partial^2\gamma}{\partial x^3\,\partial x^4}\,x^1\cdot(x^2)^2\right]\\&-\frac{1}{2}\cdot\left[\frac{\partial\beta_3}{\partial x^3}\cdot(x^1)^2+\frac{\partial\alpha_3}{\partial x^4}\cdot(x^2)^2\right]\\&+\frac{1}{2}\cdot\left(\frac{\partial\alpha_3}{\partial x^3}+\frac{\partial\beta_3}{\partial x^4}\right)x^1x^2\\&+\chi_1x^1+\chi_2x^2+\varkappa,\end{aligned}\tag{3.8}$$

where $\varkappa=\varkappa(x^3,x^4)$ is an arbitrary holomorphic function of x^3, x^4, and the functions $\gamma=\gamma(x^3,x^4)$, $\delta=\delta(x^3,x^4)$, $\alpha_3=\alpha_3(x^3,x^4)$, $\beta_3=\beta_3(x^3,x^4)$, $\chi_1=\chi_1(x^3,x^4)$, $\chi_2=\chi_2(x^3,x^4)$ constitute the general solution of the following set of equations:

$$\frac{\partial\gamma}{\partial x^3}+\frac{\partial\delta}{\partial x^4}=2\cdot(\sigma_0-a_0),\tag{3.9a}$$
$$\frac{\partial}{\partial x^3}\left[\frac{\partial\alpha_3}{\partial x^3}-\frac{\partial\beta_3}{\partial x^4}+2N\cdot(\delta x^3-\gamma x^4)\right]=-C^{(1)}\delta+2N\cdot(\sigma_0-\tfrac{2}{3}a_0)x^4,\tag{3.9b}$$
$$\frac{\partial}{\partial x^4}\left[\frac{\partial\alpha_3}{\partial x^3}-\frac{\partial\beta_3}{\partial x^4}+2N\cdot(\delta x^3-\gamma x^4)\right]=C^{(1)}\gamma-2N\cdot(\sigma_0-\tfrac{2}{3}a_0)x^3,\tag{3.9c}$$
$$\frac{\partial\chi_2}{\partial x^3}+\frac{\partial\chi_1}{\partial x^4}=N\cdot(\beta_3x^3+\alpha_3x^4),\tag{3.9d}$$

where

$$C^{(1)}=2C_{2222}=-2\cdot\left[\frac{\partial(Nx^3)}{\partial x^3}+\frac{\partial(Nx^4)}{\partial x^4}\right]\neq0\tag{3.10}$$

and C_{2222} is the only nonvanishing component of the undotted (self-dual) Weyl spinor.

Define

$$\varphi:=\frac{\partial\alpha_3}{\partial x^3}-\frac{\partial\beta_3}{\partial x^4}+2N\cdot(\delta x^3-\gamma x^4).\tag{3.11}$$

Then from (3.9b), and (3.9c) one gets

$$\begin{aligned}\gamma&=(C^{(1)})^{-1}\cdot\left[\frac{\partial\varphi}{\partial x^4}+2N\cdot\left(\sigma_0-\frac{2}{3}a_0\right)x^3\right],\\\delta&=(C^{(1)})^{-1}\cdot\left[-\frac{\partial\varphi}{\partial x^3}+2N\cdot\left(\sigma_0-\frac{2}{3}a_0\right)x^4\right].\end{aligned}\tag{3.12}$$

Consequently, (3.12) and (3.9a) lead to the equation on φ:

$$\begin{aligned}&\frac{\partial C^{(1)}}{\partial x^4}\frac{\partial\varphi}{\partial x^3}-\frac{\partial C^{(1)}}{\partial x^3}\frac{\partial\varphi}{\partial x^4}-2N\cdot\left(\sigma_0-\frac{2}{3}a_0\right)\\&\quad\cdot\left(\frac{\partial C^{(1)}}{\partial x^3}x^3+\frac{\partial C^{(1)}}{\partial x^4}x^4\right)-\left(3\sigma_0-\frac{8}{3}a_0\right)\cdot(C^{(1)})^2=0,\end{aligned}\tag{3.13}$$

and (3.11) with (3.12) gives

$$\frac{\partial\alpha_3}{\partial x^3}-\frac{\partial\beta_3}{\partial x^4}=\varphi+2N\cdot(C^{(1)})^{-1}\cdot\left(\frac{\partial\varphi}{\partial x^3}x^3+\frac{\partial\varphi}{\partial x^4}x^4\right).\tag{3.14}$$

Gathering, one concludes that the general generating function μ in the case of nonexpanding $[N]\otimes[\text{any}]$ type is defined by the formulas (3.8), (3.9d), (3.12), (3.13), and (3.14).

[Notice that if $C^{(1)}=\text{const}$, then from (3.13) it follows that $3\sigma_0-\frac{8}{3}a_0=0$ and φ is an arbitrary holomorphic function of x^3, x^4.]

3. The type $[-]\otimes[\text{any}]$ ($\equiv$heaven)

We now have

$$f=1,\quad g=0,\tag{3.15}$$

and Eq. (2.1) takes the form of the "second heavenly equation" obtained by Plebański in his fundamental work on the heavenly spaces.[12] Symmetries of this equation have been examined by Boyer and Plebański.[13]

In order to find these symmetries, we specialize the preceding considerations concerning the type $[N]\otimes[\text{any}]$ to the case $N=N(x^3,x^4)=0$. Consequently, Eqs. (3.9a)–(3.9d) yield

$$\gamma=\frac{\partial\varphi}{\partial x^4}+(\sigma_0-a_0)\cdot x^3,\quad \delta=-\frac{\partial\varphi}{\partial x^3}+(\sigma_0-a_0)\cdot x^4,\tag{3.16}$$

$$\alpha_3=\frac{\partial\psi}{\partial x^4}+b_0x^3,\quad \beta_3=\frac{\partial\psi}{\partial x^3}-b_0x^4,\quad b_0=\text{const},\tag{3.17}$$

$$\chi_2=\frac{\partial\chi}{\partial x^4},\quad \chi_1=-\frac{\partial\chi}{\partial x^3},\tag{3.18}$$

where σ_0, a_0 are constants defined as before [see (3.8)], and $\varphi=\varphi(x^3,x^4)$, $\psi=\psi(x^3,x^4)$, $\chi=\chi(x^3,x^4)$ are arbitrary holomorphic functions of the variables x^3, x^4. Then, from (3.8) with (3.16)–(3.18) one has

$$\alpha_1=\frac{\partial^2\varphi}{\partial x^3\,\partial x^4}-\frac{1}{3}a_0,\quad \alpha_2=-\frac{\partial^2\varphi}{\partial(x^4)^2},$$
$$\beta_1=\frac{\partial^2\varphi}{\partial(x^3)^2},\quad \beta_2=-\frac{\partial^2\varphi}{\partial x^3\,\partial x^4}-\frac{1}{3}a_0,$$
$$\begin{aligned}\tau=&-\frac{1}{6}\cdot\left(x^1\frac{\partial}{\partial x^3}-x^2\frac{\partial}{\partial x^4}\right)^3\varphi-\frac{1}{2}\cdot\left(x^1\frac{\partial}{\partial x^3}-x^2\frac{\partial}{\partial x^4}\right)^2\psi\\&-\left(x^1\frac{\partial}{\partial x^3}-x^2\frac{\partial}{\partial x^4}\right)\chi+\varkappa.\end{aligned}\tag{3.19}$$

This ends our analysis of the hyperheavenly equations for Ricci-flat complex space-time admitting a nonexpanding null string foliation. In the next section, we consider the "expanding case."

B. Expanding null string foliation, $\theta\neq0$

In this case, the general form of the hyperheavenly equation can be presented by (2.1) with

$$f=(x^1+x^2)^{-1},$$

$$g=\omega(x^3,x^4)+\tfrac{1}{2}\nu(x^3,x^4)\cdot(x^2-x^1)-3\rho_0 u+\tfrac{3}{2}\rho_0\cdot(x^1+x^2)\cdot(u_1+u_2)-[\tfrac{1}{4}\rho_0\cdot(x^1+x^2)^2+2\cdot(x^1+x^2)^{-1}\cdot u_2]u_{11}$$
$$-[\tfrac{1}{2}\rho_0\cdot(x^1+x^2)^2-2\cdot(x^1+x^2)^{-1}\cdot(u_1+u_2)]u_{12}-[\tfrac{1}{4}\rho_0\cdot(x^1+x^2)^2+2\cdot(x^1+x^2)^{-1}u_1]u_{22}, \quad (3.20)$$

where $\omega=\omega(x^3,x^4)$, $\nu=\nu(x^3,x^4)$ are some holomorphic functions of x^3, x^4 and ρ_0 is some constant. [To compare (3.20) with the corresponding formulas of Refs. 3 and 5 one should specialize $J_{\dot A}$ and $K_{\dot A}$ of those references to be $J_{\dot 1}=J_{\dot 2}=-1$, $K_{\dot 1}=-K_{\dot 2}=-1$. The coordinates $p^{\dot A}$, $p_{\dot A}$, $q^{\dot A}$, $q_{\dot A}$ are defined by x^i according to (3.3).] Now the following cases must be considered.

1. The Petrov–Penrose–Plebański types: [II]⊗[any] or [D]⊗[any]

Here,

$$\rho_0\neq 0, \quad \omega=0=\nu. \quad (3.21)$$

Then, long and tedious calculations lead to the following result:

$$\alpha_1=\frac{\partial\psi}{\partial x^3}, \quad \alpha_2=\frac{\partial\varphi}{\partial x^3}-a_0, \quad \alpha_3=\chi,$$

$$\beta_1=-\frac{\partial\psi}{\partial x^3}-a_0, \quad \beta_2=-\frac{\partial\varphi}{\partial x^3}, \quad \beta_3=-\chi, a_0=\text{const},$$

$$\gamma=a_0\cdot(x^3+x^4)+\varphi, \quad \delta=a_0\cdot(x^3+x^4)+\psi,$$

$$\sigma=-2\frac{\partial}{\partial x^3}(\psi-\varphi), \quad (3.22)$$

$$\tau=-\frac{1}{8}\rho_0\frac{\partial}{\partial x^3}(\psi+\varphi)\cdot(x^1+x^2)^3\cdot(x^1-x^2)$$
$$+\lambda\cdot(x^1+x^2)^3-\frac{1}{2}\frac{\partial^2\psi}{\partial(x^3)^2}\cdot(x^1)^2-\frac{1}{2}\frac{\partial^2}{\partial(x^3)^2}$$
$$\cdot(\psi+\varphi)x^1x^2-\frac{1}{2}\frac{\partial^2\varphi}{(x^3)^2}\cdot(x^2)^2-\frac{1}{2}\frac{\partial\chi}{\partial x^3}\cdot(x^1+x^2)$$
$$+\frac{1}{6}\cdot(\rho_0)^{-1}\frac{\partial^3}{\partial(x^3)^3}(\psi-\varphi),$$

where $\varphi=\varphi(x^3-x^4)$, $\psi=\psi(x^3-x^4)$, $\chi=\chi(x^3-x^4)$, $\lambda=\lambda(x^3-x^4)$ are arbitrary holomorphic functions of the variable x^3-x^4.

2. The type [III]⊗[any]

This case is realized for (see Ref. 5)

$$\rho_0=0, \quad \nu=\nu_0=\text{const}\neq 0, \quad (3.23)$$

$\omega=\omega(x^3-x^4)$, i.e., ω is a function of x^3-x^4.

With (3.23) assumed, one can integrate Eqs. (2.16)–(2.18). Thus one finds

$$\alpha_1=\frac{1}{4}\frac{\partial^2(\psi-\varphi)}{\partial(x^3)^2}\cdot(x^3+x^4)+\frac{1}{4}\frac{\partial(7\psi-3\varphi)}{\partial x^3}-a_0,$$

$$\alpha_2=\frac{1}{4}\frac{\partial^2(\psi-\varphi)}{\partial(x^3)^2}\cdot(x^3+x^4)-\frac{1}{4}\frac{\partial(\psi-5\varphi)}{\partial x^3}-a_0,$$

$$\alpha_3=-(\nu_0)^{-1}\cdot\left[\frac{1}{2}\frac{\partial^3(\psi-\varphi)}{\partial(x^3)^3}+2\omega\frac{\partial(\psi-\varphi)}{\partial x^3}+(\psi-\varphi)\cdot\frac{\partial\omega}{\partial x^3}\right],$$

$$\beta_1=-\frac{1}{4}\frac{\partial^2(\psi-\varphi)}{\partial(x^3)^2}\cdot(x^3+x^4)-\frac{1}{4}\frac{\partial(5\psi-\varphi)}{\partial x^3}-a_0,$$

$$\beta_2=-\frac{1}{4}\frac{\partial^2(\psi-\varphi)}{\partial(x^3)^2}\cdot(x^3+x^4)+\frac{1}{4}\frac{\partial(3\psi-7\varphi)}{\partial x^3}-a_0,$$

$$a_0=\text{const},$$

$$\beta_3=-\alpha_3,$$

$$\gamma=\left[\frac{1}{4}\frac{\partial(\psi-\varphi)}{\partial x^3}+a_0\right]\cdot(x^3+x^4)+\varphi,$$

$$\delta=\left[\frac{1}{4}\frac{\partial(\psi-\varphi)}{\partial x^3}+a_0\right]\cdot(x^3+x^4)+\psi,$$

$$\sigma=-\frac{7}{2}\frac{\partial(\psi-\varphi)}{\partial x^3}+2a_0,$$

$$\tau=-\left\{\frac{\nu_0}{12}\cdot\left[\frac{1}{4}\frac{\partial^2(\psi-\varphi)}{\partial(x^3)^2}\cdot(x^3+x^4)^2+\frac{\partial(\psi+\varphi)}{\partial x^3}\cdot(x^3+x^4)\right]-\lambda\right\}\cdot(x^1+x^2)^3-\frac{1}{2}\frac{\partial^2\delta}{\partial(x^3)^2}\cdot(x^1)^2$$
$$+\frac{1}{2}\frac{\partial^2(\delta+\gamma)}{\partial x^3\,\partial x^4}x^1x^2-\frac{1}{2}\frac{\partial^2\gamma}{\partial(x^4)^2}\cdot(x^2)^2$$
$$-\frac{1}{2}\frac{\partial\alpha_3}{\partial x^3}\cdot(x^1+x^2)+\varkappa, \quad (3.24)$$

where $\varphi=\varphi(x^3-x^4)$, $\psi=\psi(x^3-x^4)$, $\lambda=\lambda(x^3-x^4)$ are arbitrary holomorphic functions of the variable x^3-x^4, and $\varkappa=\varkappa(x^3,x^4)$ is an arbitrary holomorphic function of x^3, x^4. The formulas (3.24), although rather involved, give explicitly the general solution of Eqs. (2.16)–(2.18). Finally, the next case is considered.

3. The type [N]⊗[any]

We now have (compare with Ref. 5)

$$\rho_0=0=\nu, \quad \omega=b_0\cdot(x^3+x^4)+\tilde\omega(x^3-x^4), \quad (3.25)$$

where $b_0=\text{const}\neq 0$; $\tilde\omega=\tilde\omega(x^3-x^4)$ is an arbitrary holomorphic function of the variable x^3-x^4. With (3.25) assumed one finds:

$$\alpha_1=\frac{\partial\delta}{\partial x^3}+2\frac{\partial\varphi}{\partial x^3}-a_0,$$

$$\alpha_2=-\frac{\partial\delta}{\partial x^4}-\frac{\partial\varphi}{\partial x^3}, \quad \alpha_3=\chi,$$

$$\beta_1=-\frac{\partial\delta}{\partial x^3}, \quad \beta_2=\frac{\partial\delta}{\partial x^4}+3\frac{\partial\varphi}{\partial x^3}-a_0,$$

$$\beta_3=-\chi, a_0=\text{const},$$

$$\gamma = \delta - \varphi, \quad \delta = (2b_0)^{-1} \cdot \left[\frac{1}{2} \frac{\partial^3 \varphi}{\partial (x^3)^3} + 2\omega \frac{\partial \varphi}{\partial x^3} + \frac{\partial \omega}{\partial x^3} \varphi \right],$$

$$\sigma = -8 \frac{\partial \varphi}{\partial x^3} + 2a_0,$$

$$\tau = \lambda \cdot (x^1 + x^2)^3 - \frac{1}{2} \frac{\partial^2 \delta}{\partial (x^3)^2} \cdot (x^1)^2 + \frac{1}{2} \frac{\partial^2 (\delta + \gamma)}{\partial x^3 \partial x^4} x^1 x^2 - \frac{1}{2} \frac{\partial^2 \gamma}{\partial (x^4)^2} \cdot (x^2)^2 - \frac{1}{2} \cdot \left(x^1 \frac{\partial}{\partial x^3} - x^2 \frac{\partial}{\partial x^4} \right) \chi + \varkappa, \tag{3.26}$$

where $\varphi = \varphi(x^3 - x^4)$, $\lambda = \lambda(x^3 - x^4)$ are arbitrary holomorphic functions of $x^3 - x^4$; $\chi = \chi(x^3, x^4)$, $\varkappa = \varkappa(x^3, x^4)$ are arbitrary holomorphic functions of x^3, x^4.

IV. CONCLUSIONS

In this paper we have shown that every contact transformation admitted by the generalized hyperheavenly equation appears to be the first prolongation of some point transformation group. Then we are able to find the general solutions to the invariance equation (1.2) for all expanding cases and for heaven. In the case of nonexpanding $[\mathrm{III}] \otimes [\mathrm{any}]$ or $[N] \otimes [\mathrm{any}]$ types we have reduced the problem to finding the general solution of one or three, respectively, linear partial differential equations of the first order for two or five, resp., holomorphic functions of two variables.

Unfortunately, we have not succeeded in integrating the Lie equations for the groups of contact transformations admitted by the hyperheavenly equations. Nevertheless, one can easily observe that these groups are too "poor" to generate all solutions of the hyperheavenly equation from some given solution. Most likely the transformation of metric caused by the symmetry group is simply a composition of some coordinate transformation and the conformal transformation with constant conformal factor (compare Ref. 1). Consequently, the general Lie–Bäcklund transformations are expected to be more fruitful at this point. We intend to deal with this problem in the next paper.

The results of the present paper can also be considered as the starting point for the problem of searching for group-invariant solutions to the hyperheavenly equations.

Another problem of a great interest connected with our work is the relation between Killing vectors on hyperheavenly space-times and the infinitesimal operators of symmetry groups admitted by the hyperheavenly equations. Comparing the results of our paper with those of Refs. 4, 5, 13, and 14 one can formulate the following.

Conjecture: Every Killing vector field on a hyperheavenly space-time can be brought to the form $\xi^i(\partial/\partial x^i)$, where ξ^i is defined by the infinitesimal operator (2.19) of some symmetry group admitted by the appropriate hyperheavenly equation. ■

We have not succeeded in verifying the conjecture. The work on this problem is in progress.

[1] M. Przanowski, "The group theoretical analysis of gravitational instanton equations," J. Math. Phys. **31**, 300 (1990).
[2] J. F. Plebański and I. Robinson, Phys. Rev. Lett. **37**, 493 (1976).
[3] J. D. Finley, III and J. F. Plebański, J. Math. Phys. **17**, 2207 (1976).
[4] J. D. Finley, III and J. F. Plebański, J. Math. Phys. **19**, 760 (1978).
[5] C. P. Boyer, J. D. Finley, III, and J. F. Plebański, *Complex General Relativity, $\mathscr{H}$ and $\mathscr{H}\mathscr{H}$ Spaces—A Survey*, Com. Tec. Inst. Inv. Mat. Apl. Sist. UNAM, México 20, D. F., 9 No. 174 (1978). See also C. P. Boyer, J. D. Finley, III, and J. F. Plebański in *General Relativity and Gravitation*, edited by A. Held, (Plenum, New York, 1980), Einstein memorial volume, Vol. 2, pp. 241–281.
[6] J. D. Finley, III and J. F. Plebański, J. Math. Phys. **22**, 667 (1981).
[7] A. García and J. F. Plebański, Nuovo Cimento B **40**, 227 (1977).
[8] P. J. Olver, *Applications of Lie Groups to Differential Equations* (Springer, New York, 1986).
[9] L. V. Ovsiannikov, *Group Analysis of Differential Equations* (Academic, New York, 1982).
[10] R. L. Anderson and N. H. Ibragimov, *Lie-Bäcklund Transformations in Applications* (SIAM, Philadelphia, PA, 1979).
[11] N. H. Ibragimov, *Gruppy Preobrazovanii v Matematicheskoi Fizike* (Nauka, Moscow, 1983) (in Russian).
[12] J. F. Plebański, J. Math. Phys. **16**, 2395 (1975).
[13] Ch. P. Boyer and J. F. Plebański, J. Math. Phys. **18**, 1022 (1977).
[14] J. D. Finley, III and J. F. Plebański, J. Math. Phys. **17**, 585 (1976).

Effects of the shear viscosity on the character of cosmological evolution

Wung-Hong Huang
Department of Physics, National Cheng Kung University, Taiwan 700, Taiwan, Republic of China

(Received 22 November 1988; accepted for publication 4 October 1989)

Bianchi type I cosmological models are studied that contain a stiff fluid with a shear viscosity that is a power function of the energy density, such as $\eta = \alpha\epsilon^n$. These models are analyzed by describing the cosmological evolutions as the trajectories in the phase plane of Hubble functions. The simple and exact equations that determine these flows are obtained when $2n$ is an integer. In particular, it is proved that there is no Einstein initial singularity in the models of $0 \leqslant n < 1$. Cosmologies are found to begin with zero energy density and in the course of evolution the gravitational field will create matter. At the final stage, cosmologies are driven to the isotropic Friedmann universe. It is also pointed out that although the anisotropy will always be smoothed out asymptotically, there are solutions that simultaneously possess nonpositive and non-negative Hubble functions for all time. This means that the cosmological dimensional reduction can work even on matter fluid having shear viscosity. These characteristics can also be found in any-dimensional models.

I. INTRODUCTION

The investigation of relativistic cosmological models usually has the energy momentum tensor of matter as that due to a perfect fluid. To consider a more realistic model one may take into account the dissipative processes that are caused by the viscosity and that have already attracted the attention of many investigators. Misner[1] suggested that the anisotropy in an expanding universe would be smoothed out by the strong dissipative process due to the neutrino viscosity. Viscosity mechanisms in the cosmology can explain the anomalously high entropy of the present universe.[2,3] Bulk viscosity associated with the grand-unified-theory phase transition[4] may lead to an inflationary scenario.[5,6] (The inflationary cosmology invented by Guth[7] in 1981 is used to overcome several important problems arising in the standard big bang cosmology.)

Murphy[8] obtained an exactly soluble isotropic cosmological model of the zero-curvature Friedmann model in the presence of bulk viscosity. The solutions Murphy found exhibit the interesting feature that the big bang type singularity appears in infinite past. Exact solutions of the isotropic homogeneous cosmology for the open, closed, and flat universe were found by Santos *et al.*[9] for when the bulk viscosity is the power function of energy density. However, in some solutions, the big bang singularity of infinite density occurs at finite past. It is thus shown that contrary to the conclusion of Murphy, the introduction of bulk viscosity cannot avoid the initial singularity in general. Anisotropic models with bulk viscosity, which is the power function of energy density, have been discussed in detail in our previous papers.[10,11]

Belinskii and Khalatnikov[12] presented a qualitative analysis about Bianchi type I cosmological models under the influence of shear viscosity: They then found a remarkable property that near the initial singularity the gravitational field creates matter. Recently, Banerjee *et al.*[13] obtained some Bianchi type I solutions for the case of stiff matter by using the assumption that shear viscosity (η) is the power function of the energy density (ϵ), i.e., $\eta = \alpha\epsilon^n$, where α is a constant. However, Banerjee *et al.* merely analyzed the behavior of the cosmological models for some values of n.

In this paper we will investigate the cosmological models again. Since we study these models by describing the evolution of cosmologies as the flow in the phase plane of Hubble functions, we can clarify the property of the models with any value of n. In particular, we prove that there is no Einstein initial singularity in the models with $0 \leqslant n < 1$. The cosmologies have zero energy density in the initial phase and then the shear viscosity causes the gravitational field to create matter during the evolution. At the final stage, cosmologies are driven to the isotropic Friedmann universe. Although the anisotropy in the universe is smoothed out asymptotically, we point out that there are solutions that simultaneously possess nonpositive and non-negative Hubble functions for all time. In view of this fact, we then consider the higher-dimensional theory and show that the cosmological dimensional reduction may work in cosmological models which have shear viscosity. The models extended to higher dimensions have also been analyzed according to the methods described in the present paper; the results show that they all share the same characters.

The organization of this paper is as follows. In Sec. II the derivation of two dynamical evolution equations of expansion and shear scalars, which was presented in Ref. 13, is summarized for a convenient reading. In Sec. III we consider the axially symmetric Bianchi type I models in which there are only two Hubble functions. The evolutions of the cosmology are described as the flows in the phase space of the Hubble functions. The characteristics of the evolutions of the cosmological models for any n are then clarified. The Bianchi type I models with three Hubble functions are discussed in Sec. IV, where we also consider the higher-dimensional models. Section V is devoted to a summary. In the Appendix the solutions in the early stage for the $n \geqslant 1$ models, which explicitly show how the energy density approaches zero at the initial singularity, are given.

II. THE EINSTEIN FIELD EQUATION

We consider the $(1 + D)$-dimensional Bianchi type I

 0022-2488/90/030659-05$03.00

space-time with the line element

$$ds^2 = -dt^2 + \sum_{i=1}^{D} a_i^2(t)\, dx_i^2. \tag{2.1}$$

The energy-momentum tensor for a fluid with shear viscosity is

$$T_{\mu\nu} = (\epsilon + \bar{p}) u_\mu u_\nu + \bar{p} g_{\mu\nu} - \eta \kappa_{\mu\nu}, \tag{2.2}$$

$$\bar{p} \equiv p + (2/D)\eta u^\lambda{}_{;\lambda}, \tag{2.3}$$

$$\kappa_{\mu\nu} \equiv u_{\mu;\nu} + u_{\nu;\mu} + u_\mu u^\lambda u_{\nu;\lambda} + u_\nu u^\lambda u_{\mu;\lambda}, \tag{2.4}$$

where ϵ is the energy density, p is the pressure, and η is the shear viscosity, respectively. Choosing a comoving frame, where $u^\mu = \delta^\mu{}_0$, the explicit form of the Einstein equations is

$$R_{\mu\nu} - \tfrac{1}{2} g_{\mu\nu} R = T_{\mu\nu} \tag{2.5}$$

for the metric (2.1) and the energy-momentum tensor (2.2) can be written as

$$[(D-1)/2D]\, W^2 - \sigma^2 = \epsilon, \tag{2.6}$$

$$\frac{dH_i}{dt} + H_i W - \frac{1}{2}\left[2\frac{dW}{dt} + \left(1 + \frac{1}{D}\right) W^2 + 2\sigma^2\right]$$
$$= p + \left(\frac{2}{D}\eta\right) W - 2\eta H_i, \tag{2.7}$$

where the Hubble functions H_i, the expansion scalar W, and the shear scalar σ^2 are defined by

$$H_i \equiv \frac{1}{a_i}\frac{da_i}{dt}, \quad W \equiv \sum_i H_i, \quad 2\sigma^2 \equiv \sum_i H_i^2 - \frac{1}{D} W^2. \tag{2.8}$$

The trace part of Eq. (2.5) leads to

$$2\frac{dW}{dt} + \left(1 + \frac{1}{D}\right) W^2 + 2\sigma^2 = \frac{2}{(D-1)}[\epsilon - Dp]. \tag{2.9}$$

Using relation (2.6) to eliminate the σ^2 in Eq. (2.9) we obtain

$$\frac{dW}{dt} + W^2 = \frac{D}{D-1}(\epsilon - p). \tag{2.10}$$

As a consequence of the Bianchi identity we have

$$\frac{d\epsilon}{dt} + (\epsilon + p) W - 4\eta\sigma^2 = 0. \tag{2.11}$$

Equation (2.6) can yield a relation

$$\frac{d(\sigma^2/W^2)}{dt} = -\frac{d(\epsilon/W^2)}{dt}. \tag{2.12}$$

After substituting the expressions $d\epsilon/dt$ and dW/dt in Eqs. (2.10) and (2.11) into the rhs of Eq. (2.12) we finally obtain the evolutional equation of a shear scalar:

$$\frac{d(\sigma^2/W^2)}{dt} = -\left(\frac{\sigma^2}{W^2}\right)\left[\frac{2D}{D-1}\left(\frac{\epsilon - p}{W}\right) + 4\eta\right]. \tag{2.13}$$

When the universe is filled with stiff matter, i.e., $\epsilon = p$, we can, from Eq. (2.10), find the solution of an expansion scalar:

$$W = 1/t. \tag{2.14}$$

Using relation (2.14), Eq. (2.13) can be written as

$$\frac{dy}{y[(D-1)/2D - y]^n} = -4\alpha t^{-2n}\, dt, \quad y \equiv \frac{\sigma^2}{W^2}. \tag{2.15}$$

Equation (2.15) can be integrated exactly if the $2n$ is an integer, as has been shown by Banerjee *et al.*[13] However, one can only analyze the models for some values of n as a result of the fact that these integrated relations are too complex. To overcome these difficulties, in Sec. III we will analyze the cosmological evolutions by describing them as flows in the phase plane of Hubble functions.

III. SOLUTIONS IN THE PHASE PLANE

We first consider the axially symmetric Bianchi type I model in the $1 + 3$ dimension; the Hubble functions therein are denoted as

$$H_1 \equiv h, \quad H_2 = H_3 \equiv H, \tag{3.1}$$

in terms of which one can, from Eqs. (2.6) and (2.8), obtain

$$W = h + 2H, \quad \sigma^2 = \tfrac{1}{3}(h - H)^2, \quad \epsilon = H(2h + H). \tag{3.2}$$

We express h and H in terms of the variables r and θ:

$$h = r \sin\theta, \quad H = r\cos\theta. \tag{3.3}$$

Thus

$$W = r(\sin\theta + 2\cos\theta),$$
$$y = \tfrac{1}{3}[(1 - \tan\theta)^2/(2 + \tan\theta)^2], \tag{3.4}$$

and Eq. (2.15) can be written as

$$\int \frac{dy}{y(\frac{1}{3} - y)^n} = -4\alpha \int t^{-2n}\, dt$$
$$= \frac{4\alpha}{2n-1} t^{1-2n} + C$$
$$= \frac{4\alpha}{2n-1} W^{2n-1} + C$$
$$= \frac{4\alpha}{2n-1} r^{2n-1}(\sin\theta + 2\cos\theta)^{2n-1} + C, \tag{3.5}$$

where C is an integration constant. Because y is the function of the variable θ only, Eq. (3.5) tells us that we can express r as the function of the variable θ. It is through this property that one can study the cosmological evolutions by analyzing these trajectories in the phase plane.

A. Evolutions of the cosmologies

1. Fixed point

We only consider the plane where the trajectories are in the regions that satisfy the dominant energy condition,[14] i.e., $\epsilon \geqslant 0$. The evolutions of the cosmology are therefore confined to the regions $H \geqslant 0$ and $H + 2h \geqslant 0$, as Eq. (2.6) shows. The evolutions of the cosmology should start from a fixed point or infinity and then end in another fixed point or infinity in the phase plane. From Eq. (2.13) we know that the fixed points should be those with zero energy density, which are at $H = 0$ or $H + 2h = 0$.

Furthermore, from Eq. (2.13) we can obtain

$$\frac{d\sigma^2/dt}{\sigma^2} = -2W - 4\eta. \tag{3.6a}$$

Since $\eta \geqslant 0$ and $W = 1/t$, we then obtain a relation

$$\sigma^2 \leqslant t^{-2}. \tag{3.6b}$$

Equation (3.6b) tells us that the shear scalar σ^2 should be a decreasing function and that anisotropy will be smoothed out asymptotically. Therefore, the original point is an attractive fixed point.

2. Invariant lines

Equation (2.15) can lead to

$$\frac{dy}{dt} = -4\alpha W^{2n} y \left(\frac{1}{3} - y\right)^n. \tag{3.7}$$

Substituting relation (3.4) into Eq. (3.7) we then obtain

$$\frac{d\theta}{dt} = \frac{2\alpha}{3} r^{2n} (\cos\theta - \sin\theta)(2\cos\theta + \sin\theta) \times [(\cos\theta + 2\sin\theta)\cos\theta]^n. \tag{3.8}$$

From the theory of nonlinear differential equations we know that the zeros of Eq. (3.8) give invariant lines in the phase plane of $h \times H$. They are as follows.

(i) The isotropic state is $\cos\theta - \sin\theta = 0$. In practice, this shall only refer to the original point, resulting from the fact that there is no shear viscosity in the isotropic state. Furthermore, since the shear scalar is a decreasing function, the original point must be the attractive state.

(ii) The states $2\cos\theta + \sin\theta = 0$ have negative energy density and are thus neglected.

(iii) The states $\cos\theta = 0$ or $\cos\theta + 2\sin\theta = 0$ have zero energy density. Because the shear scalar is a decreasing function, the vacuum states must be the initial states.

Using results (i)–(iii) and the exact solution of the $n = 0$ model (note that we can obtain the exact solution for $2n$ = integer models) we therefore conclude that the cosmologies should begin with zero energy density in the initial phase; then the shear viscosity causes the gravitational field to create matter during the evolution; and at the final stage, cosmologies are driven to the isotropic Friedmann universe.

B. Singularity

Equation (3.2) tells us that the energy density can become infinity only if H and/or h are infinite. Equation (2.9) tells us that dW/dt can become infinity only if H and/or h are infinite. Since the Riemann scalar curvature can be expressed as

$$R = 2\frac{dW}{dt} + W^2 + 2H^2 + h^2 \tag{3.9}$$

one then sees that R can become infinity only if H and/or h are infinite. Therefore, the Einstein initial singularity can arise only if H and/or h are infinity, i.e., $r \to \infty$.

From Eq. (3.5) we know that r can become infinity only if the integrated value on the lhs is infinity. Since in this section we have discussed that only those states with zero energy density may be the infinite value of r, we therefore need only to consider the vacuum states.

When $y \to \frac{1}{3}$, then

$$\int \frac{dy}{y(\frac{1}{3}/3 - y)^n} \to 3 \int^{y \to 1/3} \frac{dy}{(\frac{1}{3} - y)^n}. \tag{3.10}$$

The integration (3.10) is infinite only if $n \geqslant 1$, as easily seen. Thus we have shown that models with $0 \leqslant n \leqslant 1$ can avoid the Einstein initial singularity.

C. Examples

To give some examples we present the following explicit solutions.

For $n = \frac{1}{2}$,

$$[r(\sin\theta + 2\cos\theta)]^{4\alpha/\sqrt{3}} = C\frac{\sqrt{3(1 + 2\tan\theta)} - (2 + \tan\theta)}{\sqrt{3(1 + 2\tan\theta)} + (2 + \tan\theta)}, \tag{3.11}$$

where C is an integration constant. Using Eq. (3.11) we can plot the trajectories in the phase plane; they then determine the cosmological evolutions. One see that r is finite when $\cos\theta = 0$ or $1 + 2\tan\theta = 0$; thus this model has no initial singularity.

For $n = 1$,

$$\ln\frac{(1 - \tan\theta)^2}{(1 + 2\tan\theta)} = \frac{4\alpha}{3} r(\sin\theta + 2\cos\theta) + C, \tag{3.12}$$

where C is an integration constant. Using Eq. (3.12) we can plot the trajectories in the phase plane; they then determine the cosmological evolutions. One see that r is infinite when $\cos\theta = 0$ or $1 + 2\tan\theta = 0$; thus this model begins with an initial singularity.

IV. MORE GENERAL SOLUTIONS

We now consider the Bianchi type I models with multiple Hubble functions. From Eq. (2.7) we can obtain

$$\frac{d\ln(H_i - H_j)}{dt} = \frac{d\ln(H_i - H_k)}{dt}. \tag{4.1}$$

Equation (4.1) yields the solutions

$$H_i = (1 - C_i)H_1 + C_i H_2, \quad i = 3,...,D, \tag{4.2}$$

where C_i are the integration constants. Relation (4.2) tells us that one can express all other D-2 Hubble functions in terms of only two Hubble functions. Through the same procedures as those described in Sec. III we find that the evolutions of the 1 + D-dimensional Bianchi type I cosmological models containing stiff fluid with shear viscosity as a power function of the energy can also be expressed as flows in the phase plane of $H_1 \times H_2$. One can also show that the characteristics of the cosmological evolutions are just those described in Sec. III, i.e., cosmologies will begin with zero energy density; in the course of evolution the gravitational field will create matter; and finally, cosmologies are driven to the isotropic Friedmann universe. In the same way, we can also prove that there is no Einstein initial singularity in models with $0 \leqslant n < 1$.

To give an illustration we consider the (1 + 3)-dimensional model with three Hubble functions. We now have the relations

$$H_3 = (1 - C)H_1 + CH_2, \quad W = (2 - C)H_1 + (1 + C)H_2,$$
$$\sigma^2 = [(1 - C + C^2)/3](H_1 - H_2)^2. \tag{4.3}$$

Expressing H_1 and H_2 in terms of the variables r and θ,

$$H_1 = r\cos\theta,\quad H_2 = r\sin\theta. \tag{4.4}$$

Then Eq. (2.15) can lead to

$$\frac{d\theta}{dt} = \frac{2\alpha}{3} r^{2n}(1 - C + C^2) \times (\cos\theta - \sin\theta)[(2 - C)\cos\theta + (1 + C)\sin\theta] \times [(1 - C)\cos^2\theta + 2\sin\theta\cos\theta + C\sin^2\theta]^n. \tag{4.5}$$

The zeros of Eq. (4.5) give invariant lines in the phase plane. They are as follows.

(i) The isotropic state is $\cos\theta - \sin\theta = 0$. As discussed in Sec. III, this shall only refer to the original point, which corresponds to the final state.

(ii) The states $(2 - C)\cos\theta + (1 + C)\sin\theta = 0$ can be shown to have negative energy density and are thus neglected.

(iii) The states $(1 - c)\cos^2\theta + 2\cos\theta\sin\theta + C\times\sin^2\theta = 0$ are the vacuum states. Because shear scalar is a decreasing function, these states must be the initial states.

Results (i)–(iii) show that the cosmologies will begin with zero energy density and then end in the isotropic Friedmann universe. One can also use the method described in Sec. III to prove that models with $0 \leqslant n < 1$ have no Einstein initial singularity.

Finally, we want to mention that although the anisotropy is smoothed out asymptotically [as Eq. (3.6b) shows] and cosmologies will be driven to the isotropic Friedmann universe eventually, there indeed are solutions that possess both nonpositive and non-negative Hubble functions. This implies that the cosmological dimensional reduction can work in cosmological models which have shear viscosity. This property has been found in our previous paper.[15] Here we show an example in order to complete our discussions of cosmological models.

Consider the $n = 1$ five-dimensional theory with two Hubble functions of h and H corresponding to those of three-space and extra space, respectively. Using Eqs. (2.14) and (2.15) we obtain

$$W = t^{-1},\quad \sigma^2 = \tfrac{3}{8}(1/t^2)[1 + Ce^{-3\alpha/2t}]^{-1}, \tag{4.6}$$

where C is a positive constant. With the definitions of W and σ^2 in Eq. (2.8) we then obtain

$$\begin{aligned} h &= (1/4t)(1 + [1 + C\exp(-3\alpha/2t)]^{-1/2}),\\ H &= (1/4t)(1 - 3[1 + C\exp(-3\alpha/2t)]^{-1/2}), \end{aligned} \tag{4.7}$$

$$\begin{aligned} h &= (1/4t)(1 - [1 + C\exp(-3\alpha/2t)]^{-1/2}),\\ H &= (1/4t)(1 + 3[1 + C\exp(-3\alpha/2t)]^{-1/2}). \end{aligned} \tag{4.8}$$

Solution (4.7), in which h is positive while H is negative for all time, can be found when $C < 8$.

We have also checked that the simultaneous existence of nonpositive and non-negative Hubble functions is found in other (including 1 + 3) space-time models with other values of n.

Thus we have clarified the effects of shear viscosity on the characteristics of cosmological evolution.

V. CONCLUSIONS

In this paper we have discussed Bianchi type I cosmological models with a viscous fluid, assuming that the shear viscosity is a power function of the energy density, such as $\eta = \alpha\epsilon^n$. We have presented a detailed study of these models by describing the evolutions of cosmologies as the flows in the phase plane of Hubble functions. We have clarified the property of these models with any value of n. In particular, we have proved that there are no Einstein initial singularities in models with $0 \leqslant n < 1$. The cosmologies have been found to begin with zero energy density; then the shear viscosity causes the gravitational field to create matter during the evolution. At the final stage, cosmologies are driven to the isotropic Friedmann universe. We have also pointed out that there are solutions that possess both nonpositive and non-negative Hubble functions for all time. In view of this fact, we then considered five-dimensional theory and gave a solution that explicitly showed that the cosmological dimensional reduction can work in cosmological models which have shear viscosity. Models extended to other dimensions can also be analyzed according to the procedures described in this paper; the results show that they all share the same characteristics.

The models considered in this paper are only concerned with stiff matter. The same models with other matter fields are certainly of interest and remain to be studied.

APPENDIX: SOLUTION IN THE EARLY STAGE

Since the behavior of the vanishing energy density at the initial singularity, which is argued in Sec. III to be a common property belonging to $n \geqslant 1$ models, seems unusual, we now give the solution in the early stage to explicitly show how the energy density approaches zero asymptotically.

Let $\epsilon = p$ and $D = 3$ in Eqs. (2.7) and (2.9). We obtain

$$\frac{dH_i}{dt} + H_i W = \frac{2}{3}\eta W - 2\eta H_i. \tag{A1}$$

From the discussions given in Sec. III we know that $n \geqslant 1$ models will begin along the invariant lines $H = 0$ or $2h + H = 0$. Thus from Eq. (A1) we find the approximate solution as $H \to 0$. (The case of $2h + H \to 0$ could be analyzed in the same way and will give the same conclusion.)

To be consistent with Eq. (2.14), $W = h + 2H = t^{-1}$, one can define

$$h = t^{-1} - \delta h, \tag{A2a}$$

$$H = \delta h/2, \tag{A2b}$$

where $0 \leqslant \delta h \ll 1$. Using the relation of Eq. (3.2), $\epsilon = H(2h + H)$, implies

$$\epsilon \cong t^{-1}\delta h. \tag{A3}$$

Let $\eta = \alpha\epsilon^n$; we then approximate Eq. (A1) as

$$\frac{\delta h}{dt} + t^{-1}\delta h \cong \frac{4\alpha}{3} t^{-1}\left(\frac{\delta h}{t}\right)^n. \tag{A4}$$

The solution for the $n = 1$ model is

$$\delta h \cong Ct^{-1}\exp(-4\alpha/3t), \tag{A5a}$$

$$\epsilon \cong Ct^{-2}\exp(-4\alpha/3t), \tag{A5b}$$

where C is an integration constant. Solution (A5) is consistent with the asymptotic form of the exact solution found in Eq. (3.9) of Ref. 13. The solution for the $n > 1$ model is

$$\delta h \cong [4\alpha(n-1)/3(2n-1)]^{1/(1-n)} t^{n/(n-1)}, \tag{A6a}$$

$$\epsilon \cong [4\alpha(n-1)/3(2n-1)]^{1/(1-n)} t^{1/(n-1)}. \tag{A6b}$$

Solutions (A5) and (A6) explicitly show that $n \geqslant 1$ models have vanishing energy density at the initial singularity.

[1]C. M. Misner, Nature **214**, 40 (1967); Astrophys. J. **151**, 431 (1967).
[2]S. Weinberg, Astrophys. J. **168**, 175 (1972).
[3]S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).
[4]P. Langacker, Phys. Rep. **72**, 185 (1981).
[5]I. Waga, R. C. Falcao, and R. Chanda, Phys. Rev. D **33**, 1839 (1986).
[6]T. Pacher, J. A. Stein-Schabes, and M. S. Turner, Phys. Rev. D **36**, 1603 (1987).
[7]A. H. Guth, Phys. Rev. D **23**, 347 (1981).
[8]G. L. Murphy, Phys. Rev. D **8**, 4231 (1973).
[9]N. O. Santos, R. S. Dias, and A. Banerjee, J. Math. Phys. **26**, 878 (1985).
[10]W.-H. Huang, Phys. Lett. A **129**, 429 (1988).
[11]W.-H. Huang, J. Math. Phys., submitted.
[12]V. A. Belinskii and V. M. Khalatnikov, Sov. Phys. JETP **42**, 205 (1976); Sov. Phys. JETP **45**, 1 (1977).
[13]A. Banerjee, S. B. Duttachoudhury, and A. K. Sanyal, J. Math. Phys. **26**, 3010 (1985).
[14]S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Spacetime* (Cambridge U.P., Cambridge, 1973).
[15]W.-H. Huang, Phys. Lett. A **136**, 21 (1989).

Interaction of null dust clouds fronted by impulsive plane gravitational waves

A. H. Taub
Department of Mathematics, University of California, Berkeley, California 94720

(Received 16 October 1989; accepted for publication 29 November 1989)

This paper discusses the energy-momentum tensor $T^{\mu\nu}$ in the region of interaction of a space-time in which two colliding plane impulsive gravitational waves, each followed by a null dust cloud, exist. It is shown that in the interaction region $T^{\mu\nu}$ is of three types: (i) that of two noninteracting null dusts; (ii) that of a scalar field (equivalently an irrotational perfect fluid with energy density equal to pressure), and (iii) that of the sum of two independent non-interacting scalar fields [equivalently a complex scalar field or an anisotropic perfect fluid with energy density w and pressures (w,π,π)].

I. INTRODUCTION

The evolution of a space-time V containing two colliding plane impulsive gravitational waves whose leading edges are followed by distributions of null dust will be discussed under the following four assumptions:

(i) V admits two commuting spacelike Killing vectors ξ^{μ}_{A} $(A=1,2;\mu=0,1,2,3)$.

The coordinate system may be chosen so that $\xi^{\mu}_{A}=\delta^{\mu}_{A}$ and the metric of V may be put into the Rosen form[1]:

$$ds^2=g_{ij}\,dx^i\,dx^j+g_{AB}\,dx^A\,dx^B, \tag{1.1}$$

where

$$\alpha,\beta=0,1,2,3,\quad i,j=0,3,\quad A,B=1,2, \tag{1.2}$$

$$g_{\alpha\beta}\delta^{\alpha}_{i}\delta^{\beta}_{j}=g_{ij}=e^{\omega}\eta_{ij}=e^{\omega}(\delta_{ij}-2\delta^{3}_{i}\delta^{3}_{j}), \tag{1.3}$$

$$g_{\alpha\beta}\delta^{\alpha}_{A}\delta^{\beta}_{B}=g_{AB}=-e^{\mu}\gamma_{AB}, \tag{1.4}$$

$$\gamma_{11}=\chi^{-1},\quad \gamma_{12}=-q_2\chi^{-1},\quad \gamma_{22}=\chi+q_2^2/\chi, \tag{1.5}$$

$$-{}^4g=-\det\|g_{\alpha\beta}\|=e^{2(\omega+\mu)}. \tag{1.6}$$

The quantities ω, μ, χ, q_2 are functions of x^0 and x^3 alone: They may also be considered as functions of the null coordinates

$$u=x^0-x^3, \tag{1.7}$$

$$v=x^0+x^3. \tag{1.8}$$

The null hypersurfaces $u=0$ and $v=0$ will be used to divide V into four subregions: region I, where $u>0$ and $v>0$; region II, where $u>0$ and $v<0$; region III, where $u<0$ and $v>0$; and region IV, where $u<0$ and $v<0$.

The hypersurface $u=0$ $(v=0)$ will be interpreted as the wave front of a gravitational wave traveling in the $x^3(-x^3)$ direction. The variables

$$2x^0=(v+u) \tag{1.9}$$

and

$$2x^3=(v-u) \tag{1.10}$$

are measures of the time from $(u=0,\,v=0)$ the instant of collision and the distance between the wave fronts, respectively.

(ii) When $g_{\mu\nu}(u,v)$ is an exact solution of the Einstein field equations holding in region I, the metric tensor in region II (III) is given by $g_{\mu\nu}(u,0)(g_{\mu\nu}(0,v))$. The metric tensor in region IV is $g_{\mu\nu}(0,0)$, so that region IV is flat.

This method of extending the solution of region I produces metrics that are continuous across the hypersurfaces $u=0$ and $v=0$, but may have discontinuous first derivatives across these hypersurfaces. If so the curvature tensor derived from $g_{\alpha\beta}$ will be distribution valued, i.e., it will contain delta functions with support on these null hypersurfaces.

In addition, space-times with metrics obtained as above are said to contain impulsive gravitational waves only if the components of the Einstein tensor (equivalently the Ricci tensor) do not contain such delta functions.

It is a consequence of assumption (ii) that in region II (III) the only nonvanishing component of the Ricci tensor can be R_{uu} (R_{vv}). Thus regions II and III are either vacuum regions or contain null dust, i.e., a medium with a stress-energy tensor of the form

$$T_{\mu\nu}=Eu_{,\mu}u_{,\nu}\quad (Ev_{,\mu}v_{,\nu})$$

in region II (III), $f_{,\mu}=\partial f/\partial x^{\mu}$. In addition, we have

$$R_{AB}=R_{\mu\nu}\delta^{\mu}_{A}\delta^{\nu}_{B}=0 \tag{1.11}$$

in regions II, III, and IV.

(iii) It is assumed that $R_{AB}=0$ in region I, that is, Eq. (1.11) is assumed to hold everywhere in V.

(iv) It is assumed that the components of $R_{\mu\nu}$ do not involve delta functions.

It has been shown in an earlier paper[2] that assumption (iv) implies that $\mu_{,u}=0$ $(\mu_{,v}=0)$ on the hypersurface $u=0$ $(v=0)$.

It is the purpose of this paper to determine the various possible energy-momentum tensors that can occur in region I, the interaction region, under assumptions (i)–(iv). We shall also discuss the relation between the media in regions II and III with the energy-momentum tensor in region I.

II. THE RICCI TENSOR

As a consequence of the fact that the coordinates of space-time may be chosen so that the vectors δ^{μ}_{1} and δ^{μ}_{2} are Killing vectors, the line element of space-time is given by Eq. (1.1), where the g_{ij} need not satisfy Eq. (1.3). It is a further consequence that the components of the Ricci tensor are such that

 0022-2488/90/030664-05$03.00

$$R^{i}_{A}=0 \tag{2.1}$$

and

$$2R^{C}_{A}=(e^{-\mu}/\sqrt{-g})(e^{\mu}\sqrt{-g}g^{ij}g^{CB}g_{CA,i})_{,j} \tag{2.2}$$

when

$$g=\det\|g_{ij}\|. \tag{2.3}$$

In case Eqs. (1.3) obtain

$$-g=e^{\omega}, \tag{2.4}$$

$$R_{uu}=\mu_{,uu}+\tfrac{1}{2}\mu_{,u}^{2}-\mu_{,u}\omega_{,u}-S_{uu}, \tag{2.5}$$

$$R_{uv}=\mu_{,uv}+\tfrac{1}{2}\mu_{,u}+\mu_{,vv}\omega_{,uv}-S_{uv}, \tag{2.6}$$

$$R_{vv}=\mu_{,vv}+\tfrac{1}{2}\mu_{,v}^{2}-\mu_{,v}\omega_{,v}-S_{uv}, \tag{2.7}$$

where

$$S_{ij}=\tfrac{1}{4}\gamma^{AB}_{,i}\gamma_{AB,j}$$
$$=-(1/2\chi^{2})(\chi_{,i}\chi_{,j}+q_{2,i}q_{2,j}), \tag{2.8}$$

γ_{AB} is given by Eq. (1.5) and γ^{AB} is such that

$$\gamma^{AB}\gamma_{BC}=\delta^{A}_{C}, \tag{2.9}$$

that is

$$\gamma^{11}=\chi+q_2^{2}/\chi,\quad \gamma^{12}=q_2/\chi,\quad \gamma^{22}=1/\chi. \tag{2.10}$$

It follows from Eqs. (2.2) and (2.4) that

$$2R^{C}_{C}=e^{-(\mu+\omega)}(e^{\mu}\eta^{ij})_{,ij}=e^{-\omega}(\mu_{,ij}+\mu_{,i}\mu_{,j})\eta^{ij} \tag{2.11}$$

and

$$2R^{A}_{B}=2R^{C}_{C}\delta^{A}_{B}+e^{-(\mu+\omega)}(e^{\mu}\gamma^{CA}\gamma_{BC,i}\eta^{ij})_{,j}. \tag{2.12}$$

The equations $R_{AB}=0$ then imply that

$$(e^{\mu})_{,uv}=e^{\mu}(\mu_{,uv}+\mu_{,u}\mu_{,v})=0 \tag{2.13}$$

and

$$(e^{\mu}\gamma_{,u}\gamma^{-1})_{,v}+(e^{\mu}\gamma_{,v}\gamma^{-1})_{,u}=0, \tag{2.14}$$

where

$$\gamma=\|\gamma_{AB}\|. \tag{2.15}$$

In view of Eqs. (1.5) and (2.10), Eq. (2.12) may be written as

$$(e^{\mu}q_{2,u}/\chi^{2})_{,v}+(e^{\mu}q_{2,v}/\chi^{2})_{,u}=0 \tag{2.16}$$

and

$$(e^{\mu}\chi_{,u}/\chi)_{,v}+(e^{\mu}\chi_{,v}/\chi)_{,u}=-2(e^{\mu}/\chi^{2})(q_{2,u}q_{2,v}). \tag{2.17}$$

In the coordinate system in which Eqs. (1.1) and (1.3) hold, the Einstein field equations supplemented by the condition $R_{AB}=0$ consist of the system of equations (2.5)–(2.7) supplemented by Eqs. (2.13), (2.16), and (2.17): In these equations the R_{ij} are assumed to be given as functions of u and v. These quantities must, of course, satisfy the integrability conditions for the determination of ω from Eqs. (2.5)–(2.7).

Note that if μ and the matrix γ are solutions of Eqs. (2.13) and (2.14) [equivalently (2.13), (2.16), and (2.18)] and ω^{0} is a solution of Eqs. (2.5)–(2.7) with $R_{ij}=0$, i.e., ω^{0} is a solution of the vacuum equations, then

$$\omega=\Omega+\omega^{0} \tag{2.18}$$

satisfies the latter equations with $R_{ij}\neq 0$ if Ω is a solution of

$$R_{uu}=-\mu_{,u}\Omega_{,u}, \tag{2.19}$$

$$R_{uv}=\Omega_{,uv}, \tag{2.20}$$

$$R_{vv}=-\mu_{,v}\Omega_{,v}, \tag{2.21}$$

with

$$e^{\mu}=1+U(u)+V(v). \tag{2.22}$$

Equations (2.19)–(2.22) have been shown to hold in the case of the subspace of V generated by the coordinates x^1 and x^2 and hence V admits the group of motions of a two-dimensional plane (cf. Ref. 3).

When μ, the matrix γ, and ω are determined as stated above, the line element in region I of V is given by Eq. (1.1). When the line element is extended to regions II–IV as described in Sec. I one finds that Eqs. (2.19)–(2.22) reduce to

$$R^{\mathrm{II}}uu=-\mu_{,u}\Omega_{,u}(u,0), \tag{2.23a}$$

$$R^{\mathrm{II}}uv=R^{\mathrm{II}}_{vv}=0, \tag{2.23b}$$

$$\mu^{\mathrm{II}}=\ln(1+U(u)) \tag{2.23c}$$

in region II.

In region III one finds

$$R^{\mathrm{III}}uu=R^{\mathrm{III}}uv=0, \tag{2.24a}$$

$$R^{\mathrm{III}}vv=-\mu_{,v}\Omega_{,v}(0,v), \tag{2.24b}$$

$$\mu^{III}=\ln(1+V(v)). \tag{2.24c}$$

In region IV one finds

$$R^{\mathrm{IV}}ij=0, \tag{2.25a}$$

$$\mu^{\mathrm{IV}}=0. \tag{2.25b}$$

The integrability conditions of Eqs. (2.19)–(2.21) are

$$((\mu_{,u})^{-1}R_{uu})_{,v}=((\mu_{,v})^{-1}R_{vv})_{,u}=-R_{uv}. \tag{2.26}$$

Equations (2.26) are equivalent to the equations $T^{\mu\nu}_{;\nu}=0$, where $T^{\mu\nu}$ is the stress-energy tensor of the medium in region I of V formed from the Ricci tensor $R_{\mu\nu}$.

III. THE ENERGY-MOMENTUM TENSOR IN REGION I

As was pointed out in Sec. I, regions II and III contain null dusts. If R^{II}_{ij} and R^{III}_{ij} are prescribed, then $\Omega^{\mathrm{I}}(u,0)$ and $\Omega^{\mathrm{I}}(0,v)$ are determined. However, this information alone is not sufficient to determine $\Omega^{\mathrm{I}}(u,v)$ uniquely in the general case. In other words, just the specification of the energy density in the null clouds that follow the colliding plane impulsive gravitational waves is not sufficient information to determine uniquely the metric of the space-time in the interaction region (that is, the evolution of the space-time after the moment of collision).

In Refs. 4 and 5 Chandrasekhar and Xanthopolous determined two different exact solutions of the Einstein field equations, each of which described two different space-times. Each of these solutions contained colliding impulsive plane gravitational waves followed by distributions of null dust, but the two space-times differed in the properties of the energy-momentum tensor in their interaction regions, regions I.

In this section we shall determine the algebraic structure of the Einstein tensor of a space-time with a metric given by Eq. (1.1) and for which assumptions (i)–(iv) inclusive obtain. Let

$$-R_{uu}=\epsilon_1, \tag{3.1a}$$

$$-R_{vv}=\epsilon_2, \tag{3.1b}$$

$$-R_{uv}=-R_{vu}=C. \tag{3.1c}$$

Then since $R_{AB}=0$,

$$\kappa T_{ij}=-G_{ij}=-(R_{ij}-\tfrac{1}{2}Rg_{ij})=\epsilon_1 u_{,i}u_{,j}+\epsilon_2 v_{,i}v_{,j}, \tag{3.2a}$$

$$\kappa T_{AB}=-G_{AB}=-(R_{AB}-\tfrac{1}{2}Rg_{AB})=-2Ce^{-\omega}g_{AB}, \tag{3.2b}$$

as follows from the fact that

$$u_{,i}v_{,j}+u_{,j}v_{,i}=g_{ij}(u^k v_{,k}), \tag{3.3a}$$

$$u^k v_{,k}=2e^{-\omega}. \tag{3.3b}$$

The notation used above and subsequently in this section is that of Ref. 2.

It may be verified that Eqs. (3.2) may be written as

$$\kappa T^{\mu\nu}=(w+\pi)U^\mu U^\nu+(p-\pi)Z^\mu Z^\nu-\pi g^{\mu\nu}, \tag{3.4}$$

where

$$g^{\mu\nu}=U^\mu U^\nu-Z^\mu Z^\nu+g^{AB}\delta^\mu_A\delta^\nu_B, \tag{3.5a}$$

$$U^\mu U_\mu=-Z^\mu Z_\mu=1, \tag{3.5b}$$

$$U^\mu=\alpha(\sqrt{\epsilon_1}u^\mu+\sqrt{\epsilon_2}v^\mu), \tag{3.5c}$$

$$Z^\mu=\alpha(\sqrt{\epsilon_1}u^\mu-\sqrt{\epsilon_2}v^\mu), \tag{3.5d}$$

$$p=w=2e^{-\omega}\sqrt{\epsilon_1\epsilon_2} \tag{3.5e}$$

$$\pi=2e^{-\omega}C=-R/2, \tag{3.5f}$$

$$\alpha^{-2}=2w. \tag{3.5g}$$

Equation (3.4) may be written as

$$\kappa T^{\mu\nu}=wU^\mu U^\nu+pZ^\mu Z^\nu-\pi g^{AB}\delta^\mu_A\delta^\nu_B, \tag{3.6}$$

as follows from Eq. (3.5a). Thus $\kappa T^{\mu\nu}$ is the energy-momentum tensor of an anisotropic fluid. Multiplying Eq. (3.6) by $g_{\mu\nu}$ and summing we have

$$\kappa T=(w-p)-2\pi;$$

hence

$$-R^{\mu\nu}=\kappa[T^{\mu\nu}-(T/2)g^{\mu\nu}]$$
$$=wU^\mu U^\nu+pZ^\mu Z^\nu-\tfrac{1}{2}(w-p)g^{\mu\nu}.$$

Since $U^A=Z^A=0$ and $R^{AB}=0$, we must have

$$w=p. \tag{3.7}$$

It follows from Eq. (3.6) that the proper values of κT^μ_ν are (w,w,π,π), corresponding to the proper vectors U^μ, Z^μ, Y^μ, X^μ, where U^μ is the velocity vector of the medium and the vectors Z^μ, Y^μ, X^μ are an orthonormal triad spanning the three-space orthogonal to U^μ. One may classify the various types of media in region I, the interaction region, as follows.

(i) Type A, for which $\pi=0=\kappa T$ (and $R_{uv}=0$). Then

$$\kappa T^{\mu\nu}=w(U^\mu U^\nu+Z^\mu Z^\nu). \tag{3.8a}$$

In view of Eqs. (3.5) we have

$$\kappa T_{ij}=\epsilon_1 u_{,i}u_{,j}+\epsilon_2 v_{,i}v_{,j}, \tag{3.8b}$$

$$\kappa T_{AB}=0. \tag{3.8c}$$

(ii) Type B, for which $\pi=w>0$. Then

$$\kappa T^{\mu\nu}=2wU^\mu U^\nu-wg^{\mu\nu}, \tag{3.9a}$$

that is, the medium in the interaction region is a perfect fluid with energy density equal to the pressure. If follows from Eq. (3.9a) that

$$-R_{\mu\nu}=2wU_\mu U_\nu, \tag{3.9b}$$

that is,

$$-R_{ij}=\tau_i\tau_j, \tag{3.9c}$$

$$\tau_i=\sqrt{2w}U_i. \tag{3.9d}$$

(iii) Type C, for which the energy-momentum tensor given by Eqs. (3.6) and (3.7) is such that

$$-w\leqslant\pi\leqslant w. \tag{3.10}$$

This requirement is equivalent to the requirement that for any timelike vector W_μ, $T^{\mu\nu}W_\nu W_\mu\geqslant 0$ and $V^\mu=T^{\mu\nu}W_\nu$ is a nonspacelike vector, that is, $T^{\mu\nu}$ satisfies the dominant energy condition.[6]

Such energy-momentum tensors have been represented by Letelier[7] as the sum of two energy-momentum tensors of type B. This is done as follows. Let

$$\Lambda^\mu=\phi^\mu+i\psi^\mu, \tag{3.11a}$$

$$\phi^\mu=\sqrt{w+\pi}U^\mu\cos\theta+\sqrt{w-\pi}Z^\mu\sin\theta, \tag{3.11b}$$

$$\psi^\mu=-\sqrt{w+\pi}U^\mu\sin\theta+\sqrt{w-\pi}Z^\mu\cos\theta. \tag{3.11c}$$

Then

$$\phi^\mu\phi^\nu+\psi^\mu\psi^\nu=(w+\pi)U^\mu U^\nu+(w-\pi)Z^\mu Z^\nu, \tag{3.12a}$$

$$\phi^\mu\phi_\mu+\psi^\mu\psi_\mu=2\pi=\Lambda^\mu\overline{\Lambda}_\mu, \tag{3.12b}$$

$$\phi^\mu\phi_\mu-\psi^\mu\psi_\mu=2w\cos 2\theta, \tag{3.12c}$$

$$\phi^\mu\psi_\mu=-w\sin 2\theta, \tag{3.12d}$$

that is,

$$2w=|\Lambda^\mu\Lambda_\mu|, \tag{3.12e}$$

$$\mathrm{Re}(\Lambda^\mu\Lambda_\mu)\tan 2\theta=-\mathrm{Im}(\Lambda^\mu\Lambda_\mu). \tag{3.12f}$$

It follows from Eqs. (3.4) and (3.7) that

$$-R_{\mu\nu}=\kappa[T_{\mu\nu}-(T/2)g_{\mu\nu}]$$
$$=(w+\pi)U_\mu U_\nu+(w-\pi)Z_\mu Z_\nu.$$

Hence

$$-R_{\mu\nu}=\phi_\mu\phi_\nu+\psi_\nu\psi_\nu, \tag{3.13}$$

as follows from Eq. (3.12a).

IV. $T^{\mu\nu}_{;\nu}=0$ FOR TYPE-A MEDIA

As shown in Ref. 2 there are only two independent equations among the four equations $T^{\mu\nu}_{;\nu}$ when $T^{\mu\nu}$ is the energy-momentum tensor in region I, the interaction region: When $T^{\mu\nu}$ is given by Eqs. (3.2) these equations are

$$2(\epsilon_1 u^\nu)_{;\nu} - \kappa T\mu_{,u} = 0, \tag{4.1a}$$

$$2(\epsilon_2 v^\nu)_{;\nu} - \kappa T\mu_{,v} = 0, \tag{4.1b}$$

where

$$u^\mu = g^{\mu\nu} u_{;\nu}, \quad v^\mu = g^{\mu\nu} v_{;\nu}.$$

Equations (4.1) are equivalent to Eqs. (2.19)–(2.21).

When the medium in region I is of type A, then equations $T^{\mu\nu}_{;\nu} = 0$ imply that

$$(\epsilon_1 u^\nu)_{;\nu} = (\epsilon_2 v^\nu)_{;\nu} = 0 \tag{4.2}$$

since $u^\mu_{;v} u^\nu = u^\mu_{;\nu} v^\nu = 0$, or equivalently that

$$((\mu_{,u})^{-1}\epsilon_1)_{,v} = ((\mu_{,v})^{-1}\epsilon_2)_{,u} = 0, \tag{4.3}$$

as follows from Eq. (2.22).

Thus when the medium in the interaction region is of type A, the energy-momentum tensor of this medium is that of the sum of two noninteracting dusts. Each dust satisfies the conservation equation given by one of Eqs. (4.2). Since $R_{uv} = 0$ in this case it follows from Eq. (2.20) that

$$\Omega^{\mathrm{I}}(u,v) = F(u) + G(v).$$

Equation (4.3) enables one to determine $F(u)$ and $G(v)$ such that $F(0) = G(0) = 0$. Then $\Omega^{\mathrm{I}}(u,v)$ is determined from $\Omega^{\mathrm{I}}(0,v)$ and $\Omega^{\mathrm{I}}(u,0)$. The latter two quantities are determined from R^{II}_{ij} andR^{III}_{ij}.

In Ref. 5 Chandrasekhar and Xanthopolous treated the case for which the medium in region I, the interaction region, is of type A. However, the physical nature of the null dusts in regions II and III was not discussed. It is expected that these dusts differ from those that are involved in the discussion of Ref. 4, where the interaction is assumed to be of type B. The discussion of Sec. V leads to the interpretation of the latter incoming null dusts as being due to scalar fields propagating with the velocity of light.

It will be shown below that the incoming dust involved in the problem treated in Ref. 5 cannot be null scalar fields unless μ is constant: Neither can they be null Maxwell fields, as has been shown in Ref. 8. This fact follows from the fact that assumption (iii) of Sec. I cannot be satisfied in an electrovac space-time admitting two commuting mutually orthogonal spacelike Killing vectors. Reference 5 does assume that assumption (iii) holds.

V. $T^{\mu\nu}_{;\nu}=0$ FOR TYPE-B MEDIA

It has been shown in Ref. 2 that the equations $T^{\mu\nu}_{;\nu}$ are equivalent to Eq. (2.26). When Eq. (3.9c) is substituted into Eqs. (2.26) and use is made of Eq. (2.22) one obtains

$$2\tau_{u,v} = 2\tau_{v,u} = -(\mu_{,u}\tau_v + \mu_{,v}\tau_u).$$

Therefore, there exists a function $\phi(u,v)$ such that

$$\tau_\mu = \phi_{,\mu} \tag{5.1}$$

and

$$2\phi_{,uv} + \mu_{,u}\phi_{,v} + \mu_{,v}\phi_{,u} = 0. \tag{5.2a}$$

Equation (5.2a) is equivalent to

$$\phi_{;\mu\nu} g^{\mu\nu} = 0 \tag{5.2b}$$

or

$$(e^\mu \phi_{,u})_{,v} + (e^\mu \phi_{,v})_{,u} = 0. \tag{5.2c}$$

Type B is characterized by an energy-momentum tensor of a perfect fluid with energy density w equal to the pressure π. Equation (5.1) states that the fluid is irrotational; Eqs. (5.2) state that the velocity potential ϕ satisfies a wave equation. The function ϕ also describes a real scalar massless field. The energy-momentum tensor of such a field is

$$\kappa T_{\mu\nu} = \phi_{,\mu}\phi_{,\nu} - \tfrac{1}{2} g_{\mu\nu}\phi_{,\rho}\phi_{,\sigma} g^{\rho\sigma} \tag{5.3}$$

since

$$2w = \tau_\mu \tau_\nu g^{\mu\nu} = \phi_{,\mu}\phi_{,\nu} g^{\mu\nu} = 2\pi. \tag{5.4}$$

The collision problem discussed by Chandrasekhar and Xanthopolous in Ref. 4 assumes that the interaction region is of type B. The problem may be interpreted as describing the collision of two plane impulsive gravitational waves, each followed by a dust cloud consisting of a massless scalar field depending on the null coordinate u or v alone.

VI. $T^{\mu\nu}_{;\nu}=0$ FOR TYPE-C MEDIA

These equations for type-C media may be evaluated by the same method as used for type-B media. Namely, one substitutes Eq. (3.13) into Eq. (2.26) and obtains

$$\phi_{,u}[2\phi_{,uv} + \mu_{,v}\phi_{,u} + \mu_{,u}\phi_{,v}]$$
$$+ \psi_{,u}[2\psi_{,uv} + \mu_{,v}\psi_{,u} + \mu_{,u}\psi_{,v}] = 0,$$
$$\phi_{,v}[2\phi_{,vu} + \mu_{,v}\phi_{,u} + \mu_{,u}\phi_{,v}]$$
$$+ \psi_{,v}[2\psi_{,vu} + \mu_{,v}\psi_{,u} + \mu_{,u}\psi_{,v}] = 0.$$

Since $\phi_{,\mu}$ and $\psi_{,\mu}$ are linearly independent vectors we must have

$$\begin{aligned} 2\phi_{,\mu\nu} + \mu_{,v}\phi_{,u} + \mu_{,u}\phi_{,v} &= 0, \\ 2\psi_{,uv} + \mu_{,v}\psi_{,u} + \mu_{,u}\psi_{,v} &= 0. \end{aligned} \tag{6.1a}$$

In other words, ϕ and ψ are each massless scalar fields satisfying

$$g^{\mu\nu}\phi_{;\mu\nu} = g^{\mu\nu}\psi_{;\mu\nu} = 0. \tag{6.1b}$$

The complex scalar

$$\Lambda = \phi + i\psi \tag{6.2}$$

aslo satisfies the wave equation. The vector Λ_μ defined in Sec. III is given by

$$\Lambda_\mu = \Lambda_{,\mu}. \tag{6.3}$$

The energy-momentum tensor of the medium is

$$\kappa T_{\mu\nu} = \phi_{,\mu}\phi_{,\nu} + \psi_{,\mu}\psi_{,\nu} - \tfrac{1}{2} g_{\mu\nu}(\phi_{,\rho}\phi_{,\sigma} + \psi_{,\rho}\psi_{,\sigma}) g^{\rho\sigma}. \tag{6.4a}$$

The energy density is given by

$$2w = |g^{\rho\sigma}\Lambda_{,\rho}\Lambda_{,\sigma}| \tag{6.4b}$$

and the pressure is given by

$$2\pi = \Lambda_\rho \overline{\Lambda}_\sigma g^{\rho\sigma}. \tag{6.4c}$$

The timelike and spacelike proper vectors of the energy-momentum tensor in the two-dimensional space spanned by the null coordinates u and v, that is, the vectors U_μ and Z_μ, are given by

$$\sqrt{w+\pi}\, U_\mu = \mathrm{Re}(e^{i\theta}\Lambda_\mu), \tag{6.5}$$

$$\sqrt{w-\pi}\, Z_\mu = \mathrm{Im}(e^{i\theta}\Lambda_\mu). \tag{6.6}$$

VII. CONCLUSIONS

The various possible interaction regions arising when two null dust clouds fronted by plane impulsive gravitational waves collide have been classified by means of the energy-momentum tensors occurring in this region. For a type-C medium this tensor may be taken to be given by Eq. (6.4a), where ϕ and ψ are two independent solutions of the scalar wave equation. Equation (6.4a) is algebraically equivalent to Eqs. (3.6) and (3.7). The latter equations are those of an anisotropic perfect fluid with energy w and pressures (w,π,π). Knowledge of the two functions ϕ and ψ, each of which satisfies the wave equation, enables one to determine the motion of the anisotropic fluid. The two independent equations $T^{\mu\nu}_{;\nu} = 0$, without additional conditions on w and π, are not sufficient to determine this motion.

Note that if $\psi = 0$ (or $\phi = 0$) $T^{\mu\nu}$ of Eq. (6.4a) becomes the same as that given by (6.3). In this sense media of type B are special cases of those of type C.

If $g^{\mu\nu}\phi_{,\mu}\phi_{,\nu} = g^{\mu\nu}\psi_{,\mu}\psi_{,\nu} = 0$, then $T^{\mu\nu}$ of Eq. (6.4a) becomes that of Eqs. (3.8b) and (3.8c) if in addition one has $\phi = \phi(u)$ and $\psi = \psi(u)$. However, in such a case we must have $\mu = \text{const}$ in order that $T^{\mu\nu}_{;\nu} = 0$. Thus media of type A are special cases of media of type C under the additional assumption that $\mu = \text{const}$.

When the space-time V satisfies assumptions (i)–(iv) the collision of two impulsive plane waves, each followed by a dust cloud, must have an interaction region of type C (which includes type B) or type A. In the former case the motion of the medium in the interaction region is determined by a complex function $\Lambda(u,v)$, which is a solution of the complex scalar wave equation. This function is in turn determined by the functions $\Lambda(0,v)$ and $\Lambda(u,0)$. The latter two functions are determined by the null dusts in regions III and II, respectively. These in turn are determined by setting $u = 0$ or $v = 0$ in $T^{\mu\nu}$ of region I, respectively.

[1] N. Rosen, Phys. Z. Sowjetunion **12**, 366 (1937).
[2] A. H. Taub, J. Math. Phys. **29**, 2622 (1988).
[3] A. H. Taub, J. Math. Phys. **29**, 690 (1988).
[4] S. Chandrasekhar and B. C. Xanthopolous, Proc. R. Soc. London Ser. A **402**, 37 (1985).
[5] S. Chandrasekhar and B. C. Xanthopolous, Proc. R. Soc. London Ser. A **403**, 139 (1986).
[6] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U.P., Cambridge, 1973).
[7] P. S. Letelier, Phys. Rev. D **22**, 807 (1980).
[8] A. H. Taub, in *Developments in General Relativity, Astrophysics and Quantum Mechanics*, Israel Physical Society, edited by F. L. Cooperstock and L. P. Horwitz (Adam Hilger, London, 1990).

The graviton propagator in maximally symmetric spaces

Michael Turyn
Department of Physics and Astronomy, Tufts University, Medford, Massachusetts 02155

(Received 25 July 1989; accepted for publication 8 November 1989)

Vacuum metric perturbation two-point functions are found for maximally symmetric backgrounds of arbitrary dimension n, using tensor mode functions on the n-sphere and a general gauge-fixing term. The gauge-invariant part of the resulting graviton propagator is isolated, and is fully evaluated for de Sitter spaces in terms of functions of the geodesic separation.

I. INTRODUCTION

A. Overview

Maximally symmetric spaces fill a special niche in the theory of quantum fields in curved space-times. Though maximal symmetry simplifies calculation enormously, making it an ideal arena in which to develop the theoretical and computational machinery of curved-background QFT, nontrivial curvature effects still manifest themselves. An easily obtained Minkowski space limit at zero scalar curvature provides for simple checks of all results against standard flat-space results.

Given the somewhat abstract nature of the above advantages, it is gratifying that an understanding of the behavior of the quantum field in a maximally symmetric background should prove to be of some practical use, most notably in the study of the inflationary cosmologies that have been proposed to explain the large-scale structure of the universe.[1] For example, questions as to the $n = 4$ de Sitter universe's stability[2,3] raise the spectre of an inflationary phase too short-lived to address the very concerns (the flatness, isotropy, and horizon problems) inflation was intended to address.

The graviton propagator for a maximally symmetric space of arbitrary dimension is of interest for several reasons. The problem of finding it has the appeal of any general case that can be handled using the lessons of the specific. There may be situations in which one might need the propagator for spaces other than dS^4, in which case having the general solution available would be of use. Sometimes the high symmetry of de Sitter space allows certain expressions to be evaluated using a reduction scheme, in which case one might need to have an expression for the propagator that can be continued away from $n = 4$.

B. Structure

The second-order variation in the graviton action will be found for a maximally symmetric space of arbitrary dimension and curvature, using a general two-parameter gauge-fixing term. This will be used to find a very general expression for the graviton propagator in terms of sums of tensor eigenfunctions of the d'Alembertian operator $\Box$ on the hypersphere, which is the Euclidean section of de Sitter space. It is then possible to continue the results so obtained off the sphere in a well-defined way[4] to obtain the full space. Flat and anti-de Sitter spaces are considered by allowing infinite and imaginary radii of curvature, respectively. Attention will be paid to the isolation of that part of the propagator relevant to gauge-independent objects, and the result related to a particular choice of gauge-fixing parameters.

All results will be expressed in a coordinate-independent fashion using the mode functions. Using the previously obtained de Sitter space vector and scalar propagators,[4] the graviton propagator there will be obtained in terms of maximally symmetric bitensors; these are functions of the geodesic separation μ alone. There is some question as to the correct set of boundary conditions for anti-de Sitter space (adS) or its covering space (CadS).[5] In Ref. 4, a choice corresponding to the reflection of the scalar and vector fields at timelike infinity was made, but only one of two independent solutions was used. The slower of the two solutions was thrown out in each case. A compelling case for reflective boundary conditions to control the flow of information at infinity has been made in Ref. 5. However, the "fastest fall-off" condition is more arbitrary, especially when experience with the de Sitter space graviton propagator has forced us to deal with an object that does not fall off at all at infinity.

Other possiblities exist. For example, it has been proposed that a set of conditions on the behavior of the Weyl tensor at timelike infinity picks out one reflecting boundary condition for the $n = 4$ graviton propagator.[6] Work in progress[7] addresses these considerations with the aid of evaluations of the various curvature tensors' fluctuation expectation values,[8] using the most general solutions for the scalar and vector propagators obeying reflecting boundary conditions. Pending these results, we will not evaluate the adS graviton two-point function further; the points at which we specialize to de Sitter space will be duly noted.

The complete expression for the de Sitter space propagator can be found at the end of Sec. IV [expressions (4.27)–(4.31)], where it is expressed using several functions defined in Sec. III. In Sec. V the long-distance behavior of the propagator and the extension of the n-sphere results to the full space are discussed.

II. BACKGROUND, ACTION, AND PROPAGATOR

A. The background and the action

An n-dimensional de Sitter space can be realized in the Lorentzian embedding space $\{\mathbf{R}^{n+1},\eta_{ab}\}$ as the hyperboloid made up of points with coordinates $\{\mathbf{X}\}$ satisfying

$$\eta_{ab}X^aX^b = a^2,$$

 0022-2488/90/030669-11$03.00

where a is real. Here η_{ab} is the Minkowski $(-+\cdots+)$ metric in $n+1$ dimensions. The Euclidean section of the space is obtained by taking $x_0 \to ix_0$ and $\eta_{ab} \to \delta_{ab}$.[9] The hypersurface becomes an n-sphere of radius a in the now Euclidean embedding space. The proper extension of the Euclidean results to the full hypersurface will be discussed in the final section of this paper.

Flat space can be obtained as the limit $a^2 \to \infty$. Results for anti-de Sitter space (adS) can be obtained by taking $a^2 = -|a|^2$, as examination of the first coordinates/line elements used for de Sitter and anti-de Sitter spaces in Ref. 10 will show. The embedding used for de Sitter space will no longer be valid; rather, adS is embeddable as the surface

$$\zeta_{ab} Y^a Y^b = a^2,$$

where $\zeta = \mathrm{diag}(--+\cdots+)$.

The Euclideanized second-order variation in the action for a maximally symmetric space can be written in a simple form. If one represents the lower components of a perturbation $\delta\mathbf{g}$ of the space's background metric $\mathbf{g}$ by h_{ab}, one has[11]

$$\delta^{(2)}S_E = -\frac{1}{2}\rho \int_V h_{ab} W^{abcd} h_{cd}\, dV, \tag{2.1}$$

where $\rho = (32\pi G)^{-1}$, and

$$\begin{aligned} W^{abcd} = {} & [-\Box + 2a^{-2}] g^{a(c} g^{d)b} \\ & + [\Box + (n-3)a^{-2}] g^{ab} g^{cd} \\ & - (g^{ab}\nabla^c\nabla^d + g^{cd}\nabla^a\nabla^b) + 2\nabla^{(a} g^{b)(c}\nabla^{d)}. \end{aligned} \tag{2.2}$$

Here, as in the rest of this paper, the derivatives are covariant with respect to the background metric $\mathbf{g}$, which will be used to raise and lower indices. The volume V is that of the full n-sphere, the default range of integration in this paper.

To obtain an invertible expression for the propagator, one can add to the second-order variation in the action (2.2) the two-parameter gauge-breaking term

$$S_{\mathrm{GF}} = -\rho \int_V \alpha_0 [\nabla_a (h^{ab} - \alpha_1 h^k{}_k g^{ab})]^2\, dV. \tag{2.3}$$

To obtain the gauge-fixed second-order action, one then replaces the gauge-invariant wave operator W^{abcd} appearing in (2.1) with the gauge-fixed

$$\begin{aligned} W^{abcd}_{\mathscr{F}} = {} & [-\Box + 2a^{-2}] g^{a(c} g^{d)b} \\ & + [(1-2\alpha_0\alpha_1{}^2)\Box + (n-3)a^{-2}] g^{ab} g^{cd} \\ & + (2\alpha_0\alpha_1 - 1)(g^{ab}\nabla^c\nabla^d + g^{cd}\nabla^a\nabla^b) \\ & + 2(1-\alpha_0)\nabla^{(a} g^{b)(c}\nabla^{d)}. \end{aligned} \tag{2.4}$$

The action used in Ref. 12 can be recovered by setting $\alpha_0 = 1$, $\alpha_1 = \frac{1}{2}$, and $n = 4$. It will be noticed immediately that for a maximally symmetric manifold of any dimension, these same choices for α_0 and α_1 give a "wave operator" $W_{\mathscr{F}}$ consisting solely of the d'Alembertian $\Box$ and constants that vanish in flat space ($a^2 \to \infty$). This is the generalization to n dimensions of the propagator found in Ref. 12 for $n = 4$. However, a much simpler result can be obtained by using the Landau gauge conditions $\alpha_0 \to \infty$, $\alpha_1 \to 1/n$.

B. Maximally symmetric bitensors

We are looking for the two-point function

$$G^{abc'd'}(x,x') = \langle 0_E | h^{ab}(x) h^{c'd'}(x') | 0_E \rangle, \tag{2.5}$$

where the states' continuations to Lorentzian de Sitter space are the Gibbons–Hawking vacuum states of Refs. 13 and 14; the recovery of the full Lorentzian space theory by the rotation $x_0 \to -ix_0$ will be an in–out propagation amplitude.[15] This two-point function is a maximally symmetric bitensor.

A bitensor $T^{a\cdots b'\cdots}{}_{c\cdots d'\cdots}(x,x')$ is any function of two points x and x' that transforms as a tensor under coordinate transformations at either point. Here, a prime on an index indicates that the index belongs to x''s tangent space, and unprimed indices to x's. A *maximally symmetric* bitensor is a bitensor that remains invariant under all isometries of the manifold.[4] The amplitude (2.5) is an example of such an object, as the vacuum chosen is maximally symmetric and the operator inside does not break the maximal symmetry (such as if it were to pick out a preferred spatial axis). One further restricts the set of possible maximally symmetric bitensors making up the propagator (2.5) by noting that they must possess the index symmetries

$$a \leftrightarrow b, \quad c' \leftrightarrow d', \quad [(ab) \leftrightarrow (c'd') \;\&\; x \leftrightarrow x'].$$

For any maximally symmetric space, there are but five distinct maximally symmetric objects with these index symmetries. The propagator is therefore expressible as a sum of the form

$$o^1 O_1^{abc'd'} + o^2 O_2^{abc'd'} + o^3 O_3^{abc'd'} + o^4 O_4^{abc'd'} + o^5 O_5^{abc'd'},$$

where $\{O_1^{abc'd'},\ldots,O_5^{abc'd'}\}$ is a basis set of five bitensors with the requisite symmetries, and $\{o^1(\mu),\ldots,o^5(\mu)\}$ are undetermined coefficient functions of the geodesic separation μ. When no geodesic connects x and x', these functions can be continued in a well-defined and consistent way in the Lorentzian embedding space. This last point will be considered in greater detail in Sec. V. When we wish to present an object with the propagator's symmetries, we will use a capital letter for the object, and the equivalent small letter for its coefficient functions. For example, we will say "$\mathbf{G_T}$ is given by $g^1_T = c^1,\ldots,g^5_T = c^5$," meaning

$$G_T^{abc'd'} = c^1 O_1^{abc'd'} + \cdots + c^5 O_5^{abc'd'}.$$

Exactly as in Ref. 12, our five basis bitensors can be built up from the metric and three simple objects. These objects are (1) the metric tensor g^{ab}; (2) the unit tangents at x and x', $n^a = \nabla^a\mu$ and $n^{a'} = \nabla^{a'}\mu$; and (3) the parallel propagator $g^{ae'}$, which has the following properties:

$$g^{ae'} n_a = -n^{e'} \text{ and } g^{ae'} n_{e'} = -n^a,$$

$$g^{ae'} g_{e'b} = g^a{}_b, \quad g^{ae'} g_{ab} = g^{e'}{}_b, \quad g^{ae'} g_{e'f'} = g^a{}_{f'}.$$

Rules for the manipulation of these fundamental objects can be found in Table I. More complicated identities can be found in Appendix A. In this paper, we will denote the basis bitensors by $\{\mathbf{O_k} : k = 1,\ldots,5\}$, defining them as

$$O_1^{abc'd'} = g^{ab} g^{c'd'}, \tag{2.6a}$$

$$O_2^{abc'd'} = n^a n^b n^{c'} n^{d'}, \tag{2.6b}$$

$$O_3^{abc'd'} = 2g^{a(c'} g^{d')b}, \tag{2.6c}$$

TABLE I. Bitensor manipulation rules.

$$\nabla^a n^b = A\,(g^{ab} - n^a n^b)$$
$$\nabla^a n^{c'} = C(g^{ac'} + n^a n^{c'})$$
$$\nabla^a g^{bc'} = -(A+C)(g^{ab} n^{c'} + g^{ac'} n^b)$$

where

$$A = a^{-1}\cot(\mu/a)$$
$$C = -a^{-1}\csc(\mu/a)$$

Note that

$$C^2 - A^2 = a^{-2}$$
$$\frac{dA}{d\mu} = -C^2$$
$$\frac{dC}{d\mu} = -AC$$

$$O_4^{abc'd'} = g^{ab} n^{c'} n^{d'} + n^a n^b g^{c'd'}, \tag{2.6d}$$
$$O_5^{abc'd'} = 4 n^{(a} g^{b)(c'} n^{d')}. \tag{2.6e}$$

Note that $O_4^{abc'd'}$ cannot be reproduced by the separate symmetrization of the primed and unprimed indices of a single bitensor term, as can the others. This makes it convenient for us to define an auxiliary object $\mathbf{O}_6$ to be

$$O_6^{abc'd'} = g^{ab} n^{c'} n^{d'} - n^a n^b g^{c'd'}. \tag{2.6f}$$

No object with our propagator's index symmetries can have a nonzero $\mathbf{O}_6$ term. Since this is so, we will not give an $\mathbf{O}_6$ coefficient for most of the objects found in this paper—it can be assumed to be 0 unless otherwise noted. In general, an expression of the form $\Box \mathbf{O}_i$ will have an $\mathbf{O}_6$ term because it does not have the $x \leftrightarrow x'$ symmetry that $(\Box + \Box')\mathbf{O}_i$ has.

C. Equation of motion and mode-sum inversion

The Euclideanized propagator $\mathbf{G}$ defined in (2.5) obeys the equation of motion[9,16]

$$W_{\mathscr{F}}^{abcd} G_{cd}{}^{e'f'} = \tfrac{1}{2} O_3{}^{abe'f'} \delta^n(x,x') \equiv \delta^{abe'f'}. \tag{2.7}$$

The factor of ρ has been set equal to 1, and will be resurrected in the final results. Using a set of tensor mode functions defined in the n-sphere (2.7) can be inverted, and the inverted expression evaluated using the vector and scalar propagators found in Ref. 4. We will obtain the propagator for a general choice of gauge as a sum of five tensor mode-sum pieces in this section, and then isolate that part of the propagator (present for any gauge choice) which can contribute to gauge-invariant physical observables. Examination of the mode-sum expression for this part of the propagator will show it to be the propagator obtained from the Landau gauge choice. The rest of this paper will be concerned with the evaluation of the Landau gauge propagator as a function of the geodesic separation $\mu(x,x')$. In Appendix B, the orthonormal mode functions $\{\chi_k^{ab}\}$, $\{W_k^{ab}\}$, $\{V_k^{ab}\}$, and $\{h_k^{ab}\}$ are introduced (a degeneracy index has been supressed in each case). These span the two-tensors on the n-sphere. One can therefore write

$$\delta^{abc'd'} = \sum_{k=0}^{\infty} h_k^{ab} h_k^{c'd'} + \sum_{k=1}^{\infty} V_k^{ab} V_k^{c'd'} + \sum_{k=2}^{\infty} W_k^{ab} W_k^{c'd'} + \sum_{k=0}^{\infty} \chi_k^{ab} \chi_k^{c'd'}. \tag{2.8}$$

The result of performing a functional integral over the metric perturbations[13] shows that the graviton propagator takes the following form:

$$G^{abc'd'} = \sum_{k=0}^{\infty} a_k [h_k^{ab} h_k^{c'd'}] + \sum_{k=1}^{\infty} b_k [V_k^{ab} V_k^{c'd'}] + \sum_{k=2}^{\infty} c_k [W_k^{ab} W_k^{c'd'}] + \sum_{k=0}^{\infty} d_k [\chi_k^{ab} \chi_k^{c'd'}] + \sum_{k=2}^{\infty} e_k [\chi_k^{ab} W_k^{c'd'} + W_k^{ab} \chi_k^{c'd'}]. \tag{2.9}$$

The general form for the propagator must contain the mixed $W^{ab}\chi^{c'd'} + \chi^{ab} W^{c'd'}$ terms found in the last part of (2.9) because the general wave operator contains term mix $\mathbf{W}_k$ and χ_k modes. The effect of the wave operator on the propagator can be expressed in terms of the tensor eigenfunctions and the eigenvalues $\lambda_k^{(s)}$ of the spin-s mode functions (for $s = 0,1,2$). Using the results of Appendix B, we find

$$W_{\mathscr{F}}^{abcd} h_{cd}^{k} = (-\lambda_k^{(2)} + 2a^{-2}) h_k^{ab}, \quad \text{for } k \geqslant 0, \tag{2.10}$$
$$W_{\mathscr{F}}^{abcd} V_{cd}^{k} = \alpha_0 (-\lambda_k^{(1)} - (n-1)a^{-2}) V_k^{ab}, \quad \text{for } k \geqslant 1, \tag{2.11}$$

and

$$W_{\mathscr{F}}^{abcd} \begin{pmatrix} W_{cd}^{k} \\ \chi_{cd}^{k} \end{pmatrix} = \begin{pmatrix} \mathscr{A}_k & \mathscr{B}_k \\ \mathscr{C}_k & \mathscr{D}_k \end{pmatrix} \begin{pmatrix} W_k^{ab} \\ \chi_k^{ab} \end{pmatrix}, \tag{2.12}$$

where a matrix/vector notation has been used to show the mode-mixing nature of $W_{\mathscr{F}}$. For $\{k = 0,1\}$, $\mathscr{A}_k = \mathscr{B}_k = \mathscr{C}_k = 0$, but, for $k \geqslant 2$,

$$\mathscr{A}_k(\alpha_0) = -[1 + (2(\alpha_0 - 1)(n-1)/n)] \times \lambda_k^{(0)} - 2\alpha_0 (n-1) a^{-2} \tag{2.13a}$$

and

$$\mathscr{B}_k(\alpha_0,\alpha_1) = \mathscr{C}_k(\alpha_0,\alpha_1) = \{(n-1)\lambda_k^{(0)}(\lambda_k^{(0)} + na^{-2})\}^{1/2} \times [(2\alpha_0\alpha_1 - 1) + (2/n)(1-\alpha_0)]. \tag{2.13b}$$

For all $k \geqslant 0$,

$$\mathscr{D}_k(\alpha_0,\alpha_1) = [(1 - 2\alpha_0\alpha_1{}^2)n + (4\alpha_0\alpha_1 - 3) + (2/n)(1-\alpha_0)]\lambda_k^{(0)} + (n-1)(n-2)a^{-2}. \tag{2.13c}$$

We now substitute the mode-function forms for the propagator $\mathbf{G}$ (2.9) and the tensor δ function (2.8) into the equation of motion (2.7). The orthonormality of the modes (C3) can then be used to pick out the following eigenvalue equations:

$$(-\lambda_k^{(2)} + 2a^{-2}) a_k = 1, \tag{2.14}$$
$$\alpha_0 (-\lambda_k^{(1)} - (n-1)a^{-2}) b_k = 1, \tag{2.15}$$

$$\mathscr{A}_k c_k + \mathscr{C}_k e_k = 1, \quad (2.16a)$$

$$\mathscr{B}_k e_k + \mathscr{D}_k d_k = 1, \quad (2.16b)$$

$$\mathscr{A}_k e_k + \mathscr{C}_k d_k = \mathscr{B}_k c_k + \mathscr{D}_k e_k = 0. \quad (2.16c)$$

Solving these equations for $\{a_k, b_k, \ldots\}$, we find that the propagator can be expressed as the sum of five terms:

$$\mathbf{G} = \mathbf{G}_{\mathrm{TT}} + \mathbf{G}_{\mathrm{T}} + \mathbf{G}_{\mathrm{L}} + \mathbf{G}_{\mathrm{PT}} + \mathbf{G}_{\mathrm{M}}, \quad (2.17)$$

where

$$G_{\mathrm{TT}}^{abc'd'} = \sum_{k=0}^{\infty} \frac{h_k^{ab} h_k^{c'd'}}{-\lambda_k^{(2)} + 2a^{-2}}, \quad (2.18)$$

$$G_T^{abc'd'} = \frac{1}{\alpha_0} \sum_{k=1}^{\infty} \frac{V_k^{ab} V_k^{c'd'}}{-\lambda_k^{(1)} - (n-1)a^{-2}}, \quad (2.19)$$

$$G_L^{abc'd'} = \sum_{k=2}^{\infty} \frac{W_k^{ab} W_k^{c'd'}}{-\mathscr{E}_k / \mathscr{D}_k}, \quad (2.20)$$

$$G_M^{abc'd'} = \sum_{k=2}^{\infty} \frac{W_k^{ab} \chi_k^{c'd'} + \chi_k^{ab} W_k^{c'd'}}{\mathscr{E}_k / \mathscr{B}_k}, \quad (2.21)$$

and

$$G_{\mathrm{PT}}^{abc'd'} = \sum_{k=0,1} \frac{\chi_k^{ab} \chi_k^{c'd'}}{\mathscr{D}_k} + \sum_{k=2}^{\infty} \frac{\chi_k^{ab} \chi_k^{c'd'}}{-\mathscr{E}_k / \mathscr{A}_k}, \quad (2.22)$$

where

$$\mathscr{E}_k = \mathscr{B}_k^2 - \mathscr{A}_k \mathscr{D}_k.$$

We will call these terms the transverse-traceless, transverse, longitudinal, mixed, and pure-trace parts of the propagator; note that the transverse-traceless term is completely independent of the choice of gauge-term parameters. This general result for the propagator agrees with the equivalent special cases discussed in Refs. 2 and 12, and can be evaluated using the techniques employed in those two papers.

D. Source terms for physical observables

Since one would like to use the propagator to find correlation functions for physical observables, it is useful to examine quantities formed from the propagator that are left unchanged by the usual gauge transformation

$$h_{\mu\nu} \to h_{\mu\nu} + 2k_{(\mu;\nu)},$$

where k_μ is a vector.

The vector and scalar eigenfunctions of $\Box$ are represented by ξ_k^a and φ_k, respectively. The shear tensor modes $\{\mathbf{V}_k\}$ are proportional to $\nabla^{(a}\xi_k^{b)}$, and so cannot contribute at all to gauge-invariant quantities. A part of the W_k^{ab} mode proportional to $\nabla^a\nabla^b\varphi_k$ will also not contribute. In addition to these explicitly noncontributing terms, one also finds in the propagator an expression of the form

$$\Phi_1 g^{ab} g^{c'd'},$$

where

$$\Phi_1(x,x') = \frac{a^2}{2\alpha_0(1-n\alpha_1)^2} \sum_{i=1}^{n+1} \varphi_1^i(x)\varphi_1^i(x'). \quad (2.23)$$

We can evaluate this sum over degenerate $k=1$ scalar modes using expression (C6). We find that $\Phi_1 \propto (1-2z)$, where $z = \cos^2(\mu/2a)$, and obeys the conformal Killing equation

$$\Phi_1 g^{ab} = -a^2 \nabla^a \nabla^b \Phi_1. \quad (2.24)$$

Terms of the form $g^{ab}\Phi_1$ are therefore also pure gauge, and should be dropped from the physically relevant propagator.

For any initial choice of gauge, subtracting all the pure-gauge modes' contributions from the propagator leaves a "reduced" propagator $\mathbf{G}_{\mathrm{red}}$ which is independent of the gauge-fixing parameters α_0 and α_1. One obtains

$$G_{\mathrm{red}}^{abc'd'} = G_{\mathrm{TT}}^{abc'd'} + G_{\substack{\mathrm{PT} \\ \mathrm{red}}}^{abc'd'}, \quad (2.25)$$

where

$$G_{\substack{\mathrm{PT} \\ \mathrm{red}}}^{abc'd'} = -[\mathscr{G}_{(0)}/(n-2)(n-1)] g^{ab} g^{c'd'}. \quad (2.26)$$

Here, the finite $\mathscr{G}_{(0)}$ is a modified scalar propagator given by

$$\mathscr{G}_{(0)} = \mathscr{G}_{(0)}(z) = \frac{\varphi_0(x)\varphi_0(x')}{-\lambda_0^{(0)} - na^{-2}} \sum_{k=2}^{\infty} \frac{\varphi_k(x)\varphi_k(x')}{-\lambda_k^{(0)} - na^{-2}}. \quad (2.27)$$

This is the scalar propagator for a field with $m^2 = -na^{-2}$, less the $k=1$ modes' contribution of Φ_1. With that term removed, this modified propagator is finite for separated points x and x'.

This is exactly the same form for the propagator as one would obtain using the Landau gauge choice $\{\alpha_0 \to \infty, \alpha_1 \to n^{-1}\}$ if α_0 and α_1 approach their limiting values as $\lim_{\zeta\to 0}\{\zeta^{-3}, \zeta + n^{-1}\}$. Other limiting behaviors can produce an infinite term proportional to $\Phi_1{}^2$; this term is a noncontributing gauge artifact, and can be of no physical significance.[13] As well as picking out the physically significant parts of the propagator, the Landau-gauge solution is also comparatively simple in form and use. For example, contracting this form for the propagator with the metric tensor or taking its divergence leaves only the pure-trace piece.

III. THE PURE-TRACE PROPAGATOR

In Sec. II, the PT piece of the propagator was found to be

$$G_{\substack{\mathrm{PT} \\ \mathrm{red}}}^{abc'd'}(x,x') = -[1/(n-2)(n-1)] \mathscr{G}_{(0)} g^{ab} g^{c'd'}, \quad (3.1)$$

where $\mathscr{G}_{(0)}$ is a modified scalar propagator on the n-sphere. Using the scalar propagator of one finds that

$$\mathscr{G}_{(0)}(z) = \lim_{\epsilon\to 0}[G([\epsilon - n]a^{-2}, z) - ((n+1)/\epsilon V_n)(2z-1)] \quad (3.2)$$

is indeed regular. Here we have used Eq. (C6) for the piece to be removed from the full propagator.

In de Sitter space, one can take the limit (3.2) to find[17]

$$\mathscr{G}_0(z) = r_0 \int_0^z \int_0^{z'} f\left(n+2, 1; \frac{n}{2}+2; z''\right) dz''\, dz' - \frac{a^2}{V_n}\left[1 + \frac{1}{n} + \left[\frac{1}{n} + \frac{1}{n+1} + \psi_0(n) + C\right](2z-1)\right], \quad (3.3)$$

where ψ_0 is the digamma function, C is the Euler–Mascheroni constant, and V_n is the volume of the n-sphere [see Eq. (A3)]. We have introduced

$$f(a,b;c;z) \equiv [\Gamma(a)\Gamma(b)/\Gamma(c)]\, {}_2F_1(a,b;c;z), \qquad (3.4)$$

which is easier to manipulate toward our ends than the standard hypergeometric function. The spatially constant

$$r_0(n) = 2^{-n}\pi^{-n/2}a^{2-n} \qquad (3.5)$$

gives the scalar propagator the correct short-distance ($\mu/a \to 0$) behavior. In practice, the term proportional to $(2z-1)$ can be ignored. Exactly like Φ_1 in Sec. II, its contribution to the propagator will be a pure gauge term of no physical relevance.

Equation (2.27) is valid for any maximally symmetric space, but evaluation of it for anti-de Sitter space would be dependent on the specific form of the scalar propagator there. As mentioned in the Introduction, this lies somewhat out of the scope of the present work, and will be taken up in work to come.

IV. THE TRANSVERSE-TRACELESS PART OF THE PROPAGATOR

We evaluate $\mathbf{G}_{\rm TT}$, the transverse-traceless part of the propagator, using the methods of Ref. 12. We know that it can be expressed as

$$\mathbf{G}_{\rm TT} = \alpha(\mu)\mathbf{O}_1 + \beta(\mu)\mathbf{O}_2 + \psi(\mu)\mathbf{O}_3 + \delta(\mu)\mathbf{O}_4 + \varepsilon(\mu)\mathbf{O}_5. \qquad (4.1)$$

We will use the equation of motion for $\mathbf{G}_{\rm TT}$, and its tracelessness and transverseness, to solve for the undetermined coefficient functions $\{\alpha,\beta,\psi,\delta,\varepsilon\}$.

A. Transverse-tracelessness

We require that $G_{\rm TT}^{abc'd'}$ be traceless on both sets of indices—primed and unprimed. Tracing on the primed indices, one finds that tracelessness requires

$$(n\alpha + 2\psi + \delta)g^{ab} + (\beta + n\delta - 4\epsilon)n^a n^b = 0, \qquad (4.2)$$

and so

$$n\alpha + 2\psi + \delta = 0, \qquad (4.3a)$$

$$\beta + n\delta - 4\varepsilon = 0. \qquad (4.3b)$$

Thus there are only three independent coefficient functions. One can define three traceless bitensor objects, in terms of which any traceless objects with the propagator's symmetries can be written. If one calls the three traceless objects $\{\mathbf{T}_1,\mathbf{T}_2,\mathbf{T}_3\}$ and writes $\mathbf{G}_{\rm TT}$ as

$$\mathbf{G}_{\rm TT} = X(\mu)\mathbf{T}_1 + Y(\mu)\mathbf{T}_2 + Z(\mu)\mathbf{T}_3, \qquad (4.4)$$

then the conditions of (4.3a) and (4.3b) are satisfied automatically. We choose the traceless objects to be

$$\mathbf{T}_1 = (1/n^2)\mathbf{O}_1 + \mathbf{O}_2 - (1/n)\mathbf{O}_4,$$
$$\mathbf{T}_2 = \mathbf{O}_3 - (2/n)\mathbf{O}_1, \qquad (4.5)$$
$$\mathbf{T}_3 = \mathbf{O}_5 + 4\mathbf{O}_2.$$

Equating (4.4) and (4.1), one finds

$$\alpha = (1/n^2)X - (2/n)Y, \quad \beta = X + 4Z,$$
$$\psi = Y, \quad \delta = -(1/n)X, \quad \varepsilon = Z. \qquad (4.6)$$

We will work with the functions X, Y, and Z for now.

$\mathbf{G}_{\rm TT}$ is also transverse; that is, it obeys

$$\nabla_a G_{\rm TT}^{abc'd'} = 0. \qquad (4.7)$$

Substituting in the form for $\mathbf{G}_{\rm TT}$ of (4.4) into (4.7), one obtains three equations in the propagator's coefficient functions. However, only two of these three equations are independent in our case, as any two of them and the tracelessness conditions can produce the third. Two independent equations so obtained are

$$2n^2CZ - 2n[Y' + n(A+C)Y] - (n-1)[X' + nAX] = 0$$

and $\qquad (4.8)$

$$Z' + nAZ - [Y' + n(A+C)Y] - (C/n)X = 0,$$

where $f' \equiv df/d\mu$, and A and C are the functions defined in Table I.

B. The TT source term

From the definition of the transverse-traceless mode functions (B1c), and the definition of the transverse-traceless part of the propagator (2.18), one can see that $\mathbf{G}_{\rm TT}$ obeys the equation of motion

$$(-\Box + 2a^{-2})G_{\rm TT}{}^{abc'd'} = \sum_{k=0}^{\infty} h_{{\rm TT}\,k}^{\;ab} h_{{\rm TT}\,k}^{\;c'd'} \equiv \Theta^{abc'd'}. \qquad (4.9)$$

The source term appearing on the right-hand side of (4.9) is *not* proportional to a delta function, but can be evaluated in terms of the modified scalar and vector propagators of Appendix C.

Defining the operator $\overline{\nabla^a\nabla^b}$ to be the traceless part of $\nabla^a\nabla^b$,

$$\overline{\nabla^a\nabla^b} \equiv (\nabla^a\nabla^b - (1/n)g^{ab}\Box),$$

relation (2.8) and the results of Appendices A and C [see especially (A8) and (A11) and (C11)–(C16)] give us $\mathbf{\Theta}$ in terms of the traceless part of the tensorial delta function of (2.7) and modified vector and scalar propagators:

$$\Theta^{ab}_{c'd'} = -\frac{1}{2}\delta^n(x,x')T^{ab}_{2c'd'} - 2\mathscr{G}^{(a;b)}_{(1)(c';d')} - [a^2/(n-1)]\,\overline{\nabla^a\nabla^b}\,\overline{\nabla_{c'}\nabla_{d'}}\,\mathscr{G}_{(0)}. \qquad (4.10)$$

As explained in Appendix C, $\mathscr{G}^{a}_{(1)c'}$ comes from an incomplete sum over the vector modes used to obtain the spin-1 propagator. The full spin-1 propagator with mass-squared m^2 has a pole at $m^2 = -2R/n$, and is obtained using a scalar function $\gamma(m^2,z)$ with a pole at the same mass value. We obtain the finite $\mathscr{G}^{a}_{(1)c'}$ from γ_0, the finite portion of $\gamma(m^2,z)$ at that pole.

C. Solution of the equation of motion

One can now find the equation of motion for the coefficient functions introduced earlier by examining the independent components of the wave equation

$$(\nabla - \kappa) G_{\rm TT}^{abc'd'} = -\Theta^{abc'd'}(z) \tag{4.11}$$

(for the moment we will use an arbitrary mass-squared κ). The evaluation of $\Box \mathbf{G}_{\rm TT}$ is simplified because

$$\Box(f\mathbf{O}_i) = (\Box f)\mathbf{O}_i + f(\Box \mathbf{O}_i), \tag{4.12}$$

for $\mathbf{O}_i$ any of our basis objects, and $f(\mu)$ any well-behaved function of μ. The expressions for $\Box \mathbf{O}_i$ can be found in Appendix A, and one can easily express $\Box f$ as a function of z or μ:

$$\Box f = \left[\frac{d^2}{d\mu^2} + (n-1)A\frac{d}{d\mu}\right] f$$
$$= a^{-2}\left[z(1-z)\frac{d^2}{dz^2} + \frac{n}{2}(1-2z)\frac{d}{dz}\right] f. \tag{4.13}$$

One can then find the five coefficient functions for $\mathbf{K} \equiv (\Box - \kappa)\mathbf{G}_{\rm TT}$. An $\mathbf{O}_6$ term vanishes identically because of one of the tracelessness conditions (4.3b).

The five objects $\mathbf{O}_1,...,\mathbf{O}_5$ are linearly independent with respect to contraction. The equation of motion (4.11) therefore implies that the bitensor coefficient functions that determine $\mathbf{K}$ and $\boldsymbol{\Theta}$ satisfy $k_i + \vartheta_i = 0$, for $i = 1,...,5$. Because of the tracelessness conditions of (4.3a) and (4.3b), there are three independent equations; we will use

$$k_2 + \vartheta_2 = 0, \tag{4.14a}$$
$$k_3 + \vartheta_3 = 0, \tag{4.14b}$$
$$k_5 + \vartheta_5 = 0. \tag{4.14c}$$

This means that

$$4(Z'' + (n-1)AZ' - [\kappa + (n+2)(A-C)^2]Z)$$
$$+ X'' + (n-1)AX'$$
$$- [\kappa + 2n(C^2 + A^2) - 8AC]X = -\vartheta_2, \tag{4.15}$$
$$Y'' + (n-1)AY' - [\kappa + 2(A+C)^2]Y$$
$$+ 4ACZ = -\vartheta_3,$$
$$Z'' + (n-1)AZ' - [\kappa + (n+2)(C^2 + A^2) - 4AC]Z$$
$$+ n(C+A)^2 Y + 2ACX = -\vartheta_5.$$

Introducing the functions $W(\mu)$ and $U(\mu)$, defined by

$$W = X + [2n/(n-1)]Y, \tag{4.16}$$
$$U = Y - Z, \tag{4.17}$$

one can decouple the five equations (4.8) and (4.15), obtaining

$$W'' + (n+3)AW' - (2na^{-2} + \kappa)W$$
$$= -\vartheta_2 - [2n/(n-1)]\vartheta_3 + 4\vartheta_5, \tag{4.18}$$
$$U = [(n-1)/2(n-1)^2 C](W' + nAW), \tag{4.19}$$
$$Y = -[(n-1)/n(n-2)(n+1)]$$
$$\times [W + (n/C)(U' + nAU)], \tag{4.20}$$

where we have used $C^2 - A^2 = a^{-2}$. These last three equations, and (4.16) and (4.17) above, completely determine the functions $\{X,Y,Z\}$ in terms of W.

Setting $\kappa = 2a^{-2}$, and performing a change of variable to $z = \cos^2(\mu/2a)$, we obtain

$$z(1-z)\frac{d^2W}{dz^2} + \left(\frac{n}{2} + 2\right)(1-2z)\frac{dW}{dz}$$
$$- 2(n+1)W = a^2 S(z). \tag{4.21}$$

This can be expressed using the usual hypergeometric operator $H = {}_2H_1$ as

$$H(a_2,b_2;c_2;z)W = a^2 S(z),$$

for $a_2 = n+1$, $b_2 = 2$, and $c_2 = n/2 + 2$. The source term $S(z)$ appearing on the right-hand side of (4.21) is

$$S(z) = -\vartheta_2 - [2n/(n-1)]\vartheta_3 + 4\vartheta_5.$$

For separated points x and x' this source term can be written as a sum of two pieces,

$$S(z) = a^{-2}[S_0(z) + S_1(z)], \tag{4.22}$$

which are obtained from spin-0 and spin-1 propagators. These pieces of the source can be found using (4.10) and expressions (A5), (A6), (A10), and (A11). One obtains

$$S_1(z) = [n^2/2(n-1)^2][2(z-1)z\partial_z^2\gamma_0 + (2^z - 1)\partial_z\gamma_0] \tag{4.23}$$

and

$$S_0(z) = [1/(n-1)]\{[z(1-z)\partial_z^2]^2$$
$$+ [n/2(n-1)]\partial_z^2\} \quad \mathcal{G}_0. \tag{4.24}$$

Equation (4.21) is an inhomogeneous hypergeometric equation for W. Its general solution is of the form

$$W(z) = W_p(z) + rW_1(z) + sW_2(z), \tag{4.25}$$

where W_1 and W_2 are linearly independent solutions of the homogeneous equation

$$H(a_2,b_2;c_2;z)W = 0$$

and W_p is a particular solution of (4.21).

At this point we will specialize to de Sitter space. In order to evaluate the source term further, one needs expressions for the scalar and vector propagators appropriate to the space in question; the attendant difficulties in the case of anti-de Sitter space have already been touched upon in the Introduction. We will proceed no further for adS, pending further investigation.

The case of de Sitter space is much easier: the scalar and vector propagators are all fixed by the requirements that the solutions be de Sitter invariant, have only one pole in the complex z plane, and have the same short distance ($\mu/a \to 0$) behaviors as the equivalent flat-space propagators.[4] We will therefore be able to evaluate the source term further; these same conditions will also be enough to fix the graviton propagator in de Sitter space.

Using the definition of γ found in Ref. 4 and of γ_0 found in Appendix C, one obtains

$$\gamma_0 = \tfrac{1}{4}\partial_z \mathcal{G}_{(0)} + \text{const},$$

where the constant in question will not matter, as examination of (4.24) will show. Using the result for $\mathcal{G}_{(0)}$ given in Sec. II, one obtains [using r_0 from Eq. (3.5)]

$$S_{dS}(z) = [-na^{-2}r_0/4(n-1)^2]$$
$$\times\{2(z-1)z[2(n-1)z-2n+1]$$
$$\times\partial_z f(a_2+1,b_2-1;c_2;z)$$
$$-[4(n-1)z^2-2(3n-2)z+n-2]$$
$$\times f(a_2+1,b_2-1;c_2;z)\}, \quad (4.26)$$

where liberal use of the hypergeometric equation satisfied by $f(a_2+1,b_2-1;c_2;z)$ has been made.

One can then find a particular solution to (4.21); this can be done easily with the aid of results in Appendix D. Using these, we find

$$W_p(z) = \frac{-nr_0}{4(n-1)^2}\left\{-\frac{\Gamma(n)}{\Gamma(n/2)}[3n-1-2(n-1)z] + \left[(n-1)z^2-(2n-1)z-n-\frac{5}{2}\right]f\left(n+2,1,\frac{n}{2}+2;z\right) + n(n+1)\left[f\left(n,1,\frac{n}{2}+1;z\right) - H^{-1}f\left(n+1,2,\frac{n}{2}+2;z\right)\right] + \frac{n}{4}(5n^2+n-10)f\left(n,1,\frac{n}{2}+2;z\right)\right\}, \quad (4.27)$$

where H^{-1} is an inverse hypergeometric operator defined in (D1).

Two independent solutions to the homogeneous equation obtained from (4.21),

$$H(a_2,b_2;c_2;z)w=0,$$

are $f(a_2,b_2;c_2;z)$ and $f(a_2,b_2;c_2;1-z)$. The second solution has a pole at $z=0$, and neither our particular solution nor the first solution has one to cancel it. We know then that it cannot contribute, and are left only with finding $r_2(n)$ for which

$$W = W_p + r_2 f(n+1,2;n/2+2;z) \quad (4.28)$$

has the correct short-distance behavior. This is done by examining the behavior of the propagator as $\mu/a\to 0$. In this limit,

$$a^2(1-z)\to\mu^2/4.$$

In n dimensions, the traceless part of the flat-space graviton propagator goes as μ^{2-n} (see Ref. 11). Moreover, an analysis of the expressions for the TT coefficient functions as functions of W shows that their limiting behaviors in this regime mirror that of W, which may therefore be regarded as a "generic" coefficient function. Use of limiting expressions for the hypergeometric functions near $z=1$ then lets us fix

$$r_2(n) = [-nr_0/(n-1)^2](n+\tfrac{3}{4}). \quad (4.29)$$

The resulting expression for the full solution W obeying de Sitter boundary conditions matches that found for previously $n=4$.[12]

D. The propagator coefficient functions

Expressed as functions of $W(z)$ and the hypergeometric operator, the original five coefficient functions for the TT part of the propagator of (4.1) are given by

$$\alpha = -\delta - [(n-1)/n^2]W,$$
$$\beta = [4(n-1)/n](1-z)$$
$$\times\{[1/n(n+1)]zH(n+1,n;2(n+1);z)+1\}W,$$
$$\psi = [(n-1)/2](\delta + (1/n)W), \quad (4.30)$$
$$\delta = [4(n-1)/n^2(n-2)(n+1)]z(1-z)$$
$$\times H(n+1,n;n+1;z)W,$$
$$\epsilon = \tfrac{1}{4}(\beta + n\delta).$$

These functions completely determine the TT part of the propagator with the bitensor objects of (2.6a)–(2.6e).

The full propagator can now be obtained by merely adding the contribution of the PT part of the propagator to $\alpha(z)$ as given above:

$$\alpha_{\rm Full} = \alpha_{\rm TT} - [1/(n-1)(n-2)]\mathscr{G}_{(0)}, \quad (4.31)$$

where $\mathscr{G}_{(0)}$ is given by (3.2). Recall that we have been working in units for which

$$32\pi G = 1,$$

which should be made explicit to restore the constant factor ignored since the beginning of Sec. II.

We have found the propagator for a de Sitter space of arbitrary dimension. Preliminary work indicates that the transverse-traceless part of the anti-de Sitter space propagator will be of the form

$$\mathbf{G}^{\rm tt}_{\rm ads}(z) = \sum_{i=0,1}\left[\mathbf{G}^{\rm tt^i}_{\rm ds}(z) + \omega^i(n)\mathbf{G}^{\rm tt^i}_{\rm ds}(1-z)\right].$$

The $\mathbf{G}^{\rm tt^i}_{\rm ds}$'s will be obtained using the S_0 and S_1 pieces of the de Sitter space source term, and the factors ω^i will be set by constraints on the long-distance behavior of the propagator and curvature fluctuation expectation values.

V. PROPAGATOR BEHAVIOR AND IMPLICATIONS

The limit $z(x,x')\to -\infty$ corresponds to large spatial separations of the two points x and x', and $z(x,x')\to\infty$ to large timelike separations. This is most easily seen by writing $z(x,x')$ in terms of the separation between the equivalent two points X and X' in the Lorentzian embedding space. Working from a result in Ref. 4, we find that, for $d^2\equiv n_{ab}(X-X')^a(X-X')^b$, one has

$$z = 1-(d/2a)^2. \quad (5.1)$$

This expression for z matches $z=\cos^2(\mu/2a)$ for μ real, and also when we pick out a time coordinate by letting $x_0\to -ix_0$. It extends z to the case of spacelike-separated x and x' that are not connected by a geodesic. The various Green's functions are obtainable from the behavior of the two-point function in the complex z plane.[3,4]

As $z\to\pm\infty$, the PT piece of the propagator diverges like $z\ln|z|$, and the TT piece like $\ln|z|$. However, the long-distance behavior of the propagator is not in itself sufficient grounds to characterize de Sitter space as unstable. The propagator is not itself observable. When physical observables are found using the propagator, the results need not have the same long-range behavior as the propagator.

For example, in Refs. 8 and 18, the two-point functions for perturbations of various curvature tensors in perturbed maximally symmetric spaces are discussed. They are found to be identically 0 or to drop off very quickly for large $|z|$. For any maximally symmetric space,

$$\langle {}^{(1)}R^{a}{}_{b}{}^{(1)}R^{e'}{}_{f'}\rangle = \langle {}^{(1)}R\,{}^{(1)}R'\rangle = 0.$$

For an $n = 4$ de Sitter space and large $|z|$,

$$\langle {}^{(1)}R^{ab}{}_{cd}{}^{(1)}R^{e'f'}{}_{r's'}\rangle \sim |z|^{-2}$$

and

$$\langle {}^{(1)}R^{a0}{}_{c0}{}^{(1)}R^{e'0'}{}_{r'0'}\rangle \sim |z|^{-3}.$$

Since this last object determines the geodesic deviations of test particles near our two points, observers anchored to distant points see only de Sitter-like motion in test particles. A similar result will hold for anti-de Sitter space; its general form will be used to establish appropriate boundary conditions for adS.[4]

In any event, the propagator is well behaved for points within one horizon length of each other. This has been used for $n = 4$ to argue that observers always see a locally undistorted de Sitter space, though the long-range divergence will show up if there is a transition to a space lacking the de Sitter particle horizon.[19] Note as well that the divergence of the propagator for large separations is not pathological for cosmologies with a de Sitter-like phase of finite duration.

ACKNOWLEDGMENTS

The author would like to thank Professor L. Ford and Professor B. Allen for many insightful and illuminating discussions in the course of preparing this paper, and Professor A. Napier for his help in the preparation of the original manuscripts.

This work was supported in part by the National Science Foundation through NSF Grants No. PHY8506593 and No. PHY8705107.

APPENDIX A: CURVATURE AND BITENSOR IDENTITIES

1. Curvature identities

In a maximally symmetric space with curvature parameter a^2, the background curvature tensors are[20]

$$\begin{aligned} R^{abcd} &= (1/a^2)[g^{ac}g^{bd} - g^{ad}g^{bc}],\\ R^{ac} &= [(n-1)/a^2]g^{ac}, \qquad\qquad (\mathrm{A1})\\ R &= n(n-1)/a^2. \end{aligned}$$

It follows that the cosmological constant Λ obeying the Einstein equation is given by

$$\Lambda = \frac{n-2}{2n}R = \frac{(n-2)(n-1)}{2a^2}. \qquad (\mathrm{A2})$$

The n-sphere volume is given by

$$V_n = \frac{2\pi^{(n+1)/2}a^n}{\Gamma((n+1)/2)} = \frac{\Gamma(n/2)}{\Gamma(n)}(4\pi a^2)^{n/2}. \qquad (\mathrm{A3})$$

2. Bitensor identities

All the identities given here can be derived from the formulas in Table I. Other helpful relations can be found in Appendix C of Ref. 4. In n dimensions, we find, for the objects defined in (2.6a)–(2.6f),

$$\begin{aligned} \Box\mathbf{O}_1 &= 0,\\ \Box\mathbf{O}_2 &= -a^{-2}\mathbf{O}_6 + 2AC\,\mathbf{O}_5\\ &\quad + (C^2+A^2)\mathbf{O}_4 + 2(4AC - n(A^2+C^2))\mathbf{O}_2,\\ \Box\mathbf{O}_3 &= n(C+A)^2\mathbf{O}_5 + 4(C+A)^2\mathbf{O}_4 - 2(C+A)^2\mathbf{O}_3,\\ \Box\mathbf{O}_4 &= -na^{-2}\mathbf{O}_6 - n(C^2+A^2)\mathbf{O}_4 + 2(C^2+A^2)\mathbf{O}_1, \qquad (\mathrm{A4})\\ \Box\mathbf{O}_5 &= 4a^{-2}\mathbf{O}_6 - (n(C^2+A^2) + 2(C+A)^2)\mathbf{O}_5\\ &\quad - 4(C+A)^2\mathbf{O}_4 + 4AC\,\mathbf{O}_3\\ &\quad + 4(n-2)(C+A)^2\mathbf{O}_2,\\ \Box\mathbf{O}_6 &= -n(C^2+A^2)\mathbf{O}_6 + 2a^{-2}\mathbf{O}_1 - na^{-2}\mathbf{O}_4. \end{aligned}$$

We now give a formula for

$$\begin{aligned} J^{abc'd'} &= a^4(\nabla^a\nabla^b - (1/n)g^{ab}\Box)\\ &\quad\times(\nabla^{c'}\nabla^{d'} - (1/n)g^{c'd'}\Box')G, \qquad (\mathrm{A5}) \end{aligned}$$

where G is any function of z. The five coefficient functions for $\mathbf{J}$ are given by

$$\begin{aligned} j^1 &= (1/n^2)[G_{(4)}(z-1)^2z^2 - \tfrac{1}{2}G_{(2)}(-4z^2+4z+n)\\ &\quad + 2G_{(3)}(z-1)z(2z-1)],\\ j^2 &= G_{(4)}(z-1)^2z^2 + 4G_{(3)}(z-1)^2z + 2G_{(2)}(z-1)^2,\\ j^3 &= \tfrac{1}{4}G_{(2)}, \qquad (\mathrm{A6})\\ j^4 &= -(1/n)[G_{(4)}(z-1)^2z^2 + 2G_{(3)}(z-1)z(2z-1)\\ &\quad + 2G_{(2)}(z-1)z],\\ j^5 &= -\tfrac{1}{2}[G_{(3)}(z-1)z + G_{(2)}(z-1)], \end{aligned}$$

where $f_{(i)} \equiv d^if/dz^i$.

The coefficient functions of

$$K^{abc'd'} = a^4\nabla^a\nabla^b\nabla^{c'}\nabla^{d'}G \qquad (\mathrm{A7})$$

are given by

$$\begin{aligned} k^1 &= \tfrac{1}{4}(2z-1)^2G_{(2)} + \tfrac{1}{2}(2z-1)G_{(1)},\\ k^2 &= (z-1)^2z^2G_{(4)} + 4(z-1)^2zG_{(3)}\\ &\quad + 2(z-1)^2G_{(2)}, \qquad (\mathrm{A8})\\ k^3 &= \tfrac{1}{4}G_{(2)},\\ k^4 &= \tfrac{1}{2}(z-1)z(2z-1)G_{(3)} + 2(z-1)zG_{(2)},\\ k^5 &= -\tfrac{1}{2}(z-1)zG_{(3)} - \tfrac{1}{2}(z-1)G_{(2)}. \end{aligned}$$

If two biscalars α and β can be written as functions of $\gamma(z)$ by

$$\begin{aligned} \alpha(\gamma) &= \left[-\frac{2}{n-1}z(1-z)\frac{\partial}{\partial z} + (2z-1)\right]\gamma,\\ \beta(\gamma) &= \alpha(\gamma) - \gamma, \qquad (\mathrm{A9}) \end{aligned}$$

then the components of

$$P^{abc'd'} = a^2\nabla^{(d'}\nabla^{(b}[\alpha g^{a)c')} + \beta n^{a)}n^{c')}] \qquad (\mathrm{A10})$$

are given by

$$p^1 = -\frac{2z-1}{2(n-1)}\frac{\partial\gamma}{\partial z},$$

$$p^2 = -2\frac{z^2(z-1)^2}{n-1}\frac{\partial^3\gamma}{\partial z^3} - 2\frac{(n+2)z(z-1)^2}{n-1}\frac{\partial^2\gamma}{\partial z^2} - 2\frac{n(z-1)^2}{n-1}\frac{\partial\gamma}{\partial z},$$

$$p^3 = \frac{z(z-1)}{2(n-1)}\frac{\partial^2\gamma}{\partial z^2} + \frac{n(2z-1)}{4(n-1)}\frac{\partial\gamma}{\partial z}, \tag{A11}$$

$$p^4 = -\frac{z(z-1)}{n-1}\frac{\partial^2\gamma}{\partial z^2},$$

$$p^5 = -\frac{z^2(z-1)^2}{2(n-1)}\frac{\partial^3\gamma}{\partial z^3} - \frac{z(z-1)(2(n+2)z-(4+n))}{4(n-1)}\frac{\partial^2\gamma}{\partial z^2} - \frac{n(z-1)^2}{2(n-1)}\frac{\partial\gamma}{\partial z}.$$

For the case of general α and β, see Ref. 12.

APPENDIX B: A MODE-FUNCTION COOKBOOK

1. The mode functions

We start with $\{\varphi^i_k\}$, $\{\xi^{ia}_k\}$, and $\{h^{iab}_k\}$, the spin-0, spin-1, and spin-2 eigenfunctions of $\Box$ on the n-sphere.[21] They are defined, for $k=0,1,2,\ldots$, by

$$\Box\varphi^i_k = \lambda^{(0)}_k\varphi^i_k, \tag{B1a}$$

$$\Box\xi^{ia}_k = \lambda^{(1)}_k\xi^{ia}_k, \tag{B1b}$$

$$\Box h^{iab}_k = \lambda^{(2)}_k h^{iab}_k, \tag{B1c}$$

$$g_{ab}h^{iab}_k = \nabla_a h^{iab}_k = \nabla_b h^{iab}_k = \nabla_a\xi^{ia}_k = 0, \quad \text{for all allowed } k,i. \tag{B1d}$$

For $s=0,1,2$, the spin-s modes' eigenvalues are given by[21]

$$\lambda^{(s)}_k = -[k^2+(n+2s-1)k+s(s+n-2)]a^{-2}. \tag{B2}$$

The general multiplicities of these modes can be found in Ref. 22; those used in the paper are given as needed. The modes are orthogonal in the following senses:

$$\int\varphi^j_m(x)\varphi^i_k(x)dV = \int\xi^{ja}_m(x)\xi^i_{k_a}(x)dV = \int h^{jab}_m(x)h^i_{k_{ab}}(x)dV = \delta^{ji}\delta_{mk}, \tag{B3}$$

where the integrals are taken over the entire n-sphere (as are all integrals in this paper involving mode functions). We will usually drop the degeneracy index i used above; all summations over the spectral index ($k=0,1,2,3,\ldots$ above) will then include an implicit sum over the degenerate modes.

Starting with the scalar and vector eigenfunctions, one can obtain tensor mode functions using the gradient operator and the background metric tensor $\mathbf{g}$ and Ricci scalar R. They are defined by

$$\chi^{ab}_k = n^{-1/2}g^{ab}\varphi_k, \quad k=0,1,2,3,\ldots, \tag{B4}$$

$$W^{ab}_k = w_k(\nabla^a\nabla^b - (1/n)g^{ab}\Box)\varphi k, \quad k=2,3,4,\ldots,$$

for

$$w_k = [\lambda^{(0)}_k([(n-1)/n]\lambda^{(0)}_n + R/n)]^{-1/2}, \tag{B5}$$

and

$$V^{ab}_k = [-\tfrac{1}{2}(\lambda^{(1)}_k + R/n)]^{-1/2}\nabla^{(a}\xi^{b)}_k, \quad n=1,2,3,\ldots. \tag{B6}$$

The χ's will be referred to as the "pure-trace" tensor modes; the traceless **W** and **V** modes will be called the longitudinal and shear tensor modes. These modes are orthonormal in the same sense as the h^{iab}_k's, that is,

$$\int\chi^{jab}_m(x)\chi^i_{k_{ab}}(x)dV = \int W^{jab}_m(x)W^i_{k_{ab}}(x)dV = \int V^{jab}_m(x)V^i_{k_{ab}}(x)dV = \delta^{ji}\delta_{mk}. \tag{B7}$$

The four types of modes χ, **W**, **V**, and **h** are orthogonal to each other in this same sense.

2. The cookbook

It will be necessary to calculate $\Box\mathbf{V}_k$ and $\Box\mathbf{W}_k$, as the eigenvalues of these modes are shifted away from those of the eigenfunctions from which they were constructed. One finds

$$\Box\mathbf{V}_k = (\lambda^{(1)}_k + (n+1)a^{-2})\mathbf{V}_k, \tag{B8}$$

and

$$\Box\mathbf{W}_k = (\lambda^{(0)}_k + 2na^{-2})\mathbf{W}_k. \tag{B9}$$

We now look at expressions of the form

$$(g^{ab}\nabla^c\nabla^d + g^{cd}\nabla^a\nabla^b)\,T_{cd},$$

where T will be the various tensor mode functions. One finds

$$(g^{ab}\nabla^c\nabla^d + g^{cd}\nabla^a\nabla^b)h^k_{cd} = (g^{ab}\nabla^c\nabla^d + g^{cd}\nabla^a\nabla^b)V^k_{cd} = 0 \tag{B10}$$

and

$$(g^{ab}\nabla^c\nabla^d + g^{cd}\nabla^a\nabla^b)\begin{pmatrix} W^k_{cd} \\ \chi^k_{cd}\end{pmatrix} = \begin{pmatrix} 0 & n^{1/2}w_k^{-1} \\ n^{1/2}w_k^{-1} & 2\lambda^{(0)}_k\end{pmatrix}\begin{pmatrix} W^{ab}_k \\ \chi^{ab}_k\end{pmatrix}, \tag{B11}$$

where a simple vector/matrix notation has been introduced to exhibit better the W-χ mixing present.

The remaining interesting expressions are

$$\nabla^{(a}g^{b)(c}\nabla^{d)}h^k_{ab} = 0, \tag{B12}$$

$$\nabla^{(a}g^{b)(c}\nabla^{d)}V^k_{cd} = \tfrac{1}{2}(\lambda^{(1)}_k + (n-1)a^{-2})V^{ab}_k, \tag{B13}$$

and

$$\nabla^{(a}g^{b)(c}\nabla^{d)}\begin{pmatrix} W^k_{cd} \\ k\chi_{cd}\end{pmatrix} = \begin{pmatrix} \lambda^{(0)-1}_k w_k^{-2} & n^{-1/2}w_k^{-1} \\ n^{-1/2}w_k^{-1} & \lambda^{(0)}_k n^{-1}\end{pmatrix}\begin{pmatrix} W^{ab}_k \\ \chi^{ab}_k\end{pmatrix}. \tag{B14}$$

It is interesting to note that the matrices appearing in (B11) and (B14) are symmetric; this will allow the graviton action to be inverted in terms of our mode functions even when it includes W-χ mixing terms.

APPENDIX C: VARIANTS OF THE SCALAR AND VECTOR PROPAGATORS

1. The scalar propagator

In Ref. 4, the scalar propagator $G(m^2;x,x')$ was found for de Sitter and anti-de Sitter spaces as a function of $\mu(x,x')$. As it obeys the equation of motion

$$(-\Box+m^2)G(m^2;x,x')=\delta^n(x,x'), \tag{C1}$$

it can also easily be found in terms of the scalar eigenfunctions of $\Box$. One obtains

$$G(m^2;x,x')=\sum_{k=0}^{\infty}\frac{\varphi_k(x)\varphi_k(x')}{-\lambda_k^{(0)}+m^2}. \tag{C2}$$

The equivalence of the mode-sum and function-of-μ forms of the scalar and vector propagators is crucial to the calculations in this paper. The mode sums arise naturally from Gaussian functional integrals,[13] but evaluation of them as functions of the geodesic separation here only proceeds using this equivalence.

At different points we will need different incomplete sums over scalar modes. We therefore introduce

$$\widetilde{G}(l,m^2,z)\equiv\sum_{k=l}^{\infty}\frac{\varphi_k(x)\varphi_k(x')}{-\lambda_k^{(0)}+m^2}. \tag{C3}$$

We will need the second term appearing in the full mode sum, which can be written as

$$\phi_1(m^2,\mu)=\sum_{i=1}^{n+1}\frac{\varphi_1^i(x)\varphi_1^i(x')}{-\lambda_1^{(0)}+m^2}. \tag{C4}$$

Since the $(n+1)$ modes $\{\varphi_1^i(x)\}$ correspond to the coordinates of the n-sphere in the embedding space $\mathbf{R}^{n+1}$,[13,22] one obtains

$$\sum_{i=1}^{n+1}\varphi_1^i(x)\varphi_1^i(x')\propto \mathbf{x}\cdot\mathbf{x}'=a^2\cos(\mu/a)$$
$$=a^2(2z-1). \tag{C5}$$

Using the orthonormality of the modes, one then obtains

$$\phi_1(m^2,z(\mu))=[(n+1)V_n^{-1}/(m^2+na^{-2})](2z-1), \tag{C6}$$

where V_n is the n-sphere volume given in Appendix A.

2. The vector propagator

The vector propagator $\mathbf{G}_{\text{spin-1}}$ obeys the equation of motion

$$(-\Box g^{ab}+R^{ab}+\nabla^a\nabla^b+m^2g^{ab})G_{bc'}^{\text{spin-1}}=\delta^n(x,x')g^a{}_{c'}. \tag{C7}$$

One can invert (C7) to obtain a mode-sum expression for $\mathbf{G}_{\text{spin-1}}$ (see Ref. 12):

$$G_{\text{spin-1}}^{ac'}(m^2;x,x')=\sum_{k=0}^{\infty}\frac{\xi_k^a(x)\xi_k^{c'}(x')}{-\lambda_k^{(1)}+R/n+m^2}+\left(\frac{1}{m^2}\right)\nabla^a\nabla^{c'}\sum_{n=1}^{\infty}\frac{\varphi_k(x)\varphi_k(x')}{-\lambda_k^{(0)}}. \tag{C8}$$

This solution exhibits the well-known divergence of the massless vector propagator in the absence of a gauge-fixing term.[9]

We now define

$$Q^{ac'}(m^2,z)\equiv\sum_{k=0}^{\infty}\frac{\xi_k^a(x)\xi_k^{c'}(x')}{-\lambda_k^{(1)}+R/n+m^2}$$
$$=G_{\text{spin-1}}^{ac'}(m^2;x,x')-(1/m^2)\nabla^a\nabla^{c'}\widetilde{G}(1,0,z). \tag{C9}$$

In Ref. 4, an expression for the vector propagator $G_{\text{spin-1}}^{ac'}(m^2,z)$ in a maximally symmetric space was found. It has the form

$$G_{\text{spin-1}}^{ac'}(m^2,z)=\alpha(\gamma)g^{ac'}+\beta(\gamma)n^an^{c'}, \tag{C10}$$

where the coefficient functions are as given in (A9), and $\gamma(m^2,z)$ obeys

$$z(1-z)\frac{d^2\gamma}{dz^2}+\left(\frac{n}{2}+1\right)(1-2z)\frac{d\gamma}{dz}+\left(m^2+\frac{2R}{n}\right)\gamma=0.$$

The de Sitter and anti-de Sitter forms for γ will, in general, be different, as they will obey different boundary conditions. Here we need only note that γ has a simple pole at $m^2=-2R/n$ in both cases.

In the text, we need to evaluate the mode sum

$$\widetilde{Q}^{ac'}(m^2;x,x')=\sum_{k=1}^{\infty}\frac{\xi_k^a(x)\xi_k^{c'}(x')}{-\lambda_k^{(1)}+R/n+m^2}. \tag{C11}$$

We can relate this incomplete sum over vector modes to a term in an expansion of $\gamma(-2R/n,z)$. Near the mass-squared value

$$m_0^2=\lambda_0^{(1)}-R/n=-2R/n, \tag{C12}$$

the $k=0$ term $\mathbf{Q}-\widetilde{\mathbf{Q}}$ diverges as a simple pole. Since this is away from $m^2=0$, examination of (C9) reveals that this pole must come from the $\mathbf{G}_{\text{spin-1}}$ contribution to $\mathbf{Q}$. Therefore, we expand $G_{\text{spin-1}}^{ac'}$ in powers of $\epsilon=m^2-m_0^2$. For small ϵ, one has

$$G_{\text{spin-1}}^{ac'}(\epsilon+m_0^2,z)=\sum_{j=-1}^{\infty}K_j^{ac'}\epsilon_j; \tag{C13}$$

one also has

$$\gamma(\epsilon+m_0^2,z)=\sum_{l=-1}^{\infty}\gamma_l(z)\epsilon^l. \tag{C14}$$

Our expression for the spin-1 propagator [(A9) and (C10)] then implies that

$$K_j^{ac'}=\alpha(\gamma_j)g^{ac'}+\beta(\gamma_j)n^an^{c'}=G_{\text{spin-1}}^{ac'}(\gamma_j). \tag{C15}$$

Removing a pole in γ removes an equivalent pole in $\mathbf{G}_{\text{spin-1}}$. This means that

$$\widetilde{Q}^{ac'}(m_0^2,z)=\mathscr{G}_{(1)}^{ac'}-(1/m_0^2)\nabla^a\nabla^{c'}\widetilde{G}(1;0,z), \tag{C16}$$

where

$$\mathscr{G}_{(1)}^{ac'}\equiv G_{\text{spin-1}}^{ac'}(\gamma_0)=\alpha(\gamma_0)g^{ac'}+\beta(\gamma_0)n^an^{c'}.$$

APPENDIX D: USEFUL HYPERGEOMETRIC IDENTITIES

We need to find a particular solution to (4.21). It is possible to rewrite this equation as

$$HW_p = H\{[Az^2 + Bz + C]f(a_2+1,b_2-1;c_2;z)\}$$
$$+ [Dz^2 + Ez]f(a_2+1,b_2-1;c_2;z),$$

where $H = H(a_2,b_2;c_2;z)$, and A, B, C, D, and E are nonzero constants. We therefore need to find

$$H^{-1}\{[Dz^2+Ez]f(a_2+1,b_2-1;c_2;z)\},$$

for

$$H(a_2,b_2;c_2;z)H^{-1}f(z) \equiv f(z). \tag{D1}$$

We first notice that $b_2 - 1 = 1$; this allows us to write the remaining piece of the source term in terms of our modified hypergeometric function defined in (3.4) using

$$z^\sigma f(a,1;c;z) = f(a-\sigma,1;c-\sigma;z) - \sum_{l=0}^{\sigma-1} \frac{\Gamma(a-\sigma+l)}{\Gamma(c-\sigma+1)} z^l, \tag{D2}$$

which holds if $a,c > \sigma$.

The rest of the source term can then be found using the contiguous relations for hypergeometric functions and the following identities:

$$H^{-1}\{zf(a_2+1,b_2-1;c_2;z)\}$$
$$= \tfrac{1}{2}[H^{-1}f(a_2,b_2;c_2;z) - f(a_2-1,b_2-1;c_2-1;z)]$$
$$+ (a_2-c_2)H^{-1}f(a_2-1,b_2-1;c_2-1;z)$$
$$- \Gamma(a_2)/\Gamma(c_2-1)]H^{-1}1, \tag{D3}$$

and

$$H^{-1}\{z^2 f(a_2+1,b_2-1;c_2;z)\}$$
$$= \tfrac{1}{2}[H^{-1}f(a_2,b_2;c_2;z) - f(a_2-1,b_2-1;c_2-1;z)]$$
$$- H^{-1}f(a_2-1,b_2-1;c_2-1;z)$$
$$- \frac{\Gamma(a_2-1)}{\Gamma(c_2-2)} H^{-1}\left[1 + \frac{a_2-1}{c_2-2} z\right]. \tag{D4}$$

Again, these are valid because our $b_2 = 2$; they can be simplified using the more general results below.

Generalizing H^{-1} to

$$H(a,b;c;z)H^{-1}f(z) = f(z),$$

we find

$$H^{-1}1 = -1/ab, \tag{D5}$$

$$H^{-1}z = [-1/(a+b+1)][z + c/ab],$$
$$\text{for} \quad a,b,(a+b+1) \neq 0. \tag{D6}$$

Note that any function $J(z)$ specified by

$$J(z) = H^{-1}(a,b;c;z)j(z)$$

is defined uniquely only up to solutions of the homogeneous equation $H(a,b;c;z)k(z) = 0$. We will drop these solutions when apparent.

For $a \neq b$, one finds

$$H^{-1}f(a,b;c;z)$$
$$= [1/(b-a)][\partial_{a'} - \partial_{b'}]f(a',b';c;z)|_{\substack{a'=a\\ b'=b}}. \tag{D7}$$

When a,b are integers and $a > b$ (as is the case for a_2,b_2 when n is an integer) we can write

$$H^{-1}f(a,b;c;z)$$
$$= \sum_{l=0}^{a-b-1} \frac{z^{-(b+l)}}{b-a} \int_0^z y^{b+l-1} f(a,b;c;y)dy. \tag{D8}$$

$$H^{-1}f(a,b;c;z)$$
$$= \sum_{l=0}^{a-b-1} \frac{z^{-(b+l)}}{b-a} \int_0^z y^{b+l-1} f(a,b;c;y)dy. \tag{D8}$$

Finally, we have

$$H^{-1}f(a-1,b-1;c-1;z) = -\tfrac{1}{2}f(a-1,b-1;c;z), \tag{D9}$$

if $a + b + 1 = 2c$, a condition met by most of the hypergeometric functions used in this paper. This is so because the differential operator

$$A(\mu)\partial_\mu = (1/2a)(1-2z)\partial_z$$

appears in every equation of motion used here.

[1]A. Guth, Phys. Rev. D **23**, 347 (1981); R. H. Brandenburger, Rev. Mod. Phys. **57**, 1 (1985); *The Very Early Universe*, edited by G. W Gibbons, S. W. Hawking, and S. T. C. Siklos (Cambridge U.P., Cambridge, 1983).

[2]I. Antoniadis, J. Illiopoulos, and T. N. Tomaras, Phys. Rev. Lett. **56**, 1319 (1986).

[3]I. Antoniadis and E. Mottola, "Graviton fluctuations in de Sitter space," CERN-TH. 4605, 1986.

[4]B. Allen and T. Jacobson, Commun. Math. Phys. **103**, 669 (1986).

[5]S. J. Avis, C. J. Isham, and D. Storey, Phys. Rev. D **18**, 3565 (1978).

[6]S. W. Hawking, Phys. Lett. B **126**, 175 (1983).

[7]M. Turyn, "Boundary conditions for the graviton propagator in anti-de Sitter space," in preparation.

[8]M. Turyn, Phys. Rev. D **39**, 2951 (1989).

[9]N. D. Birrell and P. C. W. Davies, *Quantum Fields in Curved Space* (Cambridge U.P., Cambridge, 1982).

[10]S. W. Hawking and G. F. W. Ellis, *The Large Scale Structure of Space-Time* (Cambridge U.P., London, 1980).

[11]N. H. Barth and S. M. Christensen, Phys. Rev. D **28**, 1876 (1976).

[12]B. Allen and M. Turyn, Nucl. Phys. B **292**, 813 (1987).

[13]B. Allen, Phys. Rev. D **34**, 3670 (1986).

[14]B. Allen, Phys. Rev. D **32**, 3136 (1985).

[15]R. Jordan, Phys. Rev. D **33**, 444 (1986).

[16]G. 't Hooft and M. Veltman, Ann. Inst. H. Poincaré **20**, 69 (1974).

[17]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series, and Products* (Academic, San Francisco, 1980).

[18]M. Turyn, "Graviton detectors in de Sitter space," in preparation.

[19]B. Allen, Phys. Rev. D **37**, 2078 (1988).

[20]S. Weinberg, *Gravitation and Cosmology: Principles and Applications of the Theory of General Relativity* (Wiley, New York, 1972).

[21]G. W. Gibbons and M. J. Perry, Nucl. Phys. B **146**, 90 (1978).

[22]M. A. Rubin and C. R. Ordóñez, J. Math. Phys. **25**, 2888 (1984).

Phase-spaces and dynamical descriptions of infinite mean-field quantum systems

T. Unnerstall
Institut für Theoretische Physik der Universität Tübingen, Auf der Morgenstelle 14, D-7400 Tübingen, Federal Republic of Germany

(Received 2 February 1989; accepted for publication 27 September 1989)

Following the work of P. Bona [J. Math. Phys. 29, 2223 (1988)], a class of infinite mean-field quantum lattice systems is considered within the framework of algebraic quantum mechanics, and the following question is discussed: Which are the possible dynamical descriptions of the system, i.e., in which representations π of the system's quasilocal C^*-algebra $\mathscr{A}$ do the local time evolutions converge in the thermodynamic limit to an automorphism group of a certain enlarged C^*-algebra $\mathscr{C}_\pi$ containing $\pi(\mathscr{A})$ and (some of) the classical observables of the system in π. To this end we associate a "classical state space" E_π to any representation π (which is a subset of the phase-space $E \subseteq \mathbf{R}^L$ defined by Bona for the largest possible representation), and show how it is characterized by the π-normal states, thereby obtaining a way to determine E_π explicitly in many cases. The answer to the question posed then reads: The limiting (Heisenberg) dynamics exists in π on $\mathscr{C}_\pi$ if and only if the set E_π is invariant under the flow φ_t^Q on E which describes the time evolution of classical observables. Conversely, to any invariant, closed subset B of E, there exists a dynamical description of the quantum system with classical state space B.

I. INTRODUCTION

The rigorous formulation of the dynamics of infinite quantum (lattice) systems in the presence of long range interactions has been a long-standing problem in operator-algebraic quantum mechanics. The physical relevance of such models is assured by the fact that long range forces and the corresponding mean-field Hamiltonians have proven to be excellent approximations to the real many-body systems in a number of situations, notably in the case of superconductivity. Moreover, these models are of interst from a fundamental point of view due to the occurrence of so-called "classical observables," and thus of superselection rules.[1]

In the theoretical efforts to deal with quantum mean-field systems it was realized long ago[2–4] that—in contrast to the situation for short range forces[5]—the dynamics in the thermodynamic limit cannot be formulated as a *-automorphism group of the algebra of quasilocal observables $\mathscr{A}$; i.e., that the description of this time evolution is representation dependent.

While this problem was successfully dealt with in the case of thermal equilibrium,[2,3,6–12] the task of obtaining a general dynamical description of the system is more difficult. Hepp and Lieb[13] first succeeded in establishing a picture of the dynamics of certain macroscopic (intensive) quantities, which they described as a Hamiltonian flow on a "phase-space" $\mathbf{R}^L$. In their ingenious approach, the problem of representation dependence is in a sense circumvented by taking all limits in terms of certain expectation values for a suitable large class of states. A while ago, the problem of defining the limiting dynamics as a group of automorphisms in suitable representations was discussed rather thoroughly by Morchio and Strocchi,[14] but it is only very recently that Bona[15] and independently Duffner and Rieckers[16] have been able to actually *prove* the existence of the limit of the local time evolutions for a class of models in a certain representation $\pi_{\mathscr{G}}$ of $\mathscr{A}$ (a partially universal representation belonging to a folium[17] $F_{\mathscr{G}}$ of states).[18] In Ref. 15, Bona determined explicitly the minimal C^*-algebra $\mathscr{C} \subseteq \mathscr{M}_{\pi_{\mathscr{G}}} := \pi_{\mathscr{G}}(\mathscr{A})''$ generally necessary to define this limiting Heisenberg dynamics as an automorphism group τ^Q on $\mathscr{C}$. Moreover, he developed very elegantly Hepp and Lieb's ideas by constructing a classical (generalized) phase-space E as a compact convex subset of $\mathbf{R}^L$, and a flow φ_t^Q on E, with the help of which the dynamics of the classical observables of the system (in the center of the von Neumann algebra $\mathscr{M}_{\pi_{\mathscr{G}}}$) can be described. Let us say that we consider Bona's work a milestone in the field, since it completely clarifies the general structure of the mean-field dynamics in the above class of infinite quantum systems.

In the present article, our purpose is to supplement this work by investigating possible dynamical subdescriptions of the models under consideration. The description Bona develops is maximal in the sense that he works with the largest representation $\pi_{\mathscr{G}}$ in which the limiting dynamics can (in general) possibly be defined; it is therefore natural to ask which representations quasicontained in $\pi_{\mathscr{G}}$ also allow the definition of the (Heisenberg) time evolution of the infinite system. An answer to this question is interesting for two reasons: First, often one is interested in studying the system under well-defined environmental conditions (e.g., a range of temperatures) that reduce its possible states, so that a correspondingly smaller representation is desirable from a pragmatic point of view.[19] Second, there are conceptual problems that call for smaller representations: e.g., one can show (compare Ref. 16) that in the maximal representation $\pi_{\mathscr{G}}$ the implementation (Ref. 5, Def. 3.2.54) of the generator of τ^Q by a self-adjoint operator—which could then be considered the global energy operator of the system in a certain sense—is not possible. This is due to lacking continuity properties. Thus, for the purpose of defining a global Hamiltonian, one is force to look at substructures of $\pi_{\mathscr{G}}$.

The answer we give to the problem of dynamical subdescriptions essentially has two ingredients. After a brief sketch of Bona's theory we associate a classical state space $E_\pi \subseteq E$ to each representation $\pi \leqslant \pi_{\mathscr{G}}$ ($\leqslant$ is the ordering of quasi-equivalence classes of representations) as the state-space of the algebra of macroscopic, classical observables $\mathscr{N}_\pi$ in π, and we characterize it with the help of the states in F_π (the folium associated to π), $F_\pi \subseteq F_{\mathscr{G}}$. We show that, in many important cases, this characterization allows the explicit determination of E_π. It also enables us to derive a conceptually interesting description of the subfolium structure of the folium $F_{\mathscr{G}}$: the map $F_\pi \to E_\pi$ is a surjective, order-preserving $\vee$-homomorphism between complete distributive lattices.[20] In other words, $F_{\mathscr{G}}$ is naturally classified in terms of the closed subsets of E, i.e., in terms of possible "classical structures." These connections also imply that in each such class there is a unique maximal element.

As a second step, we show that the existence of the limiting Heisenberg dynamics in a representation π with the property that it acts [besides on $\pi(\mathscr{A})$] also on the macroscopic observables $\mathscr{N}_\pi$, depends solely on E_π: it exists if and only if E_π is invariant under the classical flow φ_t^Q.

While the question of possible dynamical descriptions is thus answered quite comprehensively, in the sense that one knows in principle all representations which allow the definition of the time evolution of the infinite system with "macroscopic part," it needs to be emphasized that this result is restricted to the Heisenberg picture in which we work throughout. Hence, as we indicate at the end of the paper, one important further question is left open here: the existence of *Schrödinger* dynamics in a given representation (compare Refs. 14, 21). This problem will be addressed elsewhere.[22]

II. THE DYNAMICS OF A CLASS OF QUANTUM MEAN-FIELD SYSTEMS

Before we start with the discussion, let us briefly agree on some notation. Of central importance in our reasoning is the one-to-one correspondence between folia in $S(\mathscr{A})$ (the state space of $\mathscr{A}$), quasi-equivalence classes of representations of $\mathscr{A}$, and central projections in the universal von Neumann algebra $\mathscr{M}_u$, which preserves the partial ordering in these three sets.[17] Thus, to any representation π we associate a folium F_π and a central projection $c(\pi)$, and, conversely, to any folium F a representation π_F [up to quasi-equivalence; we have $\pi_F := \oplus_{\omega\in F}\pi_\omega$, where π_ω is the GNS-representation (the canonical cyclic representation) of $\omega\in S(\mathscr{A})$] and a central projection $c(\pi_F)$. Furthermore, for a representation π we define the von Neumann algebra $\mathscr{M}_\pi := \pi(\mathscr{A})'' = c(\pi)\mathscr{M}_u$ with center $\mathscr{Z}_\pi = c(\pi)\mathscr{Z}_u$, and the Hilbert space $\mathscr{H}_\pi := \oplus_{\omega\in F_\pi}\mathscr{H}_\omega = c(\pi)\mathscr{H}_u$. Here, $\mathscr{Z}_u$ is the center of $\mathscr{M}_u$, and $\mathscr{H}_u$ is the Hilbert space of the universal representation π_u of $\mathscr{A}$. We note also the one-to-one correspondence between the normal states on $\mathscr{M}_\pi$, and the folium F_π.

To make the paper self-contained, we shall now sketch the theory developed in Ref. 15 to the extent necessary for our present purposes; for more information and additional references, we refer the reader to the original paper.

The models under consideration are defined by three elements:

(1) The C^*-algebra

$$\mathscr{A} := \bigotimes_{n\in\mathbb{N}} \mathscr{A}_n, \quad \mathscr{A}_n = B(\mathbb{C}^m) = M_m \quad \forall n\in\mathbb{N},$$

which can be defined as the C^*-inductive limit of the local algebras

$$\mathscr{A}(\Lambda) := \bigotimes_{n\in\Lambda} \mathscr{A}_n, \quad \Lambda\in\mathscr{L} := \{\Lambda\subseteq\mathbb{N} : |\Lambda| < \infty\}.$$

Here, M_m denotes the complex $m\times m$ matrices, and $|\Lambda|$ the cardinality of Λ. $\mathscr{A}$ is simple and antiliminary,[23] yielding a very rich structure of the state space $S(\mathscr{A})$.

(2) A real Lie algebra $\mathscr{G}$ of dimension L (taken to be the Lie algebra of a compact Lie group G), which is represented on $\mathbb{C}^m$ by anti-Hermitian operators

$$(1/i)X(\beta), \quad \text{where } X(\beta)\in(M_m)_{\text{s.a.}} \quad \forall\beta\in\mathscr{G}.$$

This Lie algebra is chosen as the smallest Lie algebra, giving all elements of $(M_m)_{\text{s.a.}}$ which occur in the local Hamiltonians below. It can always be taken as $\mathscr{G}\subseteq\mathrm{Lie}(\mathrm{SU}(m))$, since $\mathrm{Lie}(\mathrm{SU}(m))$ and $(M_m)_{\text{s.a.}}$ differ only by $1\in M_m$.

(3) Local Hamiltonians

$$H^\Lambda := |\Lambda| Q(X_\Lambda(\beta^1),\ldots,X_\Lambda(\beta^L)),$$

where $Q: \mathbb{R}^L\to\mathbb{R}$ is an arbitrary, nonlinear polynomial. The density observables

$$X_\Lambda(\beta^k) := \frac{1}{|\Lambda|}\sum_{n\in\Lambda} X_n(\beta^k), \quad k = 1,\ldots,L,$$

$$X_n(\beta^k) = 1\otimes\cdots\otimes 1\otimes \underbrace{X(\beta^k)}_{\text{position } n}\otimes 1\otimes\cdots\otimes 1 \quad \in\mathscr{A}(\Lambda)$$

are defined by means of the elements

$$X(\beta^k)\in(M_m)_{\text{s.a.}};$$

$\beta^1,\ldots,\beta^L$ is a suitably chosen basis of the vectorspace $\mathscr{G}\cong\mathbb{R}^L$. In this situation, one is looking for a suitable limit $\Lambda\to\infty$ of the local time evolutions

$$\tau_t^\Lambda(x) := e^{itH^\Lambda} x e^{-itH^\Lambda}, \quad x\in\mathscr{A}.$$

It is shown in Ref. 15 (compare Ref. 14) that such limits always involve limits,

$$\lim_\Lambda X_\Lambda(\beta^k), \quad k\in\{1,\ldots,L\},$$

that do not exist in the norm topology of $\mathscr{A}$.

Therefore, one defines the largest representation $\pi_{\mathscr{G}}$ of $\mathscr{A}$ in which all such limits exist. This can be done by means of the folium

$$F_{\mathscr{G}}:=\{\omega\in S(\mathscr{A}):\ \text{stop}-\lim_{\Lambda}\pi_\omega(X_\Lambda(\beta)) = \text{ex.}\quad \forall\beta\in\mathscr{G}\};$$

the corresponding central projection $c(\pi_{\mathscr{G}})$ in $\mathscr{M}_u$ can be defined as

$$c(\pi_{\mathscr{G}}):=\sup\{P\in\mathscr{Z}_u, P \text{ projection: stop} -\lim_{\Lambda}\pi_u(X_\Lambda(\beta))P=\text{ex.}\quad \forall\beta\in\mathscr{G}\}.$$

The "kinematical structure"[5,15] in which the dynamical theory is defined is thus

$$(\mathscr{A}\cong)\pi_{\mathscr{G}}(\mathscr{A})\subseteq\pi_{\mathscr{G}}(\mathscr{A})''=:\mathscr{M}_{\mathscr{G}}\subseteq\mathscr{B}(\mathscr{H}_{\mathscr{G}}),$$

$$\mathscr{H}_{\mathscr{G}}:=c(\pi_{\mathscr{G}})\mathscr{H}_u=\bigoplus_{\omega\in F_{\mathscr{G}}}\mathscr{H}_\omega.$$

In this setting, one defines

$$X_{\mathscr{G}}(\beta):=\text{stop}-\lim_{\Lambda}\pi_{\mathscr{G}}(X_\Lambda(\beta))\in\mathscr{Z}_{\mathscr{G}}$$

and

$$\mathscr{N}:=C^\star\text{-alg. hull}\{X_{\mathscr{G}}(\beta),\ \beta\in\mathscr{G}\}\subseteq\mathscr{Z}_{\mathscr{G}}.$$

Note that $\mathscr{N}$ is obviously the smallest $C^\star$-algebra containing all macroscopic observables $X_{\mathscr{G}}(\beta)$; it can be looked at as the classical part of the description,[1] due to its being part of the center of $\mathscr{M}_{\mathscr{G}}$.

One of the important ideas in Ref. 15 is to use the SNAG-theorem to obtain from the representation

$$\mathscr{G}\to\mathscr{Z}_{\mathscr{G}},\quad \beta\to e^{iX_{\mathscr{G}}(\beta)}$$

of the additive group $\mathscr{G}$ a unique, projection-valued measure

$$\mathscr{E}_{\mathscr{G}}:\ \mathscr{B}(\mathscr{G}^\star)\to\mathscr{P}(\mathscr{Z}_{\mathscr{G}}),$$

where $\mathscr{B}(\mathscr{G}^\star)$ is the Borel σ-algebra of $\mathscr{G}^\star$, the vector space dual of $\mathscr{G}$, such that

$$X_{\mathscr{G}}(\beta)=\int_{\mathscr{G}^\star}f_\beta(F)d\mathscr{E}_{\mathscr{G}}(F)\quad\forall\beta\in\mathscr{G}\tag{1}$$

holds in the wop-topology (weak operator topology) on $\mathscr{H}_{\mathscr{G}}$. In (1), f_β is the element of the bidual $\mathscr{G}^{\star\star}$ corresponding to $\beta\in\mathscr{G}$ via the canonical isomorphism, i.e., $f_\beta(F):=F(\beta)\ \ \forall F\in\mathscr{G}^\star$.

If one lets $E:=\text{supp}\ \mathscr{E}_{\mathscr{G}}\subseteq\mathscr{G}^\star$, which is compact, and denotes with $C(E)$ the $C^\star$-algebra of continuous functions on E, one has the following result.

Theorem 1: (1) $\mathscr{N}\cong C(E)$ via the *-isomorphism

$$\mathscr{E}_{\mathscr{G}}:\ C(E)\to\mathscr{N},\qquad f\to\int_E f(F)d\mathscr{E}_{\mathscr{G}}(F),\qquad f_\beta\to X_{\mathscr{G}}(\beta).$$

(2) For $\mathscr{C}:=C^\star\text{-alg. hull}\{\pi_{\mathscr{G}}(\mathscr{A}),\mathscr{N}\}$ it holds

$$\mathscr{C}=\pi_{\mathscr{G}}(\mathscr{A})\otimes\mathscr{N}\cong\mathscr{A}\otimes C(E).$$ ■

The Gelfand representation theorem shows that E can be considered to be the pure states on $\mathscr{N}$, the algebra of classical observables. It is therefore justified to call E the (generalized) classical phase-space of the system. It depends on $\mathscr{G}$ as well as on the representation $\pi_{\mathscr{G}}$. Note that E can be characterized further: from Prop. 2.4 in Ref. 24 it follows that [$S(M_m)$ denotes the state-space of M_m]:

$$E=\{F\in\mathscr{G}^\star:\ \exists\rho\in S(M_m) \text{ with } \rho(X(\beta))=F(\beta)\ \ \forall\beta\in\mathscr{G}\}.\tag{2}$$

In particular, in the case $\mathscr{G}=\text{Lie}(SU(m))$, one has $E\cong S(M_m)$.

Turning now to the description of the dynamics, one observes that the polynomial Q: $\mathbf{R}^L\to\mathbf{R}$ can be considered an element of $C^\infty(\mathscr{G}^\star)$, since $\mathscr{G}\cong\mathscr{G}^\star\cong\mathscr{G}^{\star\star}\cong\mathbf{R}^L$ as vector spaces. It induces a flow $\{\varphi_t^Q\}_{t\in\mathbf{R}}$ on $E\subseteq\mathscr{G}^\star$ via the vector field

$$(\lambda_Q(F^1,...,F^L))^k=\sum_{j=1}^L\frac{\partial Q}{\partial F_j}(F^1,...,F^L)\left(\sum_{l=1}^L C_l^{kj}F^l\right),$$

where C_l^{kj} are the Lie structure constants of $\mathscr{G}$ with respect to the basis $\beta^1,...,\beta^L$, and $F^1,...,F^L$ the coordinates with respect to the corresponding dual basis of $\mathscr{G}^\star$.

We denote with $(\varphi_t^Q)^\star$ the induced *-automorphism group of $C(E)$: $(\varphi_t^Q)^\star(f):=f\circ\varphi_t^Q$, $f\in C(E)$.

The main result of Ref. 15 is then the following.

Theorem 2: There is a unique *-automorphism group

$$\tau_t^Q:\ \mathscr{C}\to\mathscr{C},$$

such that for "sufficiently small" $t\in\mathbf{R}$:

$$(1)\ \ \tau_t^Q(x)=\text{stop}-\lim_{\Lambda'}\pi_{\mathscr{G}}(\tau_t^{\Lambda'}(x))\ \ \forall x\in\mathscr{A}(\Lambda),\ \ \forall\Lambda\in\mathscr{L},$$

$$(2)\ \ \tau_t^Q(X_{\mathscr{G}}(\beta))=\text{stop}-\lim_{\Lambda}\pi_{\mathscr{G}}(\tau_t^\Lambda(X_\Lambda(\beta)))\ \ \forall\beta\in\mathscr{G}.$$

In addition, it holds

$$(3)\ \ \tau_t^Q|_{\mathscr{N}}=\mathscr{E}_{\mathscr{G}}\circ(\varphi_t^Q)^\star\circ\mathscr{E}_{\mathscr{G}}^{-1},$$

(4) τ_t^Q is strongly continuous in t on $\mathscr{C}$. ■

Our aim in this paper is to find all representations $\pi\leqslant\pi_{\mathscr{G}}$ such that Theorem 2 holds in π.

III. KINEMATICAL SUBDESCRIPTIONS

In this section, we consider an arbitrary subrepresentation $\pi\leqslant\pi_{\mathscr{G}}$, or, equivalently, a subfolium $F_\pi\subseteq F_{\mathscr{G}}$; we have then $c(\pi)\in\mathscr{Z}_{\mathscr{G}}$, and $\pi(x)=c(\pi)\pi_{\mathscr{G}}(x)\ \forall x\in\mathscr{A}$. All structures of the previous section can be developed for π in an analogous way. We define

$$X_\pi(\beta):=\text{stop}-\lim_{\Lambda}\pi(X_\Lambda(\beta))\ \in\mathscr{Z}_\pi,\quad\beta\in\mathscr{G},$$

and $\mathscr{N}_\pi\subseteq\mathscr{Z}_\pi$, the $C^\star$ algebra generated by these operators. The SNAG measure belonging to this representation, $\mathscr{E}_\pi$, and its support $E_\pi\subseteq\mathscr{G}^\star$ give rise to the relation

$$X_\pi(\beta)=\int_{E_\pi}f_\beta(F)d\mathscr{E}_\pi(F),$$

and to the *-isomorphism

$$\mathscr{E}_\pi: C(E_\pi) \to \mathscr{N}_\pi,$$
$$f \to \int_{E_\pi} f(F) d\mathscr{E}_\pi(F),$$
$$1 \to \mathscr{E}_\pi(E_\pi) = c(\pi).$$

The connection of these structures, which will henceforth be called a "kinematical subdescription" of the system, to the objects of Sec. II is established in the following easy lemma.

Lemma 3: One has

(i) $\mathscr{E}_\pi = c(\pi)\cdot\mathscr{E}_{\mathscr{G}}$ as a (projection-valued) measure on $\mathscr{G}^\star$,

(ii) $E_\pi \subseteq E$,

(iii) $\mathscr{N}_\pi = c(\pi)\mathscr{N}$, $\mathscr{C}_\pi := \pi(\mathscr{A}) \otimes \mathscr{N}_\pi = c(\pi)\mathscr{C}$.

Proof: We have, due to $\pi = c(\pi)\pi_{\mathscr{G}}$,

$$X_\pi(\beta) = X_{\mathscr{G}}(\beta)\cdot c(\pi)$$
$$= \left[\int_E f_\beta(F) d\mathscr{E}_{\mathscr{G}}(F)\right]\cdot c(\pi)$$

and, due to the continuity of multiplication in the wop topology,

$$X_\pi(\beta) = \int_E f_\beta(F)(d\mathscr{E}_{\mathscr{G}}(F)\cdot c(\pi)) \quad \forall \beta \in \mathscr{G}.$$

Uniqueness of the SNAG measure now gives the desired result.

Part (ii) follows immediately from (i) and the definition of the support, since

$$\mathscr{E}_\pi(E) = c(\pi)\mathscr{E}_{\mathscr{G}}(E) = c(\pi)1_{\mathscr{H}_{\mathscr{G}}} = c(\pi).$$

Part (iii) follows again from $\pi = c(\pi)\pi_{\mathscr{G}}$. ∎

Note that part (i) also implies $c(\pi)\mathscr{E}_{\mathscr{G}}(E_\pi) = \mathscr{E}_\pi(E_\pi) = c(\pi)$, hence

$$c(\pi) \leqslant \mathscr{E}_{\mathscr{G}}(E_\pi). \tag{3}$$

The "classical state space" E_π associated to the representation π plays a crucial role in the subsequent analysis

IV. CHARACTERIZATION OF E_π

Our first aim is to obtain more information about the space E_π; this will be done by deriving a connection between states in the folium F_π defining the representation π, and E_π.

Any $\omega \in F_\pi$ can be considered as a normal state on $\mathscr{M}_\pi$, which will be denoted by the same symbol. The state ω can be restricted to $\mathscr{N}_\pi \subseteq \mathscr{M}_\pi$, and thus defines—via the isomorphism $\mathscr{E}_\pi$—a state μ_ω on $C(E_\pi)$, and hence a probability measure on E_π, again denoted by μ_ω.

As a first step, consider the GNS-representation $\pi_\omega \leqslant \pi \leqslant \pi_{\mathscr{G}}$. One then gets the following.

Proposition 4: It holds for $\omega \in F_\pi$,

(i) $\mu_\omega(A) = \omega(\mathscr{E}_\pi(A)) = \omega(\mathscr{E}_{\pi_\omega}(A))$ $\forall A \in \mathscr{B}(E_\pi)$,

(ii) $\operatorname{supp}\mu_\omega = E_{\pi_\omega} := \operatorname{supp}\mathscr{E}_{\pi_\omega} \subseteq E_\pi$,

(iii) if ω is a factor state on $\mathscr{A}$, $\mu_\omega = \delta_{F_\omega}$, the Dirac measure of a point $F_\omega \in E_\pi$,

(iv) one has

$$\lim_\Lambda \omega(P(X_\Lambda(\beta^1),\ldots,X_\Lambda(\beta^L)))$$
$$= \int_{E_\pi} P(F^1,\ldots,F^L) d\mu_\omega(F^1,\ldots,F^L) = \mu_\omega(P) \tag{4}$$

for all polynomials P on $\mathbf{R}^L$.

Proof: (i) We have $\forall f \subseteq C(E)$:

$$\int_{E_\pi} f(F) d\mu_\omega(F) = \mu_\omega(f)$$
$$= \omega(\mathscr{E}_\pi(f))$$
$$= \int_{E_\pi} f(F)\omega(d\mathscr{E}_\pi(F)),$$

so that $\forall A \in \mathscr{B}(E_\pi)$,

$$\mu_\omega(A) = \omega(\mathscr{E}_\pi(A))$$
$$= \omega(c(\pi_\omega)\mathscr{E}_\pi(A))$$
$$\overset{!}{=} \omega(\mathscr{E}_{\pi_\omega}(A)), \tag{5}$$

since $c(\pi_\omega)$ is the central cover of ω. Note that (!) follows from Lemma 3, part (i).

(ii) From (5) it is clear that $\operatorname{supp}\mu_\omega \subseteq E_{\pi_\omega}$. Assume $\operatorname{supp}\pi_\omega \neq E_{\pi_\omega}$. There is then an open set $A \subseteq E_{\pi_\omega} \setminus \operatorname{supp}\mu_\omega$ such that $\mu_\omega(A) = 0$. Let $c := \mathscr{E}_{\pi_\omega}(A) \leqslant c(\pi_\omega)$; for the central projection $0 \neq c \in \mathscr{M}_\omega := \pi_\omega(A)''$ on $\mathscr{H}_\omega$ one has

$$\mu_\omega(A) = \omega(c) = \langle \Omega_\omega, c\Omega_\omega \rangle$$
$$= \langle c\Omega_\omega, c\Omega_\omega \rangle = 0 \Rightarrow c\Omega_\omega = 0.$$

But then ($-\|\cdot\|$ denotes norm closure)

$$c\mathscr{H}_\omega = c\{x\Omega_\omega: x \in \mathscr{A}\}^{-\|\cdot\|} = \{cx\Omega_\omega: x \in \mathscr{A}\}^{-\|\cdot\|}$$
$$= \{xc\Omega_\omega: x \in \mathscr{A}\}^{-\|\cdot\|} = 0,$$

so $c = 0$, which is a contradiction.

(iii) Let ω be a factor state on $\mathscr{A}$; one has

$$c(\pi_\omega) = \mathscr{E}_{\pi_\omega}(E_{\pi_\omega}).$$

If E_{π_ω} consisted of more than one point, one would have a set $A \subseteq E_{\pi_\omega}$ with

$$0 \neq E_{\pi_\omega}(A) < c(\pi_\omega);$$

this is a contradiction, since $c(\pi_\omega)$ is an atomic central projection.[23] Hence, $E_{\pi_\omega} = \{F_\omega\}$ for a $F_\omega \in E_\pi$, and thus due to (ii),

$$\mu_\omega = \delta_{F_\omega}.$$

(iv) Due to the continuity of multiplication in the stop topology, one has

$$\text{stop} - \lim_\Lambda P(X_\Lambda(\beta^1),\ldots,X_\Lambda(\beta^L))$$
$$= P(X_{\mathscr{G}}(\beta^1),\ldots,X_{\mathscr{G}}(\beta^L))$$
$$= \mathscr{E}_\pi(P(F^1,\ldots,F^L)).$$

On bounded sets, the stop and the σ-weak topology coincide, and since $P(X_\Lambda(\beta^1),\ldots,X_\Lambda(\beta^L))$ is uniformly bounded in Λ, it holds

$$\lim_{\Lambda} \omega(P(X_\Lambda(\beta^1),...,X_\Lambda(\beta^L)))$$
$$= \omega(\sigma - w - \lim_{\Lambda} P(X_\Lambda(\beta^1),...,X_\Lambda(\beta^L)))$$
$$= \omega(\mathscr{C}_\pi(P(F^1,...,F_L)))$$
$$= \mu_\omega(P). \qquad \blacksquare$$

Remarks: (a) It should be noted that μ_ω is not dependent on the representation π; $\omega \in F_\pi \subseteq F_\mathscr{G}$, so ω is also a normal state on $\mathscr{M}_\mathscr{G}$, inducing a measure $\bar{\mu}_\omega$ on E. But according to the representation-independent relation (4) (and the fact that μ_ω is uniquely determined on the norm-dense polynomials), one has $\bar{\mu}_\omega = \mu_\omega$.

(b) The inverse of (iii), the statement that each point in E_π corresponds to a factor state in F_π, is false in general. Indeed, F_π need not contain any factor states at all. Compare also Prop. 2.3 in Ref. 24.

(c) It is well possible that two disjoint factor states ω_1, ω_2 have $F_{\omega_1} = F_{\omega_2}$; this is due to the fact that, in general, $\mathscr{N}''_\pi \neq \mathscr{Z}_\pi$, and disjointness is equivalent to having different values on only one $z \in \mathscr{Z}_\pi$.

(d) Part (iv) should be compared with Ref. 13, where this relation is the starting point of the analysis. In the terminology introduced there, factor states belong to the "pure classical states."

We now proceed to the main result of this section.

Theorem 5: (i) The set $E_\pi \subseteq \mathscr{G}^\star$ can be characterized in the following way (overbar denotes closure):

$$E_\pi = \overline{\bigcup_{\omega \in F_\pi} \operatorname{supp} \mu_\omega}.$$

(ii) If $Q \subseteq S(\mathscr{A})$ is a generating set for F_π, it holds

$$E_\pi = \overline{\bigcup_{\omega \in Q} \operatorname{supp} \mu_\omega}.$$

Proof: (i) It is clear that $B := \overline{\bigcup_{\omega \in F_\pi} \operatorname{supp} \mu_\omega} \subseteq E_\pi$. We assume $B \neq E_\pi$, so that there is an open $A \subseteq E_\pi \setminus B$ with

$$\mu_\omega(A) = 0 \ \forall \omega \in F_\pi \Rightarrow \omega(\mathscr{C}_\pi(A)) = 0 \ \forall \omega \in F_\pi;$$

but since F_π is separating for $\mathscr{M}_\pi$, this implies $\mathscr{C}_\pi(A) = 0$, which is a contradiction to the definition of E_π as the support of $\mathscr{C}_\pi$.

(ii) Q is a generating set for F_π, i.e.,

$$F_\pi = \text{convhull}\{Q_c\}^{\|\cdot\|},$$

where

$$Q_c := \{\psi \in S(\mathscr{A}) : \exists \omega \in S(\mathscr{A}),\ \exists c \in \mathscr{A},\ \omega(c^\star c) = 1,$$
$$\text{such that } \psi = \omega_c := \omega(c^\star \cdot c)\}.$$

We denote $B := \overline{\bigcup_{\omega \in Q} \operatorname{supp} \mu_\omega}$ and show $\operatorname{supp} \mu_\psi \subseteq B \ \forall \psi \in F_\pi$.

(a) Let first $\psi \in Q_c$; $\Rightarrow \psi = \omega_c$, $\omega \in Q$; $\Rightarrow \psi \in F_\omega$, and Prop. 4 now implies $\operatorname{supp} \mu_\psi \subseteq E_{\pi_\omega} = \operatorname{supp} \mu_\omega \subseteq B$.

(b) Let now $\psi \in \text{convhull}\{Q_c\}$, i.e., $\psi = \Sigma_{i=1}^n \lambda_i \omega_i$, $\omega_i \in Q_c$. But that implies

$$\mu_\psi = \sum_{i=1}^n \lambda_i \mu_{\omega_i} \Rightarrow \operatorname{supp} \mu_\psi \subseteq \bigcup_{i=1}^n \operatorname{supp} \mu_{\omega_i} \subseteq B.$$

(c) Finally, consider $\omega \in F_\pi$ arbitrary. It exists a sequence $\omega_n \to \omega$ in the norm topology of $\mathscr{A}^\star$, where $\omega_n \in \text{convhull}\{Q_c\}$ $\forall n \in \mathbb{N}$. The normal extensions of these states on $\mathscr{M}_\pi$ satisfy $\omega_n \to \omega$ in the $\sigma(\mathscr{M}_\pi^\star, \mathscr{M}_\pi)$-topology (because $|\omega_n(x) - \omega(x)| \leqslant \|x\| \, \|\omega_n - \omega\|$ $\forall x \in \mathscr{M}_\pi$) and therefore $\mu_{\omega_n} \to \mu_\omega$ in $\sigma(C(E_\pi)^\star, C(E_\pi))$.

Now all μ_{ω_n} have $\operatorname{supp} \mu_{\omega_n} \subseteq B$ as shown in (b), and if one had $\operatorname{supp} \mu_\omega \not\subseteq B$, one could find a function $f \in C(E_\pi)$ with $\operatorname{supp} f \subseteq \operatorname{supp} \mu_\omega \setminus B$ and $\mu_\omega(f) \neq 0$, the first property implying $\mu_{\omega_n}(f) = 0$ $\forall n \in \mathbb{N}$. This contradicts $\sigma(C(E_\pi)^\star, C(E_\pi))$-convergence. Hence, we must have $\operatorname{supp} \mu_\omega \subseteq B$. $\blacksquare$

With this last theorem, we are in a position to determine, at least in principle, the classical state space of any kinematical subdescription of the system. In particular, consider $S^P(\mathscr{A})$, the permutation invariant states[25] on $\mathscr{A}$. Due to the fact that they exhibit the same symmetry as the local Hamiltonians (which is just the one defining mean-field models!), they are most important for the description of the system: e.g., all equilibrium states are in $S^P(\mathscr{A})$.

The classical state space of any folium generated by a subset $Q \subseteq S^P(\mathscr{A})$ can be explicitly determined. To see this, recall that every $\omega \in S^P(\mathscr{A})$ is defined by a measure ρ_ω on the extremal boundary, $\partial S^P(\mathscr{A})$, which in turns is in one-to-one correspondence to $S(M_m)$. [For every $\varphi \in \partial S^P(\mathscr{A})$ there exists a $\rho \in S(M_m)$ such that $\varphi = \otimes_{n \in \mathbb{N}} \rho$.] Hence, every $\omega \in S^P(\mathscr{A})$ is given by a measure on $S(M_m)$, which, due to Eq. (2), determines a measure $\bar{\rho}_\omega$ on E. It can be shown[22] that it holds $\bar{\rho}_\omega = \mu_\omega$. In this way, one can calculate μ_ω, and consequently $\overline{\bigcup_{\omega \in Q} \operatorname{supp} \mu_\omega} = E_{\pi_F}$, where F is the folium generated by Q (see Ref. 22 for a more detailed discussion).

V. THE STRUCTURE OF $F_\mathscr{G}$

We have seen that any subfolium $F \subseteq F_\mathscr{G}$ determines via its representation π_F a kinematical subdescription and, in particular, a state space $E_{\pi_F} \subseteq E$. It is tempting to formalize this relation. Let

$\mathscr{F} := \{F \subseteq F_\mathscr{G} :\ F \text{ folium}\}$, the set of all subfolia of $F_\mathscr{G}$,

and

$\mathscr{B} := \{B \subseteq E :\ B \text{ closed}\}$, the set of all closed subsets of E.

With the operations $F_1 \vee F_2 := F$, F the folium generated by F_1 and F_2, and $F_1 \wedge F_2 := F_1 \cap F_2$, $\mathscr{F}$ becomes a lattice which is isomorphic to the lattice of central projections in $\mathscr{M}_\mathscr{G}$, $\mathscr{P}(\mathscr{Z}_\mathscr{G})$. Hence, $\mathscr{F}$ is a complete distributive lattice. The same is of course true for $\mathscr{B}$.

We define the map

$$j:\ \mathscr{F} \to \mathscr{B},$$

$$F \to E_{\pi_F} = \overline{\bigcup_{\omega \in F} \operatorname{supp} \mu_\omega}.$$

It is obvious that j is order-preserving; additional properties are contained in the following.

Theorem 6: j is a surjective (but not injective), complete $\vee$-homomorphism, i.e.,

$$j(\bigvee_{i \in I} F_i) = \bigvee_{i \in I} j(F_i)$$

for an arbitrary family $\{F_i\colon i\in I\}\subseteq\mathscr{F}$ of folia.

Proof: (1) First, we show that j is a complete $\vee$-homomorphism. Let $\{F_i\colon i\in I\}\subseteq\mathscr{F}$ be an arbitrary family of folia.

(1.1) In order to prove $j(\vee_{i\in I}F_i)\subseteq\vee_{i\in I}j(F_i)$: $=\overline{\cup_{i\in I}j(F_i)}=:B$, we have to show that $\operatorname{supp}\mu_\omega\subseteq\overline{\cup_{i\in I}j(F_i)}$ for all $\omega\in\vee_{i\in I}F_i$ $=\operatorname{convhull}\{F_i\colon i\in I\}^{-\|\cdot\|}$, since

$$j(\vee_{i\in I}F_i)=\overline{\bigcup_{\omega\in\vee_{i\in I}F_i}\operatorname{supp}\mu_\omega}.$$

Consider first the case $\omega\in\operatorname{convhull}\{F_i\colon i\in I\}$, i.e.,

$$\omega=\sum_{m=1}^{n}\lambda_m\omega_m,\quad\text{with }\sum_m\lambda_m=1,\ \omega_m\in\{F_i\colon i\in I\}.$$

But this means $\mu_\omega=\Sigma_m\lambda_m\mu_{\omega_m}$, hence

$$\operatorname{supp}\mu_\omega\subseteq\bigcup_m\operatorname{supp}\mu_{\omega_m}\subseteq B,$$

since $\operatorname{supp}\mu_{\omega_m}\subseteq j(F_i)\subseteq B$ for some i.

The case where ω is in the norm closure of $\operatorname{convhull}\{F_i\colon i\in I\}$ can be treated as in Theorem 5, part (ii).

(1.2) We have

$$j(\vee_{i\in I}F_i)\supseteq\overline{\bigcup_{\omega\in\cup_{i\in I}F_i}\operatorname{supp}\mu_\omega}$$

$$=\overline{\bigcup_{i\in I}\bigcup_{\omega\in F_i}\operatorname{supp}\mu_\omega}$$

$$=\overline{\bigcup_{i\in I}\overline{\bigcup_{\omega\in F_i}\operatorname{supp}\mu_\omega}}$$

$$=\overline{\bigcup_{i\in I}j(F_i)}.$$

The result $j(\vee_{i\in I}F_i)=\vee_{i\in I}j(F_i)$ follows from (1.1) and (1.2).

(2) Second, we show that j is surjective. Let $B\subseteq E$ be a closed set. We need to find a folium F with $j(F)=B$. To any $\mathrm{F}\in B$ there exists a $\rho_{\mathrm{F}}\in S(M_m)$ with $\rho_{\mathrm{F}}(X(\beta))=\mathrm{F}(\beta)$ [Eq. (2)]. For the state $\varphi_{\mathrm{F}}=\otimes_{n\in\mathbb{N}}\rho_{\mathrm{F}}$, which is a factor state, one has

$$\mu_{\varphi_{\mathrm{F}}}=\delta_{\mathrm{F}},\quad\text{the Dirac measure on F.}$$

This follows from Prop. 4, (iii) and (iv), since

$$\lim_{\Lambda}\varphi_{\mathrm{F}}(X_\Lambda(\beta))\overset{!}{=}\rho_{\mathrm{F}}(X(\beta))=\mathrm{F}(\beta)\quad\forall\beta\in\mathscr{G}.$$

Here, the equality (!) can be directly verified.

According to Theorem 6, one then has

$$j(F_{\pi_{\varphi_{\mathrm{F}}}})=\operatorname{supp}\mu_{\varphi_{\mathrm{F}}}=\{\mathrm{F}\}.$$

Hence, with $F:=\vee_{\mathrm{F}\in B}F_{\pi_{\varphi_{\mathrm{F}}}}$ we get with the help of part (1):

$$j(F)=j(\bigvee_{\mathrm{F}\in B}F_{\pi_{\varphi_{\mathrm{F}}}})=\bigvee_{\mathrm{F}\in B}j(F_{\pi_{\varphi_{\mathrm{F}}}})$$

$$=\overline{\bigcup_{\mathrm{F}\in B}\{\mathrm{F}\}}=B.$$

(3) The missing injectivity of j is already clear from the remark (c) below Prop. 4: if ω_1,ω_2 are disjoint factor states with $\mu_{\omega_1}=\mu_{\omega_2}=\delta_{\mathrm{F}}$, $\mathrm{F}\in E$, one has $j(F_{\omega_1})=j(F_{\omega_2})=\mathrm{F}$ while $F_{\omega_1}\cap F_{\omega_2}=\emptyset$. This example also shows that j cannot be a $\wedge$ homomorphism. ∎

This theorem is interesting from a conceptual point of view: it makes more concrete the idea[1] that the lattice structure of $\mathscr{F}$ [of $\mathscr{P}(\mathscr{Z}_{\mathscr{G}})$, respectively] should be interpreted in the way that folia represent classical properties of the system. Note that j gives, one might say, a "coarse grained" picture of these properties. The surjectivity of j shows—and this might not have been expected—that *every* closed subset of E appears as a classical state space (as the set of pure states on $\mathscr{N}_\pi$) of some representation π.

The fact that j is not injective is not only due to the fact that, deliberately, not all macroobservables are included in the framework set by $\mathscr{G}$, it also has other reasons. Indeed, it follows from the remarks at the end of Sec. IV that it holds $j(F^P)=j(F_{\mathscr{G}})=E$, F^P being the folium generated by $S^P(\mathscr{A})$, while $F^P\subseteq F_{\mathscr{G}}$, $F^P\neq F_{\mathscr{G}}$.

Another way of looking at Theorem 6 is that the set $\mathscr{F}$ is classified in terms of the closed subsets of E. If we define accordingly for every closed $B\subseteq E$ the set

$$\mathscr{F}(B):=\{F\in\mathscr{F}\colon j(F)=B\},$$

we get the following.

Lemma 7: Let $B\subseteq E$ be a closed set. Then it holds:

(i) $\mathscr{F}(B)$ is not empty and contains a unique maximal element, the set

$$F_B:=\vee\{F\in\mathscr{F}\colon j(F)=B\}.$$

(ii) $F_B=\{\omega\in F_{\mathscr{G}}\colon\operatorname{supp}\mu_\omega\subseteq B\}$.

(iii) $\mathscr{C}_{\mathscr{G}}(B)=c(\pi_B)$, where π_B denotes the representation corresponding to F_B.

(iv) There is a state $\omega\in F_{\mathscr{G}}$ such that $F_\omega\in\mathscr{F}(B)$ (F_ω is the folium generated by ω).

Proof: (i) This is clear from Theorem 6. (ii) For any $\omega\in F_{\mathscr{G}}$, $\operatorname{supp}\mu_\omega\subseteq B$, one has [Prop. 4, part (ii)]

$$j(F_{\pi_\omega})=E_{\pi_\omega}=\operatorname{supp}\mu_\omega\subseteq B,$$

so per definition $F_{\pi_\omega}\subseteq F_B$, hence $\omega\in F_B$. The other inclusion, $F_B\subseteq\{\omega\in F_{\mathscr{G}}\colon\operatorname{supp}\mu_\omega\subseteq B\}$, is clear from $j(F_B)=B$.

(iii) From Eq. (3) one knows that $c(\pi_B)\leqslant c:=\mathscr{C}_{\mathscr{G}}(B)$. In order to prove $c\leqslant c(\pi_B)$, we show $F_c\subseteq F_B$, where F_c is the folium corresponding to the central projection c. Let $\omega\in F_c$. Then

$$1=\omega(c)=\omega(\mathscr{C}_{\mathscr{G}}(B))=\mu_\omega(B)$$

according to Prop. 4, (i). Hence, $\operatorname{supp}\mu_\omega\subseteq B$, and, due to part (ii), $\omega\in F_B$.

(iv) Take a countable, dense subset T of B, and define the measure

$$\eta:=\sum_{i=1}^{\infty}\lambda_i\delta_{x_i},\quad\sum_i\lambda_i=1,\quad\lambda_i\neq 0\quad\forall i\in\mathbb{N}$$

where $\{x_i\colon i\in\mathbb{N}\}$ is a denumeration of T. Then there are states $\varphi_i\in\partial S^P(\mathscr{A})$ with $\mu_{\varphi_i}=\delta_{x_i}$ (see the proof of Theorem 6), and thus for the state $\omega:=\Sigma_i\lambda_i\varphi_i$ it holds $\mu_\omega=\eta$ and hence [Prop. 4 (ii)]

$$j(F_\omega)=\operatorname{supp}\mu_\omega=\operatorname{supp}\eta=B.$$ ∎

VI. DYNAMICAL SUBDESCRIPTIONS

In this last section we consider the question, which of the possible kinematical subdescriptions are also dynamical ones in the sense of Theorem 2, i.e.: in which representations π do the local time evolutions τ_t^Λ converge to a *-automorphism group of the $C^\star$-algebra $\mathscr{C}_\pi := \pi(\mathscr{A}) \otimes \mathscr{N}_\pi$?

The answer is quite simple:

Theorem 8: Let $\pi \leqslant \pi_{\mathscr{G}}$, $E_\pi := j(F_\pi) \subseteq E$. Then the following conditions are equivalent.

(i) τ_t^Λ converges (in the sense of Theorem 2) to a *-automorphism group on $\mathscr{C}_\pi = \pi(\mathscr{A}) \otimes \mathscr{C}_\pi(C(E_\pi))$.

(ii) E_π is invariant under the classical flow φ_t^Q on E.

Proof: (ii) $\Rightarrow$ (i): This proof is identical to the original proof in Ref. 15 of our Theorem 2.

(i) $\Rightarrow$ (ii): Assume that (ii) is false. With the convergence of τ_t^Λ on $\mathscr{N}_\pi = \mathscr{C}_\pi(C(E_\pi))$ we mean that (at least for small $t \in \mathbb{R}$, say $|t| \leqslant r$, $r \in \mathbb{R}_+$)

$$\tau_t^Q(X_\pi(\beta)) = \text{stop} - \lim_{\Lambda} \tau_t^Q(\pi(X_\Lambda(\beta)))$$

$$\left[= \text{stop} - \lim_{\Lambda} \pi(\tau_t^\Lambda(X_\Lambda(\beta))) \right]$$

$$\forall \beta \in \mathscr{G}.$$

As shown in Ref. 15, this implies the relation

$$\tau_t^Q(P(X_\pi(\beta^1),...,X_\pi(\beta^L))) = \mathscr{C}_\pi[(\varphi_t^Q)^\star(P(F^1,...,F^L))|_{E_\pi}] \tag{6}$$

for all polynomials on $\mathbb{R}^L$. But if $\varphi_t^Q(E_\pi) \neq E_\pi$, Eq. (6) means that τ_t^Q cannot be norm continuous: It exists by assumption an $F_0 \notin E_\pi$ such that $\varphi_{-t}^Q(F_0) \in E_\pi$ for a $t \leqslant r$ (observe the group property of φ_t^Q). Also, we can find $f_1, f_2 \in C(E)$ with

$$f_1|_{E_\pi} = f_2|_{E_\pi}, \quad |f_1(F_0) - f_2(F_0)| > 2.$$

Now take polynomials P_1, P_2 with

$$\|P_i - f_i\|_{C(E)} < (\epsilon/2), \quad i = 1,2,$$

so that we have

$$\|P_1 - P_2\|_{C(E_\pi)} < \epsilon, \quad |P_1(F_0) - P_2(F_0)| > 1$$

for ϵ sufficiently small. But then one has on the one hand

$$\begin{aligned}
&\|P_1(X_\pi(\beta^1),...,X_\pi(\beta^L)) - P_2(X_\pi(\beta^1),...,X_\pi(\beta^L))\| \\
&= \|\mathscr{C}_\pi(P_1|_{E_\pi}) - \mathscr{C}_\pi(P_2|_{E_\pi})\| \\
&= \|P_1|_{E_\pi} - P_2|_{E_\pi}\|_{C(E_\pi)} \\
&< \epsilon,
\end{aligned}$$

since $\mathscr{C}_\pi$ is an isomorphism, and on the other hand,

$$\begin{aligned}
&\|\tau_t^Q(P_1(X_\pi(\beta^1),...,X_\pi(\beta^L))) \\
&\quad - \tau_t^Q(P_2(X_\pi(\beta^1),...,X_\pi(\beta^L)))\| \\
&= \|(\varphi_t^Q)^\star(P_1)|_{E_\pi} - (\varphi_t^Q)^\star(P_2)|_{E_\pi}\|_{C(E_\pi)} \\
&= \|(P_1 \circ (\varphi_t^Q))|_{E_\pi} - (P_2 \circ (\varphi_t^Q))|_{E_\pi}\|_{C(E_\pi)} \\
&= \sup\{|P_1(\varphi_t^Q(F)) - P_2(\varphi_t^Q(F))|: F \in E_\pi\} \\
&\geqslant |P_1(F_0) - P_2(F_0)| > 1;
\end{aligned}$$

hence, τ_t^Q is not norm continuous (for the above t) on the polynomials and can therefore not exist as a *-automorphism on $\mathscr{C}_\pi$ or any bigger $C^\star$-algebra.

Remark: In the case where there exist polynomials $P_1 \neq P_2$ with $P_1|_{E_\pi} = P_2|_{E_\pi}$, similar arguments show that, due to (6), τ_t^Q cannot even be well defined. ■

Theorem 8 says, in particular, that if in a representation π the τ_t^Λ converge to an automorphism group which includes the dynamics of the algebra of macroscopic observables $\mathscr{N}_\pi$, this dynamics is necessarily given by the flow φ_t^Q on the space E_π.

In general, however, a limiting automorphism group exists already on a subalgebra of $\mathscr{C}_\pi$ of the form $\pi(\mathscr{A}) \otimes \mathscr{N}'_\pi$, where $\mathscr{N}'_\pi \subseteq \mathscr{N}_\pi$ is a subalgebra.

To see this, it is necessary to look at the classical flow φ_t^Q in more detail. The nonlinear polynomial Q: $\mathbb{R}^L \to \mathbb{R}^L$ contains a certain number of variables $F^{j_1},...,F^{j_s}$, $j_k \in \{1,...,L\}$, corresponding to basis elements $\beta^{j_1},...,\beta^{j_s}$ of Lie(SU(m)).

Let $\mathscr{I} \subseteq \mathscr{G}$ denote the vector space spanned by these elements (recall that $\mathscr{G}$ was defined as the smallest Lie-algebra containing $\mathscr{I}$), and $\mathscr{J} \subseteq \mathscr{I} \subseteq \mathscr{G}$ the vector space spanned by those elements such that the corresponding variables occur *nonlinearly* in Q. We also define the space

$$\mathscr{K} := \text{lin}_{\mathbb{R}}\{\mathscr{J},\ [\mathscr{I},\mathscr{J}],\ [\mathscr{I},\ [\mathscr{I},\mathscr{J}]],...\} \subseteq \mathscr{G},$$

where $[\cdot,\cdot]$ denotes the Lie bracket and $[\mathscr{I},\mathscr{J}]$ the space of all vectors $[x,y]$, $x \in \mathscr{I}, y \in \mathscr{J}$ (which is spanned by the basis elements β^l occurring on the right-hand side of all brackets $[\beta^i,\beta^j] = \Sigma_l C_l^{ij}\beta^l$, $\beta^i \in \mathscr{I}$, $\beta^j \in \mathscr{J}$). Note that the sequence $\mathscr{J}$, $[\mathscr{I},\mathscr{J}]$,... gives no new elements after at most $L-1$ steps, since $[\mathscr{I},[\mathscr{I},\mathscr{J}]] = [\mathscr{I},\mathscr{J}]$ implies $[\mathscr{I},[\mathscr{I},[\mathscr{I},\mathscr{J}]]] = [\mathscr{I},\mathscr{J}]$, etc. We note that $\mathscr{K}$ does not necessarily contain $\mathscr{I}$.

Lemma 9: It holds

(i) $[\mathscr{I},\mathscr{K}] \subseteq \mathscr{K}$.

(ii) $\mathscr{K}$ is a Lie subalgebra of $\mathscr{G}$.

(iii) $\mathscr{K} = \mathscr{G} \Leftrightarrow \mathscr{I} \subseteq \mathscr{K}$.

Proof: (i) is obvious. (ii): One has

$$[\mathscr{K},\mathscr{J}] \subseteq \mathscr{K};\ [\mathscr{K},[\mathscr{I},\mathscr{J}]] \overset{!}{\subseteq} \text{lin}_{\mathbb{R}}\{[\mathscr{I},[\mathscr{K},\mathscr{J}]],$$

$$[\mathscr{J},[\mathscr{I},\mathscr{K}]]\} \subseteq \mathscr{K},$$

where (!) follows from the Jacobi identity. Similarly, one has $[\mathscr{K},[\mathscr{I},[\mathscr{I},\mathscr{J}]]] \subseteq \mathscr{K}$ etc., hence $[\mathscr{K},\mathscr{K}] \subseteq \mathscr{K}$. Part (iii) is clear from (ii) and the definition of $\mathscr{G}$. ■

It is clear that $\mathscr{K}$ is spanned by certain basis elements $\beta^{k_1},...,\beta^{k_r}$, $k_i \in \{1,...,L\}$, which can without loss of generality be assumed to be the first K ones, $K \leqslant L$.

Proposition 10: The variables $F^1,...,F^K$ in $\mathbb{R}^L$ are dynamically decoupled from $F^{K+1},...,F^L$ for the flow φ_t^Q, i.e.,

$$\varphi_t^Q|_{\mathbb{R}^K}: \mathbb{R}^K \to \mathbb{R}^K,$$

$$(F^1,...,F^K) \to p(\varphi_t^Q(F^1,...,F^L))$$

is well defined; here, p is the projection $\mathbb{R}^L \to \mathbb{R}^K$ onto the first K coordinates.

Proof: We have to show that the first K components of the vector field λ^Q, which are again polynomials, depend only on $F^1,...,F^K$. Let $k \leqslant K$. Recall that

$$(\lambda^{Q}(F^1,...,F^L))^k = \sum_{j=1}^{L} \frac{\partial Q}{\partial F_j}(F^1,...,F^L)\Big(\sum_{l=1}^{L} C_l^{kj} F^l\Big).$$

Note that $(\partial Q/\partial F_j)(F^1,...,F^L)$ contains only variables corresponding to the basis elements of $\mathscr{J} \subseteq \mathscr{K}$. The F^l also belong to $\mathbf{R}^K$: they correspond to the β^l occurring in $[\beta^k,\beta^j] = \Sigma_l C_l^{kj}\beta^l \subseteq [\mathscr{K},\mathscr{I}] \subseteq \mathscr{K}$. Here, we have assumed $\beta^j \in \mathscr{I}$; this is justified, since, if $\beta^j \notin \mathscr{I}$, one has per definition $\partial Q/\partial F_j = 0$. Thus, only variables in $\mathbf{R}^K$ occur in $(\lambda^{Q}(F^1,...,F^L))^k$ and the proposition is proved. ■

The proposition shows that the Lie subalgebra $\mathscr{K}$ has a profound dynamical significance.

Remark: We note that the definitions of $\mathscr{J}, \mathscr{I}, \mathscr{K}, \mathscr{G}$ are dependent on the originally chosen operators $X(\beta^1),...,X(\beta^L)$ with the help of which H^Λ was expressed (equivalently, on the coordinates $F^1,...,F^L$ in $\mathbf{R}^L$): a change of coordinates will in general lead to different spaces $\mathscr{J}, \mathscr{I}, \mathscr{K}, \mathscr{G}$. One may therefore impose the following *minimality condition*:

Let $X(\beta^1),...,X(\beta^s)$ be operators such that H^Λ can be expressed with the corresponding $X_\Lambda(\beta^1),...,X_\Lambda(\beta^s)$ and such that $\mathscr{J},\mathscr{I},\mathscr{K},\mathscr{G}$ are minimal; i.e., $\mathscr{I}(\mathscr{J})$ is the *smallest* space such that (the nonlinear part of) H^Λ can be expressed with operators $X_\Lambda(\beta), \beta \in \mathscr{I}(\mathscr{J})$.

It is not trivial that such minimal $\mathscr{J},\mathscr{I}$ and hence minimal $\mathscr{K}$ and $\mathscr{G}$ do exist. This can be shown, but here we contend ourselves with noting that in the (physically most important) case of quadratic Q (corresponding to two-particle interactions), such existence follows easily from the fact that for such Q, H^Λ can always be expressed in the form

$$H^\Lambda = |\Lambda| \Big(\sum_{j=1}^{s-1} \lambda_j X_\Lambda(\bar{\beta}^j)^2 + \sum_{j=1}^{s-1} \mu_j X_\Lambda(\bar{\beta}^j) + \eta X_\Lambda(\bar{\beta}^s)\Big)$$

($\lambda_j \neq 0\ \forall j$) with linear independent elements $\bar{\beta}^1,..., \bar{\beta}^s$ of Lie(SU(m)) and η possibly 0.

Turning now to our original purpose, we can define the objects $\mathscr{N}_\pi^{\mathscr{K}}, \mathscr{E}_\pi^{\mathscr{K}}, E_\pi^{\mathscr{K}}, \mathscr{C}_\pi^{\mathscr{K}}$, just as the $\mathscr{N}_\pi, \mathscr{E}_\pi, E_\pi, \mathscr{C}_\pi$ with $\mathscr{G}$ replaced by $\mathscr{K}$. All stated properties of the latter objects are also true for the former ones, and one has $p(E_\pi) = E_\pi^{\mathscr{K}}$, as well as $\mathscr{N}_\pi^{\mathscr{K}} \subseteq \mathscr{N}_\pi$. If now we introduce the map (with obvious notations)

$$j^{\mathscr{K}}: \mathscr{F} \to \mathscr{E}^{\mathscr{K}}$$

$$F \to E_{\pi_F}^{\mathscr{K}} = \overline{\bigcup_{\omega \in F} \operatorname{supp} \mu_\omega^{\mathscr{K}}},$$

the following holds.

Theorem 11: For every representation $\pi \leqslant \pi_{\mathscr{G}}$ the following conditions are equivalent:

(i) τ_t^Λ converges (in the sense of Theorem 2) to a *-automorphism group of $\mathscr{C}_\pi^{\mathscr{K}}$.

(ii) $j^{\mathscr{K}}(F_\pi) = E_\pi^{\mathscr{K}} \subseteq E^{\mathscr{K}}$ is invariant under the dynamics of $\varphi_t^{Q}|_{\mathbf{R}^K}$ (defined in Prop. 10).

Proof: (ii) $\Rightarrow$ (i): $\mathscr{K}$ was constructed in such a way that only the operators $X_\pi(\beta^k), \beta^k \in \mathscr{K}$, get involved in the limiting dynamics, as can be seen from Eq. 3.12 in Ref. 15: in

$$\text{stop} - \lim_{\Lambda} \pi([H^\Lambda,x]^m), \quad \forall m \in \mathbf{N},$$

$X_\pi(\beta^k)$, $\beta^k \in \mathscr{J}$, occur for $m = 1$

$X_\pi(\beta^k), \beta^k \in [\mathscr{I},\mathscr{J}]$, occur for $m = 2$, etc.

Therefore, the proof of existence of the limiting dynamics on $\mathscr{C}_\pi^{\mathscr{K}}$ goes through as before.

(i) $\Rightarrow$ (ii) This is proved as in Theorem 8. ■

Remark: In general, it seems possible that there exist even smaller algebras [possibly not of the type $\pi(\mathscr{A}) \otimes C(D)$] on which a limiting dynamics exists; such dynamics would then not be defined on a specified set of macroscopic observables $X_\pi(\beta)$.

VII. CONCLUSION

Theorem 8 (resp. 11), together with Theorem 6, means that one knows in principle all representations in which the limiting dynamics exists and has a "macroscopic part" [i.e., is defined on a set of macroscopic observables $X_\pi(\beta)$ of the representation]. It is interesting to note that these possibilities are determined solely by the "classical part" of the system. This, however, is less surprising than it might seem, because the work of Bona shows that, in fact, the *whole dynamics* is essentially determined by its classical part already.

We also observe here that, according to Theorem 6, to every "classical subdescription," i.e., every closed φ_t^{Q}-invariant set $B \subseteq E$, there exists a dynamical description of the quantum system with classical state space B. It can be chosen maximal (in the sense of the ordering $\leqslant$ of representations).

Finally, let us emphasize that one important question in this context is still open. We agree with Morchio and Strocchi in Ref. 14 that one should really require more of a dynamical description than we have required so far:

(i) the permissible representations should belong to *full* ($w^\star$-dense) folia;

(ii) the dynamical automorphism τ_t^{Q} in π should be σ-weakly continuous for all t, i.e., τ^{Q} should be extendable to a group of $W^\star$-automorphisms of $\mathscr{M}_\pi$. Only then can one define Schrödinger dynamics on $(\mathscr{M}_\pi)_{\star,+,1}$, and in particular on the Hilbert space $\mathscr{H}_\pi$; only then is F_π a physical folium in the sense of Ref. 21. A representation which does not allow for this cannot qualify as a full-blooded dynamical description of the system.

In our context, (i) is taken care of by the well-known fact that the folia of faithful representations are full; since the $C^\star$-algebra $\mathscr{A}$ is simple, *all* folia in $S(\mathscr{A})$ are full.

The second requirement, however, presents a more difficult problem. We therefore leave a discussion of this issue to another occasion,[22] where we shall show that (ii) is *not* true in general. Here we contend ourselves with noting that the question of (ii) *cannot* be settled by looking at the classical dynamics alone, in contrast to the question of existence of the limiting (Heisenberg) dynamics itself.

ACKNOWLEDGMENTS

I would like to thank A. Rieckers for continuous support and encouragement, and all members of the $C^\star$-group in Tübingen, especially B. Kümmerer, for many helpful discussions. I am also indebted to P. Bona for making available to me Refs. 15 and 23 prior to publication, and for enlightening discussions. I would also like to thank the referee for very instructive and helpful criticisms.

This work is part of a research project supported by the Deutsche Forschungsgemeinschaft. Financial support of the Studienstiftung des deutschen Volkes is gratefully acknowledged.

[1]H. Primas, *Chemistry, Quantum Mechanics and Reductionism* (Springer, Berlin, 1981).
[2]D. A. Dubin and G. L. Sewell, J. Math. Phys. **11**, 2990 (1970).
[3]G. G. Emch and H. J. F. Knops, J. Math. Phys. **11**, 3008 (1970).
[4]F. Jellinek, Commun. Math. Phys. **9**, 169 (1968).
[5]O. Bratteli and D. W. Robinson, *Operator Algebras and Statistical Mechanics I, II* (Springer, Berlin, 1979, 1981).
[6]M. B. Ruskai, Commun. Math. Phys. **20**, 193 (1971).
[7]M. Pulvirenti and B. Tirozzi, Commun. Math. Phys. **30**, 83 (1973).
[8]O. Bratteli and D. W. Robinson, Commun. Math. Phys. **50**, 133 (1976).
[9]L. van Hemmen, Fortschr. Phys. **26**, 397 (1978).
[10]H. Roos, Physica A **100**, 183 (1980).
[11]A. Rieckers, J. Math. Phys. **25**, 2593 (1984).
[12]E. Duffner, Physica A **133**, 187 (1985).
[13]K. Hepp and E. H. Lieb, Helv. Phys. Acta **46**, 573 (1973).
[14]G. Morchio and F. Strocchi, J. Math. Phys. **28**, 622 (1987).
[15]P. Bona, J. Math. Phys. **29**, 2223 (1988).
[16]E. Duffner and A. Rieckers, Z. Naturforsch. **43**, 321 (1988).
[17]R. Haag, R. V. Kadison, and D. Kaster, Commun. Math. Phys. **16**, 81 (1970).
[18]A proof of a similar statement is also announced in G. Morchio and F. Strocchi, Commun. Math. Phys. **99**, 153 (1985), Sec. 5.
[19]R. V. Kadison, Topology **3**, Suppl. 2, 177 (1965).
[20]G. Grätzer, *Lattice Theory* (Freeman, San Francisco, 1971).
[21]G. L. Sewell, Commun. Math. Phys. **33**, 43 (1973).
[22]T. Unnerstall, preprint, 1989 (submitted for publication).
[23]G. K. Pedersen, *C^*-algebras and their Automorphism Groups* (Academic, London, 1979).
[24]P. Bona, J. Math. Phys. **30**, 2994 (1989).
[25]E. Størmer, J. Funct. Anal. **3**, 48 (1969).

Frequency-dependent susceptibility of a free electron gas in D dimensions

M. Howard Lee and Jon T. Nelson
Department of Physics and Astronomy, University of Georgia, Athens, Georgia 30602

(Received 10 March 1989; accepted for publication 8 November 1989)

The real and imaginary parts of the dynamic susceptibility of a free electron gas at $T=0$ are obtained as a function of wave vector k, frequency ω, and space dimensionality D. The real part of the dynamic susceptibility is shown to be simply related to the static susceptibility. For odd-numbered dimensions there exists a general dimensional relationship for the dynamic susceptibility, originating from a similar relationship for the static susceptibility. For even-numbered dimensions there is no such general relationship, but instead a restricted one.

I. INTRODUCTION

The dynamic susceptibility of a free or ideal electron gas at $T=0$ has played an important role in the development of quantum many-body theory. The real and imaginary parts of this quantity in three dimensions are found in nearly all standard texts on many-body theory.[1] The dynamic susceptibility is also known in one and two dimensions.[2–5] These solutions show a certain similarity. To our knowledge, no one has given a unified picture. Such a picture would be possible if one had an expression for the dynamic susceptibility as a function of the spatial dimensionality D.

The dynamic susceptibility of an ideal electron gas is still a useful quantity: It appears in the current formulation of the dynamic susceptibility of a nonideal electron gas.[1] The dynamic local field, which expresses the dynamic correlation absent in mean field theory, is defined in reference to the dynamic susceptibility of an ideal electron gas. In the treatment of Hong and Lee,[6] for example, the ideal dynamic susceptibility is intimately involved in the formal expression for the nonideal dynamic susceptibility. Perhaps much more interesting is the following: According to linear response theory,[7] the static limit of the dynamic susceptibility is the static susceptibility, i.e., $\chi(k,\omega=0)=\chi(k)$, where k and ω are, respectively, wave vector and frequency. In view of this relationship, one might ask whether the properties of the static susceptibility are factorable from the dynamic susceptibility: If they are not an integral part of it, one might scale them out. Such an idea would not be useful if some static properties are somehow a basis of the dynamic susceptibility as the frequency is "turned on."

In this connection, we shall first briefly look at the static susceptibility and examine its salient properties. Recently, Sharma and Lee[8] have given an analytic expression for the static susceptibility in any spatial dimension by a technique of dimensional regularization.[9] It is thus now possible to regard D and k as two independent variables of static susceptibility, i.e., $\chi(D,k)$. According to this work viewed in the plane of D and k, there is a singularity at $D=1$ and $k=2k_F$, where k_F is the Fermi wave vector. The singularity is a simple pole if approached along D at $k=2k_F$ and it is logarithmic if approached along k at $D=1$. This plane of D and k is divided into two regions, the low- and high-k sides, by a line of $k=2k_F$. The behavior of the static susceptibility on one side is generally different from that on the other side. If D is an odd number, the structure of the static susceptibility on one side is an analytic continuation of that on the other side. If D is an even number, there is no such relationship. There is thus a marked "even–odd" effect in the behavior of the static susceptibility.[8]

The dynamic susceptibility introduces a third variable ω into this picture. How does it, if at all, affect the above described properties? Does it introduce anything new? That the dynamic susceptibility has two parts is of no particular concern. Since one part is related to the other by a Hilbert transform (the Kramers–Kronig relation[1]), it is sufficient to focus on one part, say $\mathrm{Re}\,\chi(Dk\omega)$. We are able to provide simple, but complete answers to these questions. Our solution is a unified one, from which all specialized cases can be simply and directly obtained.

It might be helpful to point out in advance some of our main results. The D dependence in the dynamic susceptibility is exactly the same as the D dependence in the static susceptibility. That is, the D coordinate is orthogonal to the ω coordinate, just as it is to the k coordinate. However, the ω coordinate is not orthogonal to the k coordinate: It is a kind of translation of the k coordinate. As a result, there is nothing intrinsically new in the dynamic susceptibility beyond what already exists in the static susceptibility. The singularity of the static susceptibility at $D=1$ and $k=2k_F$, for example, is also a singularity of the dynamic susceptibility, albeit in a more intricate way. The unifying picture of the static susceptibility also unifies the dynamic susceptibility. One can thus have a complete unified picture of the susceptibility.

II. STATIC SUSCEPTIBILITY IN D DIMENSIONS

We begin with a brief summary of the static susceptibility in D dimensions. One defines the static susceptibility as[8]

$$\chi(D,k)=2\sum_p \frac{n(p)-n(p+k)}{\omega_{pk}}, \tag{1}$$

where $\omega_{pk}=\epsilon_{p+k}-\epsilon_p$, $\epsilon_p=p^2/2m$, m is the mass of the electron, and $n(p)=n(|p|)$ is the Fermi function. At $T=0$, $n(p)=\theta(k_F-|p|)$, a step function, where k_F is the Fermi wave vector. Henceforth, all our wave vectors are scaled by k_F. We shall further normalize the susceptibility by its value at $k=0$, $\chi(D,k=0)=D\rho/2\epsilon_F$, where ρ is the number den-

 0022-2488/90/030689-06$03.00

sity and ϵ_F is the Fermi energy. (See Appendix A.) The normalized susceptibility $\tilde{\chi}(D,k) \equiv \chi(D,k)/\chi(D,0)$ is convenient for showing D dependence since $\tilde{\chi}(D,k) = 1$ if $k = 0$ for every D. Strictly speaking, our susceptibility is the susceptibility per unit volume.

Sharma and Lee[8] have shown that at $T = 0$

$$\tilde{\chi}(D,k) = \tilde{\chi}_1(D,k) = F(1,1 - D/2,\tfrac{3}{2},z), \quad \text{if } z < 1, \tag{2a}$$

$$\tilde{\chi}(D,k) = \tilde{\chi}_2(D,k) = D^{-1}z^{-1}F(1,\tfrac{1}{2},1 + D/2,z^{-1}), \quad \text{if } z > 1, \tag{2b}$$

where $z = (k/2)^2$. Here F is the hypergeometric function (hgf) defined as

$$F(a,b,c,t) = \sum_{n=0}^{\infty} \frac{a_n b_n}{c_n n!} t^n, \quad |t| < 1, \tag{3}$$

where $a_n = \Gamma(n+a)/\Gamma(a)$, etc.; $c \neq 0, -1, -2,...$, i.e., $D \neq -2, -4, -6,...$. Properties of the hgf prove to be useful in bringing out D dependence.

For example, the contiguity property gives the following relationship: If D is a positive-odd integer,

$$\tilde{\chi}(D+2,k) = [1/(D+1)][1 + (1-z)D\tilde{\chi}(D,k)], \tag{4}$$

where $z = (k/2)^2$; it is valid for any k. Exactly the same relationship holds for a positive-even integer D if $z > 1$ ($k > 2$). If $z < 1$, there is no such relationship. In this case, the susceptibility is a polynomial of z. See Appendix B for the derivation of (4). One sees immediately from (4) that

$$\tilde{\chi}(D,k=2) = 1/(D-1), \quad D > 1. \tag{5}$$

Relationship (4) indicates that the susceptibility is of two families, one composed of even-numbered dimensions and the other composed of odd-numbered dimensions. Now,

$$\tilde{\chi}(D=1,k) = k^{-1}\ln|(1+k/2)/(1-k/2)|. \tag{6}$$

Thus (4) implies that the logarithmic singularity is contained in the susceptibility of all odd-numbered dimensions. Also,

$$\tilde{\chi}(D=2,k) = 1 - (1 - z^{-1})^{1/2}, \quad \text{if } z > 1, \tag{7}$$

where $z = (k/2)^2$. Thus the square root singularity is contained in the susceptibility of all even-numbered dimensions if $k > 2$. The D- and k-dependent behavior is perhaps most clearly illustrated in Fig. 1, reproduced from the work of Sharma and Lee.[8] The $k = 2k_F$ line, referred to in Fig. 1 as a ridge, serves to demarcate the Dk plane: It turns out to play an important role as a boundary in the dynamic susceptibility.

III. DYNAMIC SUSCEPTIBILITY IN D DIMENSIONS

The dynamic susceptibility is defined as (suppressing D dependence)

$$\chi(k,\omega) = -2\sum_p \frac{n(p) - n(p+k)}{\omega - \omega_{pk} + i\eta}, \quad \eta \to 0+. \tag{8}$$

We have inserted a negative sign (opposite to the usual convention), so that $\chi(k,\omega = 0) = \chi(k)$. Now, expressing ω in units of the Fermi energy $\epsilon_F = k_F^2/2m$ with $\hbar = 1$, one can write (8) as follows:

$$\chi(k,\omega) = -(1/k\epsilon_F)[Q(k/2 + u') + Q(k/2 - u')], \tag{9}$$

where $u' = u + i\eta$, $u = \omega/2k$, and

$$Q\left(\frac{k}{2} \pm u'\right) = \sum_p \frac{n(p)}{p \cdot k + k/2 \pm u'}, \tag{10}$$

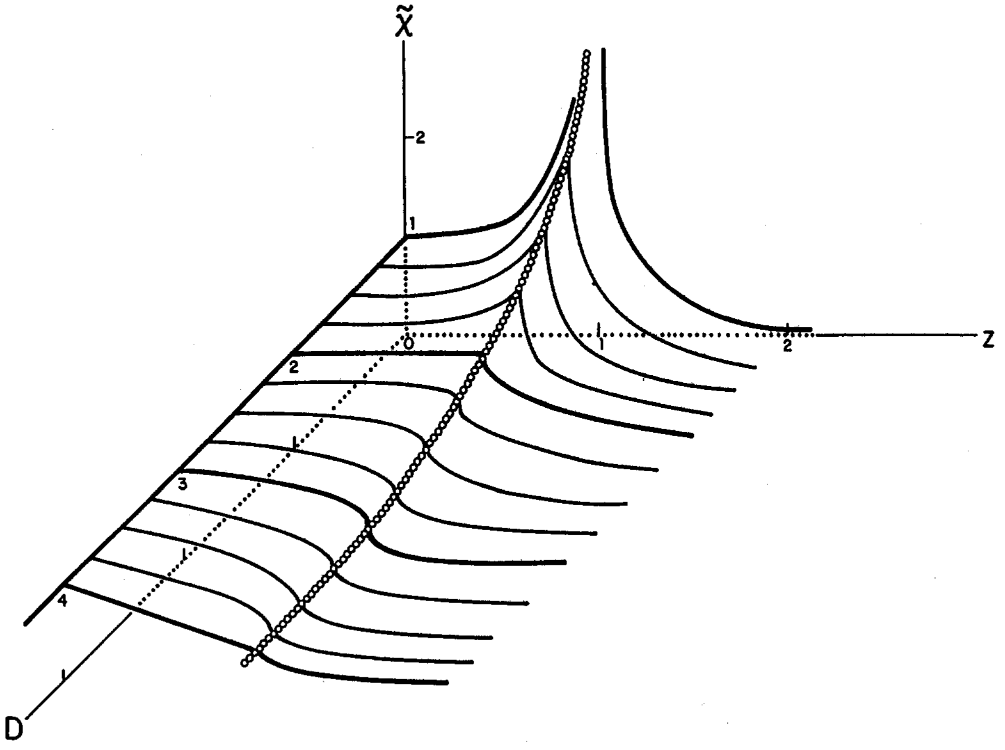

FIG. 1. The normalized static susceptibility illustrated as a surface of the volume of $\tilde{\chi}Dk$, reproduced from Ref. 8. Here, $z = (k/2)^2$, where k is in units of the Fermi wave vector k_F and D is space dimensionality. The line of small circles depicts a ridge on this surface separating the high-k side from the low-k side.

where $\hat{k} = \mathbf{k}/|k|$.

Now,

$$\chi(k,\omega = 0) = \chi(k) = -(2/k\epsilon_F)Q(k/2). \tag{11}$$

Hence,

$$Q(k/2) = -(k\epsilon_F/2)\chi(k). \tag{12}$$

Functionally, it then follows that

$$Q(k/2 \pm u') = -(\epsilon_F/2)(k \pm 2u)\chi(k \pm 2u'). \tag{13}$$

Substituting (13) into (9), we obtain

$$2k\chi(k,\omega) = (k+2u)\chi(k+2u') + (k-2u)\chi(k-2u'), \tag{14}$$

that is, the ω coordinate acts as a translation of the k coordinate. Hence, the frequency cannot produce wholly new properties. Since this translation has no effect on the D coordinate, the ω coordinate remains orthogonal to it.

One can immediately obtain the real part of the dynamic susceptibility as follows:

$$k \operatorname{Re} \chi(k\omega) = s_+\chi(2s_+) + s_-\chi(2s_-), \tag{15}$$

where $s_\pm = k/2 \pm u = (k \pm \omega/k)/2$. One can obtain the imaginary part similarly. However, it is somewhat more practical to use the following expression for analysis:

$$-k\epsilon_F \operatorname{Im} \chi(k\omega) = \operatorname{Im} Q(s_+ + i\eta) + \operatorname{Im} Q(s_- - i\eta). \tag{16}$$

It is perhaps worth noting here that relations (15) and (16) are valid for any T since our derivation does not depend on the form of the Fermi function $n(p)$ other than for its isotropy, i.e., $n(p) = n(|p|)$.

IV. REAL PART AT $T=0$

The real part of the normalized dynamic susceptibility is thus

$$k \operatorname{Re} \tilde{\chi}(k,\omega) = s_+\tilde{\chi}(2s_+) + s_-\tilde{\chi}(2s_-). \tag{17}$$

Here, $\tilde{\chi}(2s_\pm)$ is in effect a static susceptibility, as already given by Sharma and Lee[8] [see (2a) and (2b)]. Also, recall

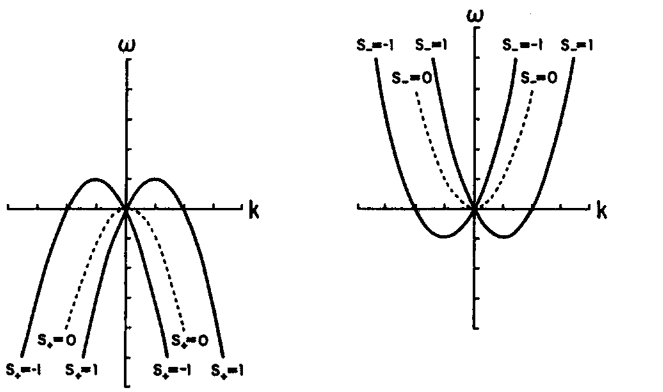

FIG. 2. Translated lines of $k/2$ in the plane of $k\omega$. Here, $s_\pm = k/2 \pm \omega/2k$, where k is given in units of the Fermi wave vector k_F and ω is given in units of the Fermi energy ϵ_F.

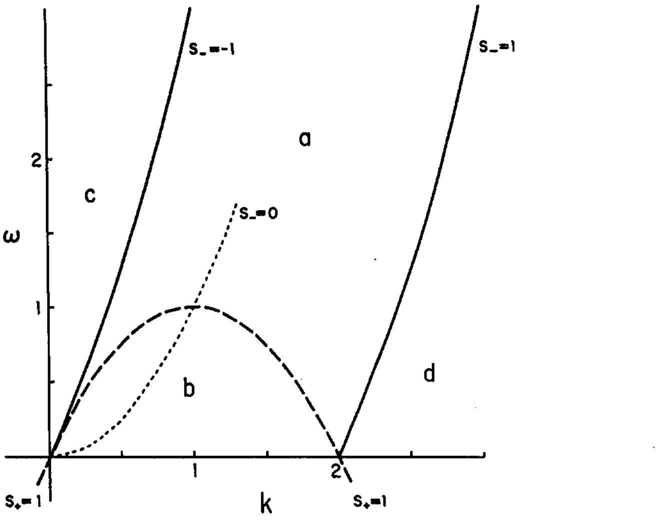

FIG. 3. Areas of the physical domain of $k\omega$ formed by boundary lines of $s_\pm$. Along the line of $s_- = 0$ (dotted line), $k \operatorname{Re} \tilde{\chi}(k\omega) = s_+\tilde{\chi}(2s_+)$ only. See Eq. (17).

that the static susceptibility $\tilde{\chi}(k)$ is characterized by a boundary line of $k/2 = 1$ in the Dk plane (see Fig. 1). Thus the real part of the dynamic susceptibility is similarly bordered by a line of $|s_\pm| = 1$, where $|s_\pm| < 1$ corresponds to the "low-k" side and $|s_\pm| > 1$ to the "high-k" side. These new boundary lines are illustrated in Fig. 2.

In Fig. 3, we have put the above boundary lines together, but restricted them to the physical region, i.e., $k \geqslant 0$ and $\omega \geqslant 0$. There are four different areas marked a–d (see, also, Table I). Area b corresponds to the low-k side and areas c and d to the high-k side. Area a is mixed. The boundary lines marking these distinct areas are lines of a singularity. Consider the point $k = 2$ and $\omega = 0$ (see Fig. 3), which is a confluence point of areas b and d. This point is a point of the logarithmic singularity if $D = 1$. If we lift ω from $\omega = 0$, it draws out two branches, each of which is a line of the logarithmic singularity. These lines separate areas a, b, and d. Other lines are also lines of the same singularity.

Since the D coordinate remains orthogonal as noted, these boundary lines are also lines of singularities of $\tilde{\chi}(D,k)$ for a given D, as illustrated in Fig. 1. If $D = 2$, for example, the corresponding boundary lines of $\operatorname{Re} \tilde{\chi}(k,\omega)$ are lines of a discontinuity in the slope of $\tilde{\chi}(D = 2,k)$ with respect to k.

We shall now use (17) to obtain the real part of the dynamic susceptibility in different dimensions. In Table II, we have indicated the applicable "low and high" forms of the static susceptibility for a given area. Consider $D = 1$. It follows from (6) that

TABLE I. Areas of the physical domain defined by lines of $s_\pm$. See, also, Fig. 1.

a	$s_+ > 1$	$\lvert s_-\rvert < 1$
b	$s_+ < 1$	$\lvert s_-\rvert < 1$
c,d	$s_+ > 1$	$\lvert s_-\rvert > 1$

TABLE II. Applicable forms of the static susceptibility in different physical areas. Here, χ_1 means the low-k form and χ_2 means the high-k form. See, also, Eq. (17).

Area	$\chi(2s_+)$	$\chi(2s_-)$
a	χ_2	χ_1
b	χ_1	χ_1
c,d	χ_2	χ_2

$$s_\pm \tilde{\chi}(D=1,2s_\pm) = \tfrac{1}{2}\ln|(1+s_\pm)/(1-s_\pm)| \equiv \tfrac{1}{2}L_\pm . \tag{18}$$

Hence,

$$\operatorname{Re}\tilde{\chi}(D=1,k\omega) = (\tfrac{1}{2}k)\cdot(L_+ + L_-), \tag{19}$$

which is a well-known result.[2-4] In a similar manner one can thus obtain the real part of the dynamic susceptibility for any D. Instead, we shall use the property that the real part of dynamic susceptibility is expressible in terms of the static susceptibility in order to obtain a relationship between different dimensions.

A. *D* odd

Relationship (4) is valid for both the low- and high-k sides and, hence, is applicable to any areas of Fig. 3. For odd-numbered dimensions, from (17) we thus obtain

$$\begin{aligned} k\operatorname{Re}\tilde{\chi}&(D+2,k\omega) \\ &= k/(D+1) + [D/(D+1)][(1-s_+^2) \\ &\quad\times s_+\tilde{\chi}(D,2s_+) + (1-s_-^2)s_-\tilde{\chi}(D,2s_-)]. \end{aligned} \tag{20}$$

Hence, with (18), one can simply and directly obtain the dynamic susceptibility in any odd-numbered dimension. As an illustration we show the susceptibility when $D=3$ and 5:

$$\begin{aligned} \operatorname{Re}\tilde{\chi}&(D=3,k\omega) \\ &= \tfrac{1}{2} + (1/4k)[(1-s_+^2)L_+ + (1-s_-^2)L_-], \end{aligned} \tag{21}$$

$$\begin{aligned} \operatorname{Re}\tilde{\chi}&(D=5,k\omega) \\ &= \tfrac{1}{4} + (3/8k)[(1-s_+^2)s_+ + (1-s_-^2)s_- \\ &\quad + \tfrac{1}{2}(1-s_+^2)^2L_+ + \tfrac{1}{2}(1-s_-^2)^2L_-]. \end{aligned} \tag{22}$$

One may recognize the well-known three-dimensional result (21).[10]

B. *D* even

If D is an even number and $k/2<1$, then the dimensional relation (4) does not apply. Hence, (4) may not be used in, e.g., area b (see Table II). However, one may still use the original relation (17) to obtain the dynamic susceptibility. Let $D=2$, for example. Then from (2a) and (2b),

$$\tilde{\chi}(D=2,k) = 1, \quad \text{if } k/2<1, \tag{23a}$$

$$\tilde{\chi}(D=2,k) = 1-(1-z^{-1})^{1/2}, \quad \text{if } k/2>1, \tag{23b}$$

where $z=(k/2)^2$.

Referring to Table II, we obtain

$$\begin{aligned} \operatorname{Re}\tilde{\chi}&(D=2,k\omega) \\ &= 1-k^{-1}[c_+(s_+^2-1)^{1/2}] \quad \text{in a}, \\ &= 1 \quad \text{in b}, \\ &= 1-k^{-1}[c_+(s_+^2-1)^{1/2}+c_-(s_-^2-1)^{1/2}] \\ &\quad \text{in c and d}, \end{aligned} \tag{24}$$

where $c_\pm = s_\pm/|s_\pm|$. One may recognize the above as the well-known result due to Stern.[5]

To find $D=4$, we begin with

$$\tilde{\chi}(D=4,k) = 1-2z/3, \quad \text{if } z<1, \tag{25a}$$

$$\qquad\qquad = 1-2z/3(1-z^{-1})^{3/2}, \quad \text{if } z>1, \tag{25b}$$

where $z=(k/2)^2$. See (2a) and (2b) and Sharma and Lee[8] (their Table I). Using (25a) and (25b) with Table II, we obtain

$$\begin{aligned} \operatorname{Re}\tilde{\chi}&(D=4,k\omega) \\ &= 1-(2/3k)(s_+^3+s_-^3) \\ &\quad + k^{-1}[c_+(s_-^2-1)^{3/2}] \quad \text{in a}, \\ &= 1-(2/3k)(s_+^3+s_-^3) \quad \text{in b}, \\ &= 1-(2/3k)(s_+^3+s_-^3)+k^{-1}[c_+(s_+^2-1) \\ &\quad + c_-(s_-^2-1)] \quad \text{in c and d}. \end{aligned} \tag{26}$$

A comparison of our results for $D=2$ and 4 shows that in region c, for example, there is a certain obvious connection: It is, of course, because of the dimensional relationship (4) which is applicable in that region.

V. IMAGINARY PART AT $T=0$

The imaginary part of the dynamic susceptibility is in principle known given the real part of it. To obtain it via Hilbert transforms is, however, not so trivial generally. It is much simpler to obtain the imaginary part directly from (16) by evaluating the imaginary part of Q defined by (10). In fact, one can thereby provide information on Hilbert transforms.

Using (10), we obtain

$$\begin{aligned} Q(s_\pm \pm i\eta) &= A\int_0^1 dp\, p^{D-1}\int_0^\pi d\theta(\sin\theta)^{D-2} \\ &\quad \times(p\cos\theta + s_\pm \pm i\eta)^{-1}, \end{aligned} \tag{27}$$

where $A=(k_F/2\pi)^D 2(\Gamma(\tfrac{1}{2}))^{D-1}/\Gamma((D-1)/2)$. Hence,

$$\begin{aligned} \operatorname{Im}&Q(s_\pm \pm i\eta) \\ &= \mp\pi A\int_0^1 dp\, p^{D-2}\int_{-1}^1 d\mu(1-\mu^2)^{(D-3)/2} \\ &\quad \times\delta\Big(\mu+\frac{s_\pm}{p}\Big). \end{aligned} \tag{28}$$

Clearly, both Im Q vanish in areas c and d (see Fig. 3); they do not vanish in area b or on its boundary line. In area a, Im $Q(s_+)$ vanishes, but Im $Q(s_-)$ does not. Integrating the angle variable first, we obtain the following for the nonvanishing Im $Q(s_-)$:

TABLE III. Coefficients accompanying the imaginary part of $\tilde{\chi}(k\omega)$, tabulated for even- and odd-numbered dimensions.

D	$B(D)$	D	$B(D)$
1	$\pi/2$	2	1
3	$\pi/4$	4	$\frac{2}{3}$
5	$3\pi/16$	6	$\frac{8}{15}$
7	$5\pi/32$	8	$\frac{16}{35}$

$$\begin{aligned}\operatorname{Im} Q(s_\pm \pm i\eta) &= \mp \pi A \int_{|s_\pm|}^{1} dp\, p^{D-2}\left(1-\frac{s_\pm^2}{p^2}\right)^{(D-3)/2} \\ &= \mp \frac{\pi A}{D-1}(1-s_\pm^2)^{(D-1)/2}.\end{aligned} \tag{29}$$

Substituting result (29) in (16) and after normalizing as before by $\chi(D,k=0)$, we obtain the imaginary part of the dynamical susceptibility

$$\begin{aligned}\operatorname{Im} \tilde{\chi}(k\omega) &= -k^{-1}B(1-s_-^2)^{(D-1)/2} \quad \text{in a}, \\ &= k^{-1}B\,[(1-s_+^2)^{(D-1)/2} \\ &\quad -(1-s_-^2)^{(D-1)/2}] \quad \text{in b}, \\ &= 0 \quad \text{in c and d},\end{aligned} \tag{30}$$

where $B=\Gamma(\frac{1}{2})\Gamma(D/2)/2\Gamma((D+1)/2)$, as tabulated in Table III.

Observe that the imaginary part has an unusual property: When $D=1$, $\operatorname{Im}\tilde{\chi}(k\omega)\neq 0$ only in area a. Area a is bordered on all sides by a line of the logarithmic singularity, across which one may not continue. Recall that the static susceptibility is not continuous at $k/2=1$ if $D=1$. One can further show that this particular result for the imaginary part in $D=1$ is already implied by Eq. (19) via a Hilbert transform.[11]

In Table IV, a few examples of the imaginary part are shown, separately for even- and odd-numbered dimensions. The different behavior of the two families can be traced to the static susceptibility. Within the same family there is a dimensional relationship, although it is sometimes hidden.

TABLE IV. Imaginary part of $\tilde{\chi}(k\omega)$ in the physical areas (a) and (b), tabulated for even- and odd-numbered dimensions.

D	$\operatorname{Im}\tilde{\chi}(k\omega)/k^{-1}B$ (a)	(b)
1	-1	0
3	$-(1-s_-^2)$	$(1-s_+^2)-(1-s_-^2)$
5	$-(1-s_-^2)^2$	$(1-s_+^2)^2-(1-s_-^2)^2$
2	$-(1-s_-^2)^{1/2}$	$(1-s_+^2)^{1/2}-(1-s_-^2)^{1/2}$
4	$-(1-s_-^2)^{3/2}$	$(1-s_+^2)^{3/2}-(1-s_-^2)^{3/2}$
6	$-(1-s_-^2)^{5/2}$	$(1-s_+^2)^{5/2}-(1-s_-^2)^{5/2}$

For the family of odd-numbered dimensions, the same dimensional relation (20) for the real part is also implied for the imaginary part via Hilbert transforms. Referring to Table IV again, we note that $\operatorname{Im}\tilde{\chi}(k\omega)/k^{-1}B=-1$ in area a if $s_-=0$, i.e., D independent. Along the line of $s_-=0$ (see Fig. 3), where $\omega=k^2$, this quantity is also at its minimum.

ACKNOWLEDGMENTS

This work has been supported in part by NSF and ARO. We thank Dr. N. L. Sharma for his interest in this work.

APPENDIX A: $\chi(D,k=0)=D\rho/2\epsilon_F$

From (1) it follows that

$$\begin{aligned}\chi(D,k=0) &= -2\lim_{k\to 0}\sum_p \frac{n(p+k)-n(p)}{\epsilon_{p+k}-\epsilon_p} = -2\sum_p \frac{\partial n(p)}{\partial \epsilon_p} \\ &= \left(\frac{L}{2\pi}\right)^D 2m\,u_D \int_0^\infty dp\, p^{D-2}\delta(|p|-k_F),\end{aligned} \tag{A1}$$

where $u_D=2(\Gamma(\frac{1}{2}))^D/\Gamma(D/2)$. Hence, $\chi(D,k=0)/L^D=2m\,u_D k_F^{D-2}/(2\pi)^D$. Now the number density in D dimensions is $\rho=(k_F/2\pi)^D 2u_D/D$. Hence,

$$\chi(D,k=0)/L^D=D\rho/2\epsilon_F. \tag{A2}$$

Q.E.D.

APPENDIX B: DERIVATION OF Eq. (4)

The static susceptibility is expressible in terms of the hgf [see (2a) and (2b)]. Of the three parameters a, b, c of the hgf [see (3)] only parameter b is variable (i.e., D dependent) if $k/2<1$. Then we use the following contiguity property:

$$F_{b-1}=[\tfrac{1}{2}-(1-b)(1-t)F_b]/(\tfrac{3}{2}-b), \tag{B1}$$

where $F_b=F(1,b,\frac{3}{2},t)$. Let $b=1-D/2$ and $t=z=(k/2)^2$. Also, identify $F_{1-D/2}(z)=\tilde{\chi}(D,k)$. Then (B1) becomes

$$\tilde{\chi}(D+2,k)=[1/(D+1)][1+(1-z)D\tilde{\chi}(D,k)]. \tag{B2}$$

If $k/2<1$, parameter c is variable. Then we use the following contiguity property:

$$F_{c+1}=[c/((c-\tfrac{1}{2})t)][1-(1-t)F_c], \tag{B3}$$

where $F_c=F(1,\frac{1}{2},c,t)$. Now let $c=1+D/2$ and $t=z^{-1}=(k/2)^{-2}$. Also, identify $F_{1+D/2}(z^{-1})=zD\tilde{\chi}(D,k)$. Then (B3) becomes (B2). Hence, the dimensional relation is valid for any value of k. Note that the contiguity property (B1) does not hold if b is a negative integer. Hence, the dimensional relation does not apply if D is an even integer and $k/2<1$. It is valid, however, if $k/2>1$. Relationship (B2) shows that the susceptibility of, e.g., odd-numbered dimensions is related to the susceptibility of *other* odd-numbered dimensions. There is no "mixing" of the two families.

The dimensional relation (B2) contains some useful properties. Let $z=(k/2)^2=1$ and $D\to D-2$. Then we obtain

$$\tilde{\chi}(D,k=2) = 1/(D-1). \tag{B4}$$

Property (B4) was previously obtained by Sharma and Lee[8] by directly evaluating the hgf. Also, since $\tilde{\chi}(D+2,k) \geqslant 0$, we obtain the inequality for $D>0$:

$$1 + (1-z)D\tilde{\chi}(D,k) \geqslant 0. \tag{B5}$$

For $z = (k/2)^2 > 1$, we obtain the bounds

$$0 \leqslant \tilde{\chi}(D,k) \leqslant 1/((z-1)D). \tag{B6}$$

For the large-k behavior of $\chi(k)$ and $\mathrm{Re}\,\chi(k\omega)$, see Appendix C.

APPENDIX C: LARGE-k BEHAVIOR OF $\chi(k,\omega)$

The knowledge of large-k behavior of the dynamical susceptibility is necessary when evaluating asymptotic properties of, e.g., the dynamic local field. For $\mathrm{Re}\,\chi(k,\omega)$, we use relation (15):

$$\mathrm{Re}\,\chi(k,\omega) = k^{-1}[s_+\chi(2s_+) + s_-\chi(2s_-)]. \tag{C1}$$

Hence, it is sufficient to know the large-k behavior of $\chi(2s_\pm)$. Sharma and Lee[8] have shown that

$$\chi(k\to\infty) = C(k^{-2} + O(k^{-4})), \tag{C2}$$

where $C = 2\rho/\epsilon_F$.

Hence,

$$\lim_{k\to\infty} \chi(2s_\pm) = C[k^{-2}(1 \pm \omega/k^2)^{-2} + \cdots]. \tag{C3}$$

Thus

$$\lim_{k\to\infty} \mathrm{Re}\,\chi(k\omega)/\chi(k) = 1 + \omega^2/k^4 + O(k^{-6}). \tag{C4}$$

If $k\to\infty$ while ω is fixed, one moves out of areas a or b into area d (see Fig. 3). Hence,

$$\lim_{k\to\infty} \mathrm{Im}\,\chi(k\omega)/\chi(k) = 0. \tag{C5}$$

Also, observe that because of the relationship between k and ω, i.e., $2s_\pm = k \pm \omega/k$, the $k\to\infty$ limit is formally equivalent to the $\omega\to 0$ limit. This kind of dependent property follows from the fact that the ω coordinate has been made into a translation of the k coordinate.

[1] G. D. Mahan, *Many Particle Physics* (Plenum, New York, 1981); see Chap. 5.

[2] P. F. Williams and A. N. Bloch, Phys. Rev. B **10**, 1097 (1974).

[3] M. Apostal, Z. Phys. B **22**, 279 (1975).

[4] W. I. Friesen and B. Bergersen, J. Phys. C **13**, 6627 (1980).

[5] F. Stern, Phys. Rev. Lett. **18**, 546 (1967).

[6] J. Hong and M. H. Lee, Phys. Rev. Lett. **55**, 2375 (1985) and J. Phys. Condens. Matter **1**, 3867 (1989).

[7] See, e.g., W. Marshall and R. D. Loude, Rep. Prog. Phys. **31**, 705 (1968).

[8] N. L. Sharma and M. H. Lee, J. Math. Phys. **27**, 1618 (1986).

[9] For recent examples, see V. Elias, R. B. Mann, A. M. Chowdhury, G. McKeon, S. Samant, and S. B. Phillips, Can. J. Phys. **63**, 1453 (1985); B. Duplantier, J. Phys. (Paris) **47**, 569 (1986).

[10] See, e.g., A. L. Fetter and J. D. Walecka, *Quantum Theory of Many-Particle Systems* (McGraw-Hill, New York, 1971), p. 159.

[11] See, e.g., *Table of Integral Transforms II*, edited by A. Erdelyi (McGraw-Hill, New York, 1954), p. 252 (No. 42).

Spherically symmetric black holes in Kaluza–Klein unification

Y. M. Cho and D. H. Park
Department of Physics, Seoul National University, Seoul 151-742, Korea

(Received 29 June 1989; accepted for publication 13 September 1989)

Exact time-independent spherically symmetric black hole solutions of the $(4+3)$-dimensional Kaluza–Klein theory are obtained by imposing the isometry group E_2 on the unified metric. The solutions include the five-dimensional Dobiasch–Maison–Lee solutions as a special case. The physical implication of the solutions is discussed.

I. INTRODUCTION

Recently some authors[1,2] have obtained the most general static spherically symmetric black hole solutions of the five-dimensional Kaluza–Klein theory. What makes the solutions so interesting is that some of them can exert a repulsive fifth force[3,4] which is generated by the Kaluza–Klein dilaton. The antigravity constitutes the first example that demonstrates how drastically the internal gravity[5,6] can modify gravitational attraction in a higher-dimensional unification. The purpose of this paper is to present the most general spherically symmetric black hole solutions in a seven-dimensional Kaluza–Klein unification that has E_2 as the internal isometry, and to establish *the existence of a universal fifth force that can be attractive as well as repulsive* in the unification. In particular, we show that the long-range nature of the fifth force is characterized by "the dilatonic charge of the black hole" that can be nonvanishing even when it is neutral. The existence of the novel dilatonic charge demonstrates that the conventional wisdom of the Birkoff theorem is no longer valid for the "black holes" of the higher-dimensional unification.

II. SOLUTIONS

A best way to construct a spherically symmetric Kaluza–Klein black hole solution is to impose a proper internal isometry[7,8] to the unified metric and reduce the higher-dimensional equation of motion to a set of four-dimensional ones. In this dimensional reduction by isometry the higher-dimensional metric is decomposed into the four-dimensional metric, the internal metric, and a well-defined set of non-Abelian gauge fields, whose internal-dependence is completely fixed by the isometry. Consider a $(4+n)$-dimensional unified space P endowed with a metric g_{AB} $(A,B=1,2,\ldots,4+n)$ which has a freely acting isometry G

$$\mathscr{L}_{\xi_a} g_{AB}=0,$$
$$[\xi_a,\xi_b]=(1/\kappa) f_{ab}{}^c\xi_c \quad (a,b,c=1,2,\ldots,n) \tag{1}$$

where ξ_a are the Killing vector fields and κ is the scale parameter that characterizes the size of the internal space.[7,8] With $M=P/G$ as the four-dimensional space-time one can identify P as a principal fiber bundle $P(M,G)$ on which G acts as the structure group. On the bundle the metric g_{AB} defines a connection and thus a horizontal-lift basis (D_μ,ξ_a) in a natural way, where $\mu,\nu(\mu,\nu=1,2,3,4)$ and a,b are space-time and the internal indices. In this basis the remaining part of the metric always has the block-diagonal form $g_{AB}=\gamma_{\mu\nu}\otimes\phi_{ab}$, where the internal metric ϕ_{ab} becomes gauge covariant.

Now suppose $G=E_2$ which can be obtained with the Wigner contraction of SU(2). The reason why we consider the isometry E_2 in this paper is that it naturally appears as a spontaneous contraction[9] of SU(2) which breaks the symmetry further down to U(1). In this case choosing a proper gauge one can always diagonalize the internal metric ϕ_{ab}, $(a,b=1,2,3)$

$$\phi_{ab}=\phi^{1/3}\rho_{ab}=\phi^{1/3}\mathrm{diag}(\rho_1^{-1},\rho_2^{-1},\rho_1\rho_2), \tag{2}$$

where $\phi=|\det\phi_{ab}|$ is the Kaluza–Klein dilaton field. This tells us that the internal metric has three physical degrees of freedom, the dilaton and two additional scalar fields ρ_1 and ρ_2. In this gauge, the seven-dimensional line element takes the form

$$\begin{aligned} ds_7^2 &= \phi^{-1/2}g_{\mu\nu}\,dx^\mu\,dx^\nu+\kappa\phi^{1/3}\rho_1^{-1}(\omega^1+eB_\mu^1\,dx^\mu)^2\\ &\quad+\kappa\phi^{1/3}\rho_2^{-1}(\omega^2+eB_\mu^2\,dx^\mu)^2\\ &\quad+\kappa\phi^{1/3}\rho_1\rho_2(\omega^3+eB_\mu^3dx^\mu)^2, \end{aligned} \tag{3}$$

where $g_{\mu\nu}=\sqrt{\phi}\gamma_{\mu\nu}$ is the Einstein part of the unified metric,[5,6] ω^a are the left-invariant one-forms of the internal isometry E_2 [with ω^3 as the one-form of the U(1) subgroup], and B_μ^a are the corresponding gauge potentials.

The vacuum solution of the seven-dimensional Einstein equation with the flat $g_{\mu\nu}$ and the vanishing B_μ^a is given by[9]

$$\bar\phi_{ab}=\mathrm{diag}\,(1,1,1), \tag{4}$$

so that at the vacuum the internal space becomes flat. The vacuum breaks the gauge symmetry E_2 down spontaneously to U(1) so that the first two bosons (B_μ^1 and B_μ^2) acquire a mass while the third one (B_μ^3) remains massless.[9]

With the above preliminaries we now proceed to obtain the desired solutions. For this we choose the ansatz

$$\rho_1=\rho_2=\rho, \quad B_\mu^a=A_\mu\delta_3^a, \tag{5}$$

and obtain the following equations of motion:

 0022-2488/90/030695-04$03.00

$$R_{\mu\nu}-\tfrac{1}{2}g_{\mu\nu}R=-8\pi G\phi^{5/6}\rho^2(g^{\alpha\beta}F_{\mu\alpha}F_{\nu\beta}-\tfrac{1}{4}g_{\mu\nu}F^2)$$
$$-\frac{5}{24}\left\{\left(\frac{\partial_\mu\phi}{\phi}\right)\left(\frac{\partial_\nu\phi}{\phi}\right)-\frac{1}{2}g_{\mu\nu}\left(\frac{\partial_\alpha\phi}{\phi}\right)^2\right\}$$
$$-\frac{3}{2}\left\{\left(\frac{\partial_\mu\rho}{\rho}\right)\left(\frac{\partial_\nu\rho}{\rho}\right)-\frac{1}{2}g_{\mu\nu}\left(\frac{\partial_\alpha\rho}{\rho}\right)^2\right\},$$
$$\left(\frac{\partial_\alpha\phi}{\phi}\right)^2-\frac{\nabla^2\phi}{\phi}=-8\pi G\phi^{5/6}\rho^2F^2,$$
$$\left(\frac{\partial_\alpha\rho}{\rho}\right)^2-\frac{\nabla^2\rho}{\rho}=-\frac{8\pi G}{3}\phi^{5/6}\rho^2F^2, \tag{6}$$
$$\nabla^\mu F_{\mu\nu}+\frac{5}{6}\frac{\partial^\mu\phi}{\phi}F_{\mu\nu}+2\frac{\partial^\mu\rho}{\rho}F_{\mu\nu}=0,$$

where $F_{\mu\nu}$ is the field strength of the U(1) subgroup. To obtain a static spherically symmetric solution we further assume

$$ds_4^2=-A(r)dt^2+B(r)(dr^2+r^2\,d\Omega^2),$$
$$A_\mu=\begin{cases}\Phi(r), & \mu=t,\\ m(1-\cos\theta), & \mu=\varphi,\\ 0, & \text{otherwise},\end{cases} \tag{7}$$
$$\phi=\phi(r),\quad \rho=\rho(r),$$

where m is the magnetic charge of the U(1) subgroup. With this the equations of motion are reduced to

$$\frac{A''}{A}-\frac{A'}{A}\left(\frac{A'}{2A}-\frac{B'}{2B}-\frac{2}{r}\right)$$
$$=8\pi G\phi^{5/6}\rho^2\left(\frac{\Phi'^2}{A}+\frac{m^2}{r^4B}\right),$$
$$\frac{A''}{A}-\frac{A'^2}{2A^2}+2\frac{B''}{B}-2\frac{B'^2}{B^2}$$
$$-\frac{A'B'}{2AB}+\frac{2B'}{rB}+\frac{5}{12}\frac{\phi'^2}{\phi^2}+3\frac{\rho'^2}{\rho^2}$$
$$=8\pi G\phi^{5/6}\rho^2\left(\frac{\Phi'^2}{A}+\frac{m^2}{r^4B}\right),$$
$$\frac{B''}{B}+\frac{B'}{B}\left(\frac{A'}{2A}-\frac{B'}{B}+\frac{3}{r}\right)+\frac{A'}{rA}$$
$$=-8\pi G\phi^{5/6}\rho^2\left(\frac{\Phi'^2}{A}+\frac{m^2}{r^4B}\right), \tag{8}$$
$$\frac{\phi''}{\phi}+\frac{\phi'}{\phi}\left(\frac{A'}{2A}+\frac{B'}{2B}-\frac{\phi'}{\phi}+\frac{2}{r}\right)$$
$$=-16\pi G\phi^{5/6}\rho^2\left(\frac{\Phi'^2}{A}-\frac{m^2}{r^4B}\right),$$
$$\frac{\rho''}{\rho}+\frac{\rho'}{\rho}\left(\frac{A'}{2A}+\frac{B'}{2B}-\frac{\rho'}{\rho}+\frac{2}{r}\right)$$
$$=-\frac{16\pi G}{3}\phi^{5/6}\rho^2\left(\frac{\Phi'^2}{A}-\frac{m^2}{r^4B}\right),$$
$$\Phi'=-\frac{q}{r^2}\frac{1}{\phi^{5/6}\rho^2}\left(\frac{A}{B}\right)^{1/2},$$

where the primes denote the derivative with respect to r and q is an integration constant which we will later identify as the electric charge of the U(1) subgroup. From these we obtain

$$AB=(1-\mu^2/r^2)^2 \tag{9}$$

and

$$\phi^{1/3}\rho^{-1}=((r-\mu)/(r+\mu))^{6c/\mu}, \tag{10}$$

where μ^2 and c are integration constants. To simplify (8) we put

$$\Sigma_\pm^3=A^{-3}\phi^{\pm 5/6}\rho^{\pm 2} \tag{11}$$

and

$$f'=4\pi G\frac{q^2}{r^2-\mu^2}\frac{\Sigma_-}{\Sigma_+^2},\quad g'=4\pi G\frac{m^2}{r^2-\mu^2}\frac{\Sigma_+}{\Sigma_-^2}. \tag{12}$$

The remaining equations then become

$$(r^2-\mu^2)\frac{\Sigma_+'}{\Sigma_+}=-4f,$$
$$(r^2-\mu^2)\frac{\Sigma_-'}{\Sigma_-}=-4g, \tag{13}$$
$$(r^2-\mu^2)(f'+g')=4(f^2-fg+g^2+15c-\mu^2),$$

$$\Phi=\frac{1}{4\pi Gq}(f|_{r\to\infty}-f).$$

With this we now present the solutions by the values of q and m.

(i) $q=m=0$. In this case f and g become constants, $f=a$ and $g=b$. So from (13) we have

$$\Sigma_+=\left(\frac{r-\mu}{r+\mu}\right)^{-2a/\mu},\quad \Sigma_-=\left(\frac{r-\mu}{r+\mu}\right)^{-2b/\mu},$$
$$a^2-ab+b^2+15c^2=\mu^2.$$

Introducing the new parameters α,β,γ by

$$\alpha=a+b,\quad \beta=-2(a-b-4c),$$
$$\gamma=-\tfrac{2}{3}(a-b+5c), \tag{14}$$

we obtain the following solution:

$$A=\left(\frac{r-\mu}{r+\mu}\right)^{\alpha/\mu},\quad \phi=\left(\frac{r-\mu}{r+\mu}\right)^{\beta/\mu},\quad \rho=\left(\frac{r-\mu}{r+\mu}\right)^{\gamma/\mu}, \tag{15}$$

where $\alpha^2+\frac{5}{12}\beta^2+3\gamma^2=4\mu^2$. So one can always assume $\mu>0$ in this case. Notice that when $\gamma=\beta/3$ the above solutions reduce to the five-dimensional solutions.[1,2] In particular, when $2\alpha=\beta=3\gamma=2\mu$ and $\alpha=\kappa/8$ the solution represents the five-dimensional regular soliton.[3] All of these solutions except the Schwarzschild solution ($\beta=\gamma=0$ and $\alpha=2\mu$) have a four-dimensional naked point singularity at $r=\mu$.

(ii) $q=0$. In this case f becomes constant, $f=a$. So from (13) we have

$$\Sigma_+=((r-\mu)/(r+\mu))^{-2a/\mu},$$
$$2g=a+D\{[(2b-a+D)+(2b-a-D)$$
$$\times((r-\mu)/(r+\mu))^{2D/\mu}]$$
$$\times[(2b-a+D)-(2b-a-D)$$
$$\times((r-\mu)/(r+\mu))^{2D/\mu}]^{-1}\},$$

where

$$D^2=4\mu^2-3a^2-60c^2,\quad b=g|_{r\to\infty}.$$

Using this we find

$$\Sigma_- = \left\{\frac{2b-a+D}{2D} - \frac{2b-a-D}{2D}\left(\frac{r-\mu}{r+\mu}\right)^{2D/\mu}\right\}$$
$$\times((r-\mu)/(r+\mu))^{-(a+D)/\mu}.$$

Inserting $\Sigma_\pm$ and g in (12) we have

$$(2b-a)^2 - D^2 = 4\pi G m^2;$$

thus we obtain the following solution:

$$A = \Psi^{-1/2}((r-\mu)/(r+\mu))^{\alpha/\mu},$$
$$\phi = \Psi^{-1}((r-\mu)/(r+\mu))^{\beta/\mu}, \quad (16)$$
$$\rho = \Psi^{-1/3}((r-\mu)/(r+\mu))^{\gamma/\mu},$$

where

$$\alpha^2 + \tfrac{5}{12}\beta^2 + 3\gamma^2 - 4\pi G m^2 = 4\mu^2,$$
$$\Psi = \frac{\eta+D}{2D}\left(\frac{r-\mu}{r+\mu}\right)^{(\eta-D)/\mu} - \frac{\eta-D}{2D}\left(\frac{r-\mu}{r+\mu}\right)^{(\eta+D)/\mu},$$
$$D^2 = \eta^2 - 4\pi G m^2, \quad \eta = (6\alpha + 5\beta + 12\gamma)/12.$$

Here α, β, γ, and m can be chosen as the independent parameters so that μ can become pure imaginary. Notice that when $\gamma = \beta/3$, the above solutions reduce to the five-dimensional solutions.[1,2] In particular, when $2\alpha = \beta = 3\gamma = \sqrt{4\pi G}\, m$ and $\alpha = \kappa/8$ the solution represents the Gross–Perry–Sorkin monopole[3] with magnetic charge $m = 1/2e$. This monopole is perfectly regular as a five-dimensional object, but from the four-dimensional point of view the monopole has a naked singularity at $r = \mu$.

(iii) $m = 0$. In this case g becomes constant, $g = b$. So from (13) we have

$$\Sigma_- = ((r-\mu)/(r+\mu))^{-2b/\mu},$$
$$2f = b + D\{[(2a-b+D) + (2a-b-D)$$
$$\times((r-\mu)/(r+\mu))^{2D/\mu}]$$
$$\times[(2a-b+D) - (2a-b-D)$$
$$\times((r-\mu)/(r+\mu))^{2D/\mu}]^{-1}\},$$

where

$$D^2 = 4\mu^2 - 3b^2 - 60c^2, \quad a = f|_{r\to\infty}.$$

From this we have

$$\Sigma_+ = \left\{\frac{2a-b+D}{2D} - \frac{2a-b-D}{2D}\left(\frac{r-\mu}{r+\mu}\right)^{2D/\mu}\right\}$$
$$\times((r-\mu)/(r+\mu))^{-(b+D)/\mu}.$$

Inserting $\Sigma_\pm$ and f in (12) we find

$$(2a-b)^2 - D^2 = 4\pi G q^2;$$

thus we obtain the following solution

$$A = \Psi^{-1/2}((r-\mu)/(r+\mu))^{\alpha/\mu},$$
$$\phi = \Psi((r-\mu)/(r+\mu))^{\beta/\mu},$$
$$\rho = \Psi^{1/3}((r-\mu)/(r+\mu))^{\gamma/\mu},$$
$$\Phi = \frac{q}{2}\left[\left(\frac{r-\mu}{r+\mu}\right)^{-D/\mu} - \left(\frac{r-\mu}{r+\mu}\right)^{D/\mu}\right]$$
$$\times\left[(\eta+D)\left(\frac{r-\mu}{r+\mu}\right)^{-D/\mu} - (\eta-D)\left(\frac{r-\mu}{r+\mu}\right)^{D/\mu}\right]^{-1}, \quad (17)$$

where

$$\alpha^2 + \tfrac{5}{12}\beta^2 + 3\gamma^2 - 4\pi G q^2 = 4\mu^2,$$
$$\Psi = \frac{\eta+D}{2D}\left(\frac{r-\mu}{r+\mu}\right)^{(\mu-D)/\mu} - \frac{\eta-D}{2D}\left(\frac{r-\mu}{r+\mu}\right)^{(\mu+D)/\mu},$$
$$D^2 = \eta^2 - 4\pi G q^2, \quad \eta = (6\alpha - 5\beta - 12\gamma)/12.$$

Here the independent parameters are α, β, γ, and q. When $\gamma = \beta/3$, the above solutions reduce to the five-dimensional solutions.[1]

(iv) $q \neq 0$ and $m \neq 0$. This is the most general case. From (12) and (13), we have

$$\Sigma_+ \frac{d^2\Sigma_+}{dz^2} - \left(\frac{d\Sigma_+}{dz}\right)^2 + 16\pi G q^2 \Sigma_- = 0,$$
$$\Sigma_- \frac{d^2\Sigma_-}{dz^2} - \left(\frac{d\Sigma_-}{dz}\right)^2 + 16\pi G m^2 \Sigma_+ = 0,$$

where

$$z = (1/2\mu)\log[(r-\mu)/(r+\mu)].$$

The solutions of the above equations are given by[10]

$$\Sigma_+ = 16\pi G(m^2q^4)^{1/3}$$
$$\times\left[\frac{e^{w_1(z-z_1)}}{(w_1-w_2)(w_1-w_3)} + \frac{e^{w_2(z-z_2)}}{(w_2-w_1)(w_2-w_3)} + \frac{e^{w_3(z-z_3)}}{(w_3-w_1)(w_3-w_2)}\right], \quad (18)$$
$$\Sigma_- = 16\pi G(m^4q^2)^{1/3}$$
$$\times\left[\frac{e^{-w_1(z-z_1)}}{(w_1-w_2)(w_1-w_3)} + \frac{e^{-w_2(z-z_2)}}{(w_2-w_1)(w_2-w_3)} + \frac{e^{-w_3(z-z_3)}}{(w_3-w_1)(w_3-w_2)}\right],$$

where w_i and z_i ($i = 1,2,3$) are integration constants that satisfy

$$w_1 + w_2 + w_3 = 0, \quad w_1z_1 + w_2z_2 + w_3z_3 = 0. \quad (19)$$

Asymptotically flat boundary conditions require

$$\Sigma_\pm |_{z=0} = 1. \quad (20)$$

Inserting $\Sigma_\pm$ in (13) we find

$$w_1^2 + w_2^2 + w_3^2 = 32(\mu^2 - 15c^2); \quad (21)$$

thus we obtain the following complete set of solution:

$$A = (\Sigma_+\Sigma_-)^{-1/2},$$
$$\phi = \left(\frac{\Sigma_+}{\Sigma_-}\right)\left(\frac{r-\mu}{r+\mu}\right)^{8c/\mu}, \quad \rho = \left(\frac{\Sigma_+}{\Sigma_-}\right)^{1/3}\left(\frac{r-\mu}{r+\mu}\right)^{-10c/3\mu},$$
$$\Phi = \frac{1}{16\pi G q}\left[\frac{1}{\Sigma_+}\frac{d\Sigma_+}{dz} - \frac{1}{\Sigma_+}\frac{d\Sigma_+}{dz}\bigg|_{z=0}\right]. \quad (22)$$

Here the ten parameters μ, m, q, w_i, z_i, and c are constrained by the five equations (19)–(21) so that the solutions have five-independent parameters. When $c = 0$ the solutions reduce to the five-dimensional solutions.[2] Notice that all the solutions except the Dobiasch–Maison solutions[1] have a four-dimensional naked singularity.

III. CONCLUSION

This completes our list of the static and spherically symmetric black hole solutions. Strictly speaking, not all the solutions can be called black holes because many of them have a naked singularity at the origin. For simplicity, however, we will continue to call them black holes. A characteristic feature of the above solutions is that they carry additional "charges" that are forbidden by the Birkoff theorem. To see this, consider the "neutral" black hole (15). Even though it is neutral in the sense that it carrier no "electromagnetic" charges of the unbroken U(1) gauge group, it is characterized by three (not one) parameters α, β, and γ. Of these, α determines the inertial mass of the black hole because it describes the asymptotic behavior of the metric. But β and γ are new parameters. Since they describe the asymptotic behavior of ϕ and ρ_{ab}, they may be interpreted as the dilatonic and internal charges, respectively. With this interpretation, the physical meaning of the five parameters of the most general spherically symmetric black hole (22) becomes clear. They are the inertial mass, the electric, magnetic, dilatonic, and internal charges of the solution.

To discuss the physical meaning of the new dilatonic charge let us consider the motion of a neutral test particle around the black hole (15). In this case, the higher-dimensional geodesic equation gives us the following four-dimensional equation[11]:

$$\frac{dp^\mu}{ds} + \Gamma_{\alpha\beta}{}^{\mu} p^\alpha p^\beta = \frac{1}{4}(\epsilon_4 g^{\mu\alpha} + p^\mu p^\alpha)\,\frac{\partial_\alpha \phi}{\phi}, \tag{23}$$

where p^μ and s are the four-dimensional momentum and affine parameter of the test particle, and ϵ_4 is the signature factor of the four-dimensional trajectory. Notice that *the right-hand side of (23) characterizes the universal fifth force generated by the dilaton.* To understand the effect of the fifth force, notice that the inertial mass of the black hole (15) (obtained with the Landau pseudo-energy–momentum tensor) is given by

$$m_i = \alpha/G. \tag{24}$$

On the other hand in the Newtonian limit (23) gives the following effective gravitational mass of the black hole:

$$m_g = (2\alpha - \beta)/2G. \tag{25}$$

From this we deduce that the dilaton generates an attractive (or repulsive) fifth force when its charge β is negative (or positive). Clearly this is a generalization of the five-dimensional antigravity effect of Gross and Perry,[3,4] which they obtained when $2\alpha = \beta$.

One may wonder whether the above result implies a violation of the equivalence principle in the Kaluza–Klein unification. According to the equivalence principle, the motion of a neutral test particle under the presence of gravitation should be described by a geodesic. In this sense it is obvious that (23) violates the equivalence principle. Indeed the result (25) tells that the *observed* "gravitational" mass of the black hole must include the contribution of the dilaton. On the other hand, it must also be emphasized that this apparent violation of the equivalence principle is *not* due to the gravitation, but due to the fifth force. In fact, (23) guarantees that the gravitation never violates the equivalence principle, as far as one can find a way to separate the pure gravitation from the fifth force. The real question then is whether this separation can be made possible. Our analysis shows that this is not possible with a neutral test particle. Using a charged test particle, however, one can show that in principle one can separate the fifth force from the pure gravitation.[11] This is so because the fifth force is charge dependent in general. This assures that there is no violation of the equivalence principle for the gravitation.

Finally we point out the existence of the dilatonic charge may not be a generic feature of the Kaluza–Klein black holes. Indeed our result (25) is a direct consequence of the fact that the dilaton in the above solution is massless. This allows the dilaton to generate the long-range fifth force that can directly compete with the gravitational attraction. In a realistic model, however, the dilaton may acquire a mass due to a spontaneous breaking of the scaling invariance.[3,5] Once the dilaton acquires a mass the fifth force becomes short ranged, so that the dilatonic charge will vanish. Nevertheless the implication of the above analysis is unmistakable. In any higher-dimensional unification (supersymmetric or not) in which the Kaluza–Klein dilaton becomes very light, the appearance of a fifth force that can alter the gravitational attraction significantly is unavoidable. A detailed analysis will be published elsewhere.[11]

[1] P. Dobiasch and D. Maison, Gen. Rel. Grav. **14**, 231 (1982); A. Chodos and S. Detweiler, *ibid.* **14**, 879 (1982).

[2] H. M. Lee and S. C. Lee, Phys. Lett. B **149**, 95 (1984); S. C. Lee, *ibid.* **149**, 98 (1984); I. G. Angus, Nucl. Phys. B **264**, 349 (1986).

[3] D. J. Gross and M. J. Perry, Nucl. Phys. B **226**, 29 (1983); R. Sorkin, Phys. Rev. Lett. **51**, 87 (1983).

[4] D. Pollard, J. Phys. A **16**, 565 (1983); J. Scherk, Phys. Lett. B **88**, 265 (1979).

[5] Y. M. Cho, Phys. Rev. D **35**, 2628 (1987).

[6] Y. M. Cho, Phys. Lett. B **199**, 358 (1987).

[7] Y. M. Cho, J. Math. Phys. **16**, 2029 (1975); Y. M. Cho and P. G. O. Freund, Phys. Rev. D **12**, 1711 (1975); Y. M. Cho and P. S. Jang, Phys. Rev. D **12**, 3789 (1975).

[8] Y. M. Cho, Phys. Lett. B **186**, 38 (1987); Y. M. Cho and D. S. Kimm, J. Math. Phys. **30**, 1570 (1989).

[9] Y. M. Cho, Phys. Rev. Lett. **55**, 2932 (1985); Y. M. Cho and P. Y. Pac, Mod. Phys. Lett. A **3**, 1473 (1988).

[10] B. Kostant, Adv. Math. **34**, 195 (1979).

[11] Y. M. Cho and P. Y. Pac, submitted to Phys. Lett. B; Y. M. Cho and D. H. Park, submitted to Gen. Rel. Grav.

Affine fields and operator representations for the nonlinear σ model

Christopher J. Isham
The Blackett Laboratory, Imperial College, Prince Consort Road, South Kensington, London SW7 2BZ, United Kingdom
John R. Klauder
Departments of Physics and Mathematics, University of Florida, Gainesville, Florida 32611

(Received 5 September 1989; accepted for publication 27 September 1989)

The problem of operator representations and associated Hamiltonians for quantum fields and their conjugates appropriate to the nonlinear σ model is discussed. To respect the target-space constraints, canonical field momenta are rejected in favor of alternative affine conjugate fields that serve to generate constraint-preserving group transformations in the target space. Various operator representations are discussed at length, some of which do not admit canonically conjugate fields. After establishing certain general properties for Hamiltonians, there is an analysis of how operator field representations and Hamiltonians can be linked.

I. INTRODUCTION

As a mathematical model for classical and quantum field theory, the nonlinear σ model continues to be of considerable interest. It frequently serves as a paradigm for constrained systems and is of particular relevance to those working in quantum gravity since its perturbative renormalizability properties are the same as those of general relativity. The σ model is in fact nonrenormalizable in space-time dimensions greater than two, and our study of possible nonperturbative quantizations is strongly motivated by the desire to find a nonperturbative approach to quantum gravity. The purpose of the present paper is to begin such a program based on an investigation of the unitary representations of certain groups which are naturally adapted to the global structure of the highly nonlinear configuration space of the classical σ model.

The simplest such model is when the field variables $u^1(x),u^2(x),...,u^N(x)$ (where $1<N<\infty$ and x is a point in the s-dimensional spatial manifold Σ) are forced to lie on the surface of an $(N-1)$-dimensional sphere S^{N-1} in $\mathbb{R}^N$ via the constraint

$$\sum_{a=1}^{N} u^a(x)u^a(x) = c \tag{1.1}$$

for some positive constant c. Conventional canonical fields $\varphi^a(x), a=1,...,N$ possess canonical conjugates $\pi^b(x)$ satisfying the classical Poisson brackets

$$\{\varphi^a(x),\pi^b(y)\} = \delta^{ab}\delta(x,y), \quad a,b=1,...,N, \tag{1.2}$$

where the Dirac delta function $\delta(x,y)$ is defined with respect to a suitable volume element ϑ on the spatial manifold Σ. However, for the nonlinear σ model, the adoption of such fields $\pi^b(x)$ (which essentially induce translations in the target space) would be incompatible with the constraint (1.1) and an alternative set of conjugate fields is required. Of course, one possibility would be to solve the constraint for the "physical" configuration variables and then to construct the conventional conjugates for this reduced set. This is essentially what is achieved by the use of Dirac brackets or related methods. However, this can only be done locally on the target space and is also likely in practice to involve a weak field perturbative expansion via, for example, the use of normal coordinates.[1] Both these restrictions are undesirable from our perspective and we wish instead to retain the full set $u^a(x), a=1,...,N$ of variables. This can be done consistently if we select canonical "conjugates" that induce *rotations* in the target space rather than *translations*. This would involve a set of fields $J_{ab}(x)(=-J_{ba}(x))$, $a,b=1,...,N$ which satisfy the Poisson brackets

$$\{J_{ab}(x),J_{cd}(y)\} = (\delta_{ad}J_{bc}(x) - \delta_{bd}J_{ac}(x) + \delta_{bc}J_{ad}(x) - \delta_{ac}J_{bd}(x))\delta(x,y) \tag{1.3}$$

and which induce the infinitesimal rotations on the $u^a(x)$ variables given by

$$\{u^a(x),J_{cd}(y)\} = (\delta^a{}_c u_d(x) - \delta^a{}_d u_c(x))\delta(x,y)\,. \tag{1.4}$$

Unlike (1.2), these new Poisson brackets are manifestly consistent with the constraint (1.1).

Together with

$$\{u^a(x),u^b(y)\} = 0\,, \tag{1.5}$$

Eqs. (1.3), (1.4) form the fundamental kinematical brackets for the S^{N-1} valued nonlinear σ model, just as (1.2) and the relations $\{\varphi^a(x),\varphi^b(y)\} = \{\pi^a(x),\pi^b(y)\} = 0$ represent the fundamental kinematical brackets for conventional, unconstrained fields. It is important to observe that the variable

$$T(x) := \sum_{a=1}^{N} u^a(x)u^a(x) \tag{1.6}$$

has a vanishing Poisson bracket with $u^a(x)$ and $J_{bc}(y)$ and is thus a Casimir operator for the algebra represented by (1.3)–(1.5).

Just as the aim of conventional canonical quantization is to find operator realizations of the canonical fields $\varphi^a(x)$

 0022-2488/90/030699-13$03.00

and $\pi^b(y)$ that satisfy the commutator analog of the fundamental kinetmatical brackets so, it can be argued, the aim of quantization of the nonlinear σ model is to find operator realizations of the fields $u^a(x)$ and $J_{bc}(y)$ that satisfy the commutator analog of the fundamental brackets (1.3)–(1.5). The purpose of this paper is to analyze several operator representatives of precisely this type.

This procedure can be further justified by noting that, in effect, we are considering a special case of a rather general approach to quantizing a system whose classical phase space $\mathfrak{S}$ is a homogeneous space, i.e., one that admits a transitively acting group $\mathfrak{G}$ of symplectic transformations. As has been discussed at length,[2] the act of constructing unitary representations of this group (or perhaps of a central extension) can be regarded as the correct analog for $\mathfrak{S}$ of the familiar quantization of a system whose phase space is a (topologically trivial) vector space and where the canonical group $\mathfrak{G}$ is simply the associated Weyl–Heisenberg group. In the case of the nonlinear σ model, the configuration space is the infinite-dimensional manifold $C(\Sigma,G/H)$ of maps from Σ into the homogeneous space G/H, where G is a compact Lie group; for example, when the target space is the $(N-1)$ sphere S^{N-1}, G is $\mathrm{SO}(N)$, and H is an $\mathrm{SO}(N-1)$ subgroup. Generally speaking the maps concerned will be smooth and of compact support, although occasionally it will be useful to consider a wider class.

The target space G/H can always be embedded in some real vector space W on which G acts linearly in such a way that one of its orbits is precisely G/H, i.e., H appears as the "little group" of some vector in W.[3,4] Typically, the orbit G/H will be the zero set of some set of G-invariant functions on W which thus generate the constraints of the system; for S^{N-1} the carrier space W is simply the Euclidean space $\mathbb{R}^N$. Note that since G is compact, there is no loss of generality in assuming that there is an inner product on W with respect to which the action of G [denoted $L(g)$, $g\in G$] is via orthogonal operators. These will leave invariant the spheres embedded in W, and hence the "quadratic Casimir" constraint (1.1) will always be among the set of constraints defining the model. We will concentrate entirely on this quadratic Casimir in the present paper.

The space $C(\Sigma,G/H)$ is topologically quite complicated, but if we restrict our attention to the subspace $C_{(0)}(\Sigma,G/H)$ of maps from Σ into G/H that can be *lifted* to G, then it can be shown[5] that the group $C(\Sigma,G)$ acts transitively on this subspace; a group which acts transitively on the entire phase space $T^*C_{(0)}(\Sigma,G/H)$ is then $\mathfrak{G}:=C(\Sigma,W^*\circledS G)\approx C(\Sigma,W^*)\circledS C(\Sigma,G)$, where the semidirect product $\circledS$ is with respect to the conjugate action of G on the dual vector space W^*. [N.B. This is an analog for function spaces of the argument in Ref. 2 which shows that the canonical group for a system whose classical phase space is $\mathfrak{S}=T^*(G/H)$ is $W^*\circledS G$.]

For the case where the target space is S^{N-1}, the corresponding canonical group will be $C(\Sigma,\mathbb{R}^N)\circledS C(\Sigma,\mathrm{SO}(N))$ whose Lie algebra is indeed precisely that of the Poisson bracket relations (1.3)–(1.5). For the general G/H-valued σ model, we must find unitary representations of the group $C(\Sigma,W^*)\circledS C(\Sigma,G)$ with the unitary operators $V(\Lambda)$, $U(\alpha)$, $\Lambda\in C(\Sigma,G)$, $\alpha\in C(\Sigma,W^*)$ satisfying

$$V(\Lambda_1)V(\Lambda_2)=V(\Lambda_1\Lambda_2)\,, \tag{1.7}$$

$$V(\Lambda)U(\alpha)V(\Lambda)^{-1}=V(\widetilde{L}(\Lambda)\alpha)\,, \tag{1.8}$$

$$U(\alpha_1)U(\alpha_2)=U(\alpha_1+\alpha_2)\,, \tag{1.9}$$

where $\widetilde{L}(g)$ denotes the conjugate representation of $g\in G$ on the vector space W^* defined by $\langle\widetilde{L}(g)u,w\rangle:=\langle u,L(g)w\rangle$, $u\in W^*$, $w\in W$, and $\langle\,,\rangle$ is the usual pairing between W and its dual W^*. Note that in (1.7)–(1.8) the group products are understood to be pointwise in Σ; for example, $\widetilde{L}(\Lambda)\alpha$ means the element in $C(\Sigma,W^*)$ defined by $(\widetilde{L}(\Lambda)\alpha)(x):=\widetilde{L}(\Lambda(x))\alpha(x)$. The Lie algebra of the canonical group $\mathfrak{G}$ is the space of smooth maps from Σ into the Lie algebra $\mathfrak{L}(W^*\circledS G)$ of $W^*\circledS G$ and the associated, self-adjoint operators $J(f)$ and $u(\alpha)$, where $f\in C(\Sigma,\mathfrak{L}(G))$, $\alpha\in C(\Sigma,\mathfrak{L}W^*))$, are defined via

$$V(\exp f)=e^{-iJ(f)},\quad U(\alpha)\equiv U(\exp\alpha)=e^{-iu(\alpha)}\,, \tag{1.10}$$

where, as usual for an Abelian group, we identify a Lie algebra element with its exponential in the Lie group. The corresponding commutation relations are

$$[J(f_1),J(f_2)]=iJ([f_1f_2])\,, \tag{1.11}$$

$$[J(f),u(\alpha)]=iu(\widetilde{L}(f)\alpha)\,, \tag{1.12}$$

$$[u(\alpha_1),u(\alpha_2)]=0\,, \tag{1.13}$$

with $[f_1f_2](x):=[f_1(x)f_2(x)]$, where $[\ ]$ denotes the Lie bracket in $\mathfrak{L}(G)$. With respect to a basis set $(e^1,e^2,\ldots,e^N)$ of W^*, and a basis set $\{E_1,E_2,\ldots,E_{\dim(G)}\}$ of $\mathfrak{L}(G)$, the unsmeared form of these relations is

$$[J_i(x),J_j(y)]=iC_{ij}{}^k\delta(x,y)J_k(x), \tag{1.11'}$$

where $[E_iE_j]=C_{ij}{}^kE_k$, and

$$[J_i(x),u^a(y)]=iu^b(x)L_{ib}{}^a\delta(x,y)\,, \tag{1.12'}$$

$$[u^a(x),u^b(y)]=0\,, \tag{1.13'}$$

where $L_{ib}{}^a$ is the set of real constants describing the action of $\mathfrak{L}(G)$ on W^*.

Commutation relations like (1.11)–(1.13) must be understood as referring (implicitly) to a common dense domain for the operators concerned. In our case, domains of this type will usually be composed of finite sums of coherent states or the like. However, we will not be unduly pedantic about this point since we will mainly be studying unitary representations of the *group* $\mathfrak{G}$, in which case the self-adjoint generators can be constructed in the usual way with the aid of Stone's theorem.

The algebra (1.11)–(1.13) belongs to a larger class of related algebras, sometimes called "affine algebras" and whose associated quantum fields are known as "affine fields." Thus our study is part of a larger program to find operator realizations of quantum fields of this type. Commutation relations like (1.11)–(1.13) are distinguished sharply from conventional canonical commutators by the absence of a central extension term, a difference which produces radical changes in the representation theory. Affine relations are not only relevant for constrained systems like the nonlinear σ model, but they are also appropratie for certain models in which the fields satisfy constraints that are *inequalities*, rather than equalities. Elsewhere it has been argued that the positive-definiteness of the three-metric in canonically quantized gravity can best be preserved through the use of affine fields and affine commutation relations.[6–9] Thus, at least tangentially, our present study of operator realizations of the nonlinear σ model may have some bearing on the deeper question of the quantization of the gravitational field.

A key issue in what follows will be the existence or nonexistence of an operator equivalent of the Casimir function $T(x)$ in (1.6), or its generalizations to the G/H-valued σ model. For the canonical commutation relations, the standard goal is to search for irreducible representations of the canonical fields, since these are traditionally regarded as relevant for model problems (although there are counterexamples to this conventional wisdom[10,11]). By analogy, one would expect to seek solutions of the nonlinear σ model affine commutation relations for which the fields $u(h)$, $J(f)$ are represented irreducibly. In the case of, for example, quantum theory for a system whose configuration space is the *finite*-dimensional sphere S^{N-1}, the canonical group is $\mathbb{R}^N$ (s) $\mathrm{SO}(N)$, and the operator $u^a u^a$ is a well-defined Casimir for this group, and must therefore be a multiple of the unit operator if the representation is to be irreducible. This is the means whereby the classical constraint is coded into the quantum theory. However, in the infinite-dimensional case the situation is somewhat different and in the limited cases which we will discuss, we find two classes of representation as far as the Casimir function $T(x)$ (1.6) is concerned. In one of these, $T(x)$ is a well-defined, nonconstant, unbounded local operator as it stands. However, in the second class $T(x)$ is only defined as an operator after a fairly severe regularization.

Section II of the paper deals with a class of representations in which the canonical momenta for the fields $u(h)$ also exist as well-defined operators. We shall display two types of representation here: one irreducible and one reducible. Section III, on the other hand, deals with representations of the affine commutation relations for which *no* canonically conjugate momenta exist. These are direct extensions of the ultralocal representations that have been used in other contexts to study idealized model problems in which terms in the Hamiltonian involving spatial gradients are dropped. Finally, in Sec. IV we consider the extent to which the Hamiltonian for an affine theory is determined by the cyclic representation of the affine algebra.

II. REPRESENTATIONS WITH CONJUGATE MOMENTA

A. A class of irreducible representations

We are interested in studying the representations of infinite-dimensional groups of the form $\mathfrak{G} := C(\Sigma, W^*$ (s) $G) \approx C(\Sigma, W^*)$ (s) $C(\Sigma, G)$. Representations of groups of the type $C(\Sigma, K)$ (where K is a Lie group) have been of interest to theoretical physicists since the days of current algebras and have been investigated in a variety of ways. Continuous tensor products[12,13] have been an important tool, as has the (not unrelated) line of research initiated in Gel'fand *et al.*[14–16] And of course, in recent years, there has been enormous interest in the special case of loop groups (i.e., where $\Sigma \approx S^1$). However, much (albeit not all) of this work has been concerned with the case where the target group K is compact, whereas for us, not only is $K = W^*$ (s) G noncompact, but it has the suggestive form of a semidirect product, as does the function group $\mathfrak{G} = C(\Sigma, K) \approx C(\Sigma, W^*)$ (s) $C(\Sigma, G)$ itself.

The simplest finite-dimensional analog is the Euclidean group $E^n \approx \mathbb{R}^n$ (s) $\mathrm{SO}(n)$, all of whose irreducible representations can be obtained using induced representation theory. This at once raises the question of the feasibility of using this method for the infinite-dimensional semidirect product group $\mathfrak{G}$. This question cannot be addressed in the usual way since infinite-dimensional groups like $\mathfrak{G} \approx C(\Sigma, W^*)$ (s) $C(\Sigma, G)$ are not locally compact and hence the standard theorems[17] of induced representation theory are not applicable. *A priori*, there is no reason to expect that an induced representation of $\mathfrak{G}$ will be irreducible or that all representations could be obtained this way. The analogy of the finite-dimensional case suggests at best that it might be worth studying the action of $C(\Sigma, G)$ on the topological dual $C(\Sigma, W^*)'$ of the real vector space $C(\Sigma, W^*)$ but, unlike the analogous case for E^n, one should not expect the representation functions to be necessarily concentrated on a single orbit of this action. Indeed, as we shall show, a more typical situation is one in which the action of $C(\Sigma, G)$ on $C(\Sigma, W^*)'$ is *strictly ergodic* with respect to some measure μ on $C(\Sigma, W^*)'$. Thus the action is not transitive on the support of μ but nevertheless the only measurable proper subsets of the support that are group-invariant have measure zero.

The general form of a representation of $C(\Sigma, W^*)$ (s) $C(\Sigma, G)$ can be found most easily if we assume that the topology on $C(\Sigma, W^*)'$ has been chosen to make it into a nuclear space (fortunately, this is not too difficult to arrange in general). We can also assume, with no real loss of generality, that the representation of $\mathfrak{G}$ is cyclic with respect to the Abelian subgroup $C(\Sigma, W^*)$. Then a straightforward adaptation of the well-known theorem for the infinite-dimensional Weyl–Heisenberg group (for example, Ref. 18) shows that if Ω is any cyclic vector, there exists a probability measure μ on $C(\Sigma, W^*)'$ such that the Hilbert space of the representation is isomorphic to $L^2(C(\Sigma, W^*)', d\mu)$ and the following applies.

(i) The measure μ is determined uniquely by its "Fourier transform" which satisfies

$$\langle \Omega, U(\alpha) U \rangle = \int_{C(\Sigma, W^*)'} e^{-i\langle \chi, \alpha \rangle} \, d\mu(\chi) . \tag{2.1}$$

Here $\langle \chi, \alpha \rangle$ denotes the value of distribution $\chi \in C(\Sigma, W^*)'$ when acting on the function $\alpha \in C(\Sigma, W^*)$.

(ii) The representation of $\mathfrak{G}$ on $\Psi \in L^2(C(\Sigma, W^*)', d\mu)$ is of the form

$$(U(\alpha)\Psi)(\varkappa) = e^{-i\langle\varkappa,\alpha\rangle}\Psi(\varkappa)\,, \tag{2.2}$$

$$(V(\Lambda)\Psi)(\varkappa) = \gamma_\Lambda(\varkappa)\Psi(L^*(\Lambda)\varkappa)\,, \tag{2.3}$$

where $\langle L^*(\Lambda)\varkappa,\alpha\rangle := \langle\varkappa,\widetilde{L}(\Lambda)\alpha\rangle$ and where μ is quasi-invariant under the action of $C(\Sigma,G)$ with a Radon–Nikodym derivative satisfying

$$d\mu(L^*(\Lambda)\varkappa) = |\gamma_\Lambda(\varkappa)|^2 d\mu(\varkappa)\,. \tag{2.4}$$

(iii) The representation of the group $C(\Sigma,W^*)$ Ⓢ $C(\Sigma,G)$ on $L^2(C(\Sigma,W^*)',d\mu)$ is irreducible if and only if the measure μ is ergodic with respect to the action of $C(\Sigma,G)$ on $C(\Sigma,W^*)'$.

Thus, as in the case of conventional induced representation theory, the problem is to find either transitive orbits carrying quasi-invariant measures or strictly ergodic group actions. To motivate our first example it is useful to consider the simple finite-dimensional case in which we are given an irreducible representation of the canonical commutation relations

$$[q_a,q_b] = [p_a,p_b] = 0, \quad [q_a,p_b] = i\delta_{ab}; \quad a,b = 1,\ldots,n\,. \tag{2.5}$$

Then the set of operators

$$J_{ab} := q_a p_b - q_b p_a \tag{2.6}$$

together with $q_1,q_2,\ldots,q_n$ provides a representation of the Euclidean group E^n. This representation is of course reducible and decomposes into a direct integral of irreducible representations labelled by the continuous eigenvalue of the Casimir operator $q_a q_a$. This simple example motivates us to study first the case in which the field variables $u(\alpha)$ admit a conjugate set $\pi(\beta)$ such that

$$[u(\alpha),\pi(\beta)] = i(\alpha,\beta)\,, \tag{2.7}$$

where (,) denotes some inner product which is bicontinuous on the real vector space $C(\Sigma,W^*)$. The simplest representation is the irreducible Fock representation which is specified uniquely by the Gaussian measure on $C(\Sigma,W^*)'$ with Fourier transform

$$\langle\Omega,U(\alpha)\Omega\rangle = \int_{C(\Sigma,W^*)'} e^{-i\langle\varkappa,\alpha\rangle}\, d\mu(\varkappa) = \exp - \tfrac{1}{4}(\alpha,\alpha)$$
$$\text{for all } \alpha\in C(\Sigma,W^*)\,. \tag{2.8}$$

The canonical fields act as

$$(e^{-iu(\alpha)}\Psi)(\varkappa) = e^{-i\langle\varkappa,\alpha\rangle}\Psi(\varkappa)\,, \tag{2.9}$$

$$(e^{-i\pi(\beta)}\Psi)(\varkappa) = \rho_\beta(\varkappa)\Psi(\varkappa-\beta)\,, \tag{2.10}$$

where $\alpha,\beta\in C(\Sigma,W^*)$ and $\rho_\beta(\varkappa) = (d\mu(\varkappa-\beta)/d\mu(\varkappa))^{1/2}$ is the square root of the Radon–Nikodym derivative of the measure μ on $C(\Sigma,W^*)'$ that is quasi-invariant under translations by $C(\Sigma,W^*)$. The critical question for us is whether this representation of the canonical commutation relations can be extended to include the group $\mathfrak{G}$. From what was said above it is clear that a necessary and sufficient condition for this is that, as well as being quasi-invariant under translations by $C(\Sigma,W^*)$, the Gaussian measure μ should also be quasi-invariant under the action of $C(\Sigma,G)$. We will consider first the case when the inner product (,) is invariant under $C(\Sigma,G)$; this is particularly simple since it implies that the measure μ is actually *invariant* (i.e., the Radon–Nikodym factor is trivial). As was mentioned earlier, since G is *compact*, we can assume without loss of generality that the representation $\widetilde{L}$ of G on W^* is orthogonal with respect to some inner product $(\ ,\ )_{W^*}$ on W^*. Then a natural $C(\Sigma,G)$ invariant inner product on $C(\Sigma,W^*)$ can be defined by

$$(\alpha,\beta) := \int_\Sigma (\alpha(x),\beta(x))_{W^*}\, d\vartheta(x)$$
$$= \int_\Sigma \alpha_a(x)\beta_a(x)\,d\vartheta(x)\,, \tag{2.11}$$

where ϑ is the volume element on Σ and where the basis $\{e^1,e^2,\ldots,e^N\}$ of W^* has been chosen so that $(e^a,e^b)_{W^*} = \delta^{ab}$.

The problem now is to decide whether or not the ensuing representation is irreducible. The formal expression for the Casimir is $T(x) := \Sigma u^a(x)u^a(x)$, but we can see at once that this is unlikely to be a well-defined operator since, in this Gaussian representation, $(\Omega,u(\alpha)u(\beta)\Omega) = \frac{1}{2}(\alpha,\beta)$, and hence

$$(\Omega,u^a(x)u^b(y)\Omega) = \tfrac{1}{2}\delta^{ab}\delta(x,y)\,. \tag{2.12}$$

Thus the putative Casimir would appear to be proportional to $\delta(0)$. A natural step at this stage might be to define a regularized Casimir as the operator $:\Sigma u^a(x)u^a(x):$ where $::$ denotes the usual normal ordering operation. Since this new operator differs from the unordered operator by a c number it would appear to have the same commutation properties as the latter, and in particular therefore to commute with the generators of the affine group. However, although $:\Sigma u^a(x)\, u^a(x):$ has finite matrix elements between, for example, any pair of coherent states, it does not define a genuine operator (even when smeared with a test function) but only a quadratic form. As such, it cannot serve as a proper Casimir for the representation.

However, another possibility is to define a regularized operator by "dividing" by $\delta(0)$ (rather than subtracting the singularity as in normal ordering) but this requires a more careful analysis of the singularity. Since we are only interested in the short distance behavior, we can illustrate the main idea by assuming that the s-dimensional manifold Σ is compact and endowed with a Riemannian metric whose associated volume element is ϑ. Let $w_1(x)$, $w_2(x)$,... denote the eigenfunctions of the Laplacian operator with eigenvalues $-\lambda_1, -\lambda_2,\ldots$. This set is orthonormal with respect to the inner product $(f,g) = \int f(x)g(x)d\vartheta(x)$ and is a basis set for the Hilbert space completion of $C(\Sigma,\mathbf{R})$ with respect to this inner product. Thus, formally

$$\delta(x,y) = \sum_{n=1}^{\infty} w_n(x)w_n(y)\,, \tag{2.13}$$

which enables us to define the "regularized" Casimir operator (sum over a being understood) as

$$T_M(r) := \sum_{\lambda_n<M^2} u^a(rw_n)u^a(w_n) \tag{2.14}$$

where the smearing function $r(x) > 0$ and where the sum is over all eigenvalues less than the regulating parameter M^2. (Note that M has the units of inverse length.) As $M \to \infty$, this family of operators converges formally to the smeared form of $T(x)$. The "vacuum" expectation value of $T_M(r)$ is given by (2.12) as

$$(\Omega, T_M(r)\Omega) = \frac{N}{2} \sum_{\lambda_n < M^2} \int_\Sigma d\vartheta(x) r(x) w_n(x) w_n(x) \tag{2.15}$$

and then Carleman's asymptotic relation[19]

$$\sum_{\lambda_n < M^2} w_n(x) w_n(x) = \frac{(M^2/4\pi)^{s/2}}{\Gamma(1+s/2)} (1 + O(M^{-1})) \tag{2.16}$$

shows that as $M \to \infty$, the matrix element $(\Omega, T_M \Omega)$ diverges as M^s. Since Ω is a cyclic vector with respect to the representation of the subgroup $C(\Sigma, W^*)$, the most general matrix element of T_M can be obtained from

$$(U(\alpha)\Omega, T_M(r) U(\beta)\Omega) = \int_{C(\Sigma, W^*)'} e^{-i(\chi, \beta - \alpha)} \sum_{\lambda_n < M^2} \chi^a(r w_n) \chi^a(w_n) d\mu(\chi). \tag{2.17}$$

But

$$\int e^{-i(\chi,\alpha)} \chi^a(x) \chi^b(y) d\mu(\chi) = \tfrac{1}{2} \delta^{ab} (\delta(x,y) - \tfrac{1}{2}\alpha(x)\alpha(y)) e^{-(1/4)(\alpha,\alpha)} \tag{2.18}$$

and hence

$$(U(\alpha)\Omega, T_M(r) U(\beta)\Omega) = \frac{N}{2} \int_\Sigma r(x) d\vartheta(x) \times \left(\frac{(M^2/4\pi)^{s/2}}{\Gamma(1+s/2)} (1 + O(M^{-1})) - (\beta - \alpha, \beta - \alpha) \right) e^{-(1/4)(\beta - \alpha, \beta - \alpha)} \tag{2.19}$$

and thus the matrix element (2.17) has the same M^s behavior as $(\Omega, T_M(r)\Omega)$.

This suggests that a "regularized" Casimir operator might be defined as

$$T_{\text{reg}} := \operatorname*{Lt}_{M \to \infty} M^{-s} T_M \tag{2.20}$$

and then (2.19) implies that, between the basic states $U(\alpha)\Omega$, T_{reg} is a multiple of the unit operator. It is an easy extension of the argument above to show that the sequence of operators in (2.20) does indeed converge strongly to a multiple of the unit operator on the dense subset of finite sums of vectors of the form $U(\alpha)\Omega$, $\alpha \in C(\Sigma, W^*)$. This in turn suggests that, in the special case when $W \approx \mathbf{R}^N$ and $G \approx SO(N)$ [so that $T(x)$ is the only Casimir to be expected] the representation of the affine field operators given in (2.2), (2.3) may actually be *irreducible* in this Fock space. Surprising as this may seem, it is in fact true.

In proving this result it is convenient to utilize the canonical coherent states defined as usual in termes of the annihilation operator $A(f) := (u(f) + i\pi(f))/\sqrt{2}$ [so that $[A(f), A^\dagger(g)] = (f,g)$] as

$$|f\rangle := \exp - \tfrac{1}{2}(f,f) \exp A^\dagger(f)\Omega. \tag{2.21}$$

Such vectors are normalized, $\langle f|f\rangle = 1$, and can be defined for all complex functions f in $L^2(\Sigma, W^*)$. However, it will be sufficient to confine our attention to the subset of coherent states for which $f \in C(\Sigma, W^*)$ and is real since finite linear combinations of such vectors are dense in the Hilbert space of interest. We shall make repeated use of the eigenrelations $A(f)|g\rangle = (f,g)|g\rangle$ appropriate for such states.

In order to demonstrate irreducibility it is sufficient to show that the operators $A^a(x)$ and $A^{a\dagger}(x)$ are suitable functions of $u^b(x)$ and $J_i(x)$. Before giving a precise statement it might be useful to first outline the argument formally. Thus $J_i(x) = -L_{iab} u^a(x) \pi^b(x) = (i/2) L_{iab} A^{a\dagger}(x) A^b(x)$, which is well defined between an arbitrary pair of coherent states $|f\rangle, |g\rangle$ as

$$\langle f | J_i(x) | g\rangle = i L_i{}^{ab} f_a(x) g_b(x). \tag{2.22}$$

Then

$$J_i(x) u^c(x) = i L_{iab} A^{a\dagger}(x) A^b(x) (A^c(x) + A^{c\dagger}(x))/\sqrt{2} \tag{2.23}$$

$$= (i/\sqrt{2})(L_{ia}{}^c A^{a\dagger}(x)\delta(0) + {:}J_i(x) u^c(x){:}), \tag{2.24}$$

where, as usual, : : denotes normal ordering. From (2.24), it follows that, between arbitrary coherent states,

$$L_{ia}{}^b \langle f | A^{a\dagger}(x) | g\rangle \delta(0) = -i\sqrt{2} \langle f | J_i(x) u^b(x) | g\rangle + \text{regular terms}. \tag{2.25}$$

However, for this present case of the rotation group SO (N) acting on $\mathbf{R}^N$, the matrix elements $L_{ia}{}^b$ obey the relation $C^{ij} L_{ia}{}^b L_{jb}{}^c = \nu \delta_a{}^c$ for some $\nu > 0$, where C^{ij} is the quadratic Casimir for the group SO(N). Thus $A^{a\dagger}(x)$ can be isolated in the form

$$\langle f | A^{a\dagger}(x) | g\rangle \delta(0) = (-i\sqrt{2}/\nu) C^{ij} L_{ib}{}^a \langle f | J_j(x) u^b(x) | g\rangle + \text{regular terms}, \tag{2.26}$$

which demonstrates the point provided we can "divide" by $\delta(0)$ in some appropriate way and hence make the "regular terms" irrelevant.

To make this kind of argument precise let us study the quantities

$$\mathfrak{U}_M := (-i\sqrt{2}/\nu) C^{ij} L_{ib}{}^a \sum_{\lambda_n < M^2} \varkappa_M J_j(f_a w_n) u^b(w_n), \tag{2.27}$$

where $\{\varkappa_M | M = 1,2,...\}$ is a sequence of positive numbers which suitably tend to zero and which will be determined shortly. For coherent states $|g\rangle$, it is straightforward to show that

$$\|(\mathbb{1}_M - A^\dagger(f))|g\rangle\|^2$$
$$= \int d\vartheta(x)\, f_a(x) f_a(x)\Big(1 - \sum_{\lambda_n < M^2} \varkappa_M(w_n(x))^2\Big)^2$$
$$+ \int\int d\vartheta(x) d\vartheta(y)\, f_a(x) g_a(x) f_b(y) g_b(y)$$
$$\times\Big(1 - \sum_{\lambda_n < M^2} \varkappa_M(w_n(x))^2\Big)$$
$$\times\Big(1 - \sum_{\lambda_n < M^2} \varkappa_M(w_n(y))^2\Big) + \text{regular terms}, \tag{2.28}$$

where "regular" refers to terms that would remain bounded were $\varkappa_M = 1$ for all M, i.e., when the limit as $M \to \infty$ of $\Sigma_{\lambda_n < M^2} \varkappa_M w_n(x) w_n(y)$ is the δ function $\delta(x,y)$. Now from (2.16) we see that if $\varkappa_M$ is defined to be

$$\varkappa_M := \Gamma(1 + s/2)(4\pi/M^2)^{s/2}, \tag{2.29}$$

then

$$\operatorname*{Lt}_{M\to\infty} \|(\mathbb{1}_M - A^\dagger(f))|g\rangle\|^2 = 0, \tag{2.30}$$

which implies the same result for arbitrary finite linear combinations of coherent states. But the set of all such combinations is dense in the Hilbert space, and this is the sense in which $A^{\dagger a}(x)$ is indeed a function of the operators $u^a(x)$ and $J_i(x)$. The relation $A^a(x) = \sqrt{2}u^a(x) - A^{a\dagger}(x)$ completes the isolation of the creation and annihilation operators and demonstrates the irreducibility of the affine algebra for this representation.

By virtue of the theorem quoted earlier, the result just proved also shows that the action of the group $C(\Sigma,\mathrm{SO}(N))$ on $C(\Sigma,\mathbb{R}^N)'$ is ergodic with respect to the measure μ. But note that this action is certainly *not* transitive since it is easy to write down a number of disjoint orbits. Thus, in demonstrating the irreducibility of the affine group representation, we have at the same time shown that this $C(\Sigma,\mathrm{SO}(N))$ action on $C(\Sigma,\mathbb{R}^N)'$ belongs to the interesting (but rather elusive) class of *strictly* ergodic actions. For more general groups G we would need to study the extra Casimir operators that are needed to select the G/H orbit embedded in W.

B. A class of reducible representations

If we wish to find a Fock representation in which $T(x)$ is a well-defined operator without regularization, it is clear that we must use a measure that is concentrated on less singular configurations than is the one defined by (2.7), (2.8). One natural way is to introduce a Sobolev-type structure and consider the Gaussian measure with covariance specified by the inner product $(\alpha,F\beta)$, where $F := (-\Delta + \rho)^{-\tau}$, τ and ρ are positive real numbers, and Δ is the Laplacian operator used above. Generally speaking, the larger is τ, the wider is the class of allowed test functions α, β, and, correspondingly, the less singular are the distributions contained in the support of the measure μ $(:= \mu_F)$.

One unavoidable property of this regulated inner product is that it is no longer invariant under the action of the group $C(\Sigma,G)$ on $C(\Sigma,W^*)$ and hence the measure μ will, at best, be quasi-invariant. Note also, that in order to maintain the canonical commutation relations associated with (2.8), as well as irreducibility thereof, it is necessary to define the conjugate variable $\pi(\beta)$ so that

$$(\Omega, e^{i\pi(\beta)}\Omega) = \exp - \tfrac{1}{4}(\beta, F^{-1}\beta) \tag{2.31}$$

which pairs with

$$(\Omega, e^{-iu(\alpha)}\Omega) = \exp - \tfrac{1}{4}(\alpha, F\alpha) \tag{2.32}$$

in the form

$$(\Omega, e^{-iu(\alpha)} e^{i\pi(\beta)}\Omega)$$
$$= \exp(i/2)(\alpha,\beta)\exp - (1/4)((\alpha,F\alpha) + (\beta,F^{-1}\beta)). \tag{2.33}$$

It is not easy to check directly the $C(\Sigma,G)$ quasi-invariance of this new Gaussian measure, and we will tackle the problem via a study of the existence of the operators which we would expect to be the generators of the group, were the measure to be quasi-invariant. Thus for test functions f^i, we consider

$$J_M(f) := -L_{iab} \sum_{\lambda_n < M^2} u^a(f^i w_n)\pi^b(w_n) \tag{2.34}$$

and seek to define $J(f)$ as $M \to \infty$. The commutation relations and (2.33) give

$$(\Omega, u^a(\alpha_1)\pi^b(\beta_1)u^c(\alpha_2)\pi^d(\beta_2)\Omega)$$
$$= -i\delta^{bc}(\alpha_2,\beta_1)(\Omega, u^a(\alpha_1)\pi^d(\beta_2)\Omega)$$
$$+ (\Omega, u^a(\alpha_1)u^c(\alpha_2)\pi^b(\beta_1)\pi^d(\beta_2)\Omega)$$
$$= \tfrac{1}{4}(\delta^{bc}\delta^{ad}(\alpha_1,\beta_2)(\alpha_2,\beta_1) - \delta^{ab}\delta^{cd}(\alpha_1,\beta_1)(\alpha_2,\beta_2)$$
$$+ \delta^{ac}\delta^{bd}(\alpha_1,F\alpha_2)(\beta_1,F^{-1}\beta_2)) \tag{2.35}$$

and then, writing $\mathrm{Tr}\, L_i L_j$ as D_{ij}, (2.34), (2.35) give

$$\| J_M(f)\Omega\|^2$$
$$= \frac{1}{4} D_{ij} \sum_{\lambda_n < M^2} \sum_{\lambda_m < M^2} ((f^i w_n, w_m)(f^j w_m, w_n)$$
$$- (f^i w_n, F f^j w_m)(w_n, F^{-1} w_m)). \tag{2.36}$$

To simplify the estimates, we will take as an example the special case where Σ is the flat s torus $\mathbf{R}^s/2\pi\mathbf{Z}^s$ with the Fourier transform relations

$$h(x) = \sum_n e^{in\cdot x} h(n),$$
$$h(n) = (2\pi)^{-s} \int_0^{2\pi} \cdots \int_0^{2\pi} e^{-in\cdot x} h(x)\, dx, \tag{2.37}$$

where n and x now denote s vectors and whose inner product is written as $n\cdot x$. Then

$$\|J_M(f)\Omega\|^2$$
$$= \frac{1}{4}(2\pi)^{2s} D_{ij} \sum_{\lambda_n < M^2} \sum_{\lambda_m < M^2} \Big[f^i(n-m) f^j(m-n)$$
$$\times\Big(1 - \Big(\frac{\lambda_m + \rho}{\lambda_n + \rho}\Big)^\tau\Big)\Big], \tag{2.38}$$

which, for large M, becomes

$$\| J_M(f)\Omega\|^2 = \frac{1}{4}(2\pi)^{2s}D_{ij}\sum_n f^i(n)f^{j*}(n) \times \sum_{\lambda_m<M^2}\left(\frac{\lambda^\tau_{m+n}-\lambda^\tau_m}{\lambda^\tau_{m+n}}\right). \tag{2.39}$$

Now, on a torus, $\lambda_m = m\cdot m$ and so, provided $\tau\neq 0$, the sum in (2.39) can be approximated by

$$\sum_{\lambda_m<M^2}\left(\frac{\lambda^\tau_{m+n}-\lambda^\tau_m}{\lambda^\tau_{m+n}}\right)\sim\int^{M^2}\lambda^{-1/2}\rho(\lambda)d\lambda\,, \tag{2.40}$$

where $\rho(\lambda)$ is the "density of eigenvalues," i.e., $\int^\Lambda\rho(\lambda)d\lambda$ is the number of eigenvalues whose value is less than Λ. However, Weyl's well-known theorem[19] shows that the latter depends asymptotically on Λ as $\Lambda^{s/2}$, and hence $\rho(\lambda)\sim\lambda^{s/2-1}$. Hence the right-hand side of (2.39) behaves like $M^{(s-1)}$ and it follows that one cannot define an operator $J(f)$ as $M\to\infty$ unless the dimension s of Σ is one. In this case, a more detailed calculation shows that for any coherent state $|g\rangle$, the matrix elements $\langle g|(J_M(f)-J_N(f))^2|g\rangle$ satisfy

$$\langle g|(J_M(f)-J_N(f))^2|g\rangle\leqslant c(M^{-1}+N^{-1}) \tag{2.41}$$

for some constant c. This bound shows that the sequence of operators $J_M(f)$ converges strongly on all coherent states; it is then a trivial matter to prove the strong convergence on the dense set of all finite linear combinations of such states, and in that sense, $J(f) := \mathrm{Lt}\, M\to\infty J_M(f)$ defines a densely defined operator. It is clear that this operator satisfies the correct commutation relations (1.11)–(1.13) and hence we have shown that the modified Gaussian measure defined by (2.32) is indeed quasi-invariant under the (connected component) of the group $C(\Sigma,G)$ provided that $\dim\Sigma = 1$.

For the example under consideration ($s=1$), we have thus obtained a representation of the *loop group* $C(S^1,W^*)\,\text{ⓢ}\,C(S^1,G)\approx C(S^1,W^*\text{ⓢ}G)$. Note that

$$(\Omega,u^a(x)u^b(y)\Omega)=\frac{\delta^{ab}}{8\pi}\sum_n(n\cdot n+\rho)^{-\tau}e^{in\cdot(x-y)}\,, \tag{2.42}$$

and hence if we choose a suitable value of τ, for example $\tau=1$, it follows that the left-hand side is actually *finite* when $x=y$. As a consequence, it follows (using coherent states, for example) that $T(x)=u^a(x)u^a(x)$ is a well-defined operator for each x (*without* the need for smearing). Since $T(x)$ commutes with $u^a(x)$ and $J_i(x)$, and is not a multiple of the identity, the reducibility of this representation is thereby established. Moreover, the spectrum of $T(x)$ is absolutely continuous so that the representation of the affine fields is a direct integral of inequivalent representations labelled by the spectral values of $T(x)$. Thus in this special case, we reproduce something resembling the decomposition theory of the analogous representation for the finite-dimensional Euclidean group.

III. ULTRALOCAL REPRESENTATIONS

In this section we present a class of inequivalent representations of the affine commutation relations which are unlike those of the preceding section in that *no* canonical momenta exist[20] for the fields in question. Representations with just this property suggest themselves as being especially relevant for affine algebras associated with nonlinear σ models since translations of the field variables associated with canonical momenta are manifestly incompatible with the putative constraint $u^a(x)u^a(x)=$ constant. The existence of conjugate variables in the form of "rotation" generators that preserve the constraint, as well as the absence of canonical generators that violate the constraint, thus seems particularly appealing.

The affine field representations discussed in this section are constructed in the fashion used for ultralocal models, and we shall refer to the existing literature[11,21,22] for most of the details of such representations, contenting ourselves with the essentials needed for our present purpose. We begin by considering the Hilbert space $\mathfrak{h}=L^2(\Sigma\times W,d\vartheta\otimes dv)$, where dv is some G-invariant measure on the finite-dimensional real vector space W in which the G/H orbit is embedded; for example, the Lebesgue measure on $\mathbb{R}^n$ is invariant under the action of SO(n). Now define the unitary operators

$$(\mathfrak{v}(\Lambda)\psi)(x,\lambda):=\psi(x,L(\Lambda(x)^{-1})\lambda)\,, \tag{3.1}$$

$$(\mathfrak{u}(\alpha)\psi)(x,\lambda):=e^{-i\langle\alpha(x),\lambda\rangle}\psi(x,\lambda)\,, \tag{3.2}$$

where $(x,\lambda)\in\Sigma\times W$ and $\langle\,,\rangle$ denotes the pairing of W^* and W. It is easy to check that (3.1), (3.2) gives a representation of the canonical group $\mathfrak{G}=C(\Sigma,W^*)\,\text{ⓢ}\,C(\Sigma,G)$. However, it is extremely reducible. For example, let A be a subset of Σ with $\vartheta(A)>0$ and let P_A denote the corresponding projection operator on $\mathfrak{h}$ defined by $(P_A\psi)(x,\lambda):=\psi(x,\lambda)$, if $x\in A$ and $(P_A\psi)(x,\lambda):=0$, otherwise. Then P_A commutes with the operators in (3.1), (3.2) which reflects the "independence" of the group operations on neighboring spatial points. Similarly, the operators in (3.1), (3.2) commute with the projection operator Q_B associated with any G-invariant subset B of W with $v(B)>0$.

A class of representations naturally associated with the above can be obtained by forming the finite tensor products $\overset{n}{\otimes}\mathfrak{h}$ of $\mathfrak{h}$ with itself. Direct sums of these representations can also be taken, leading in the limit to the various Fock spaces over $\mathfrak{h}$. For our purposes, the most relevant will be the bosonic Fock space $\mathfrak{H}$ formed from the direct sum of all the symmetrized tensor products Sym $\otimes\,\mathfrak{h}$. It will be convenient for us to employ the (equivalent) definition of $\mathfrak{h}$ as the *exponential of* $\mathfrak{h}$ (Refs. 11, 23, 24). This exploits "exponential" vectors of the form

$$\exp\psi:=1\oplus\psi\oplus(1/\sqrt{2!})\psi\otimes\psi\oplus(1/\sqrt{3!})\psi\otimes\psi\otimes\psi+\cdots, \tag{3.3}$$

whose inner products are defined by

$$\langle\exp\varphi,\exp\psi\rangle_{\mathfrak{H}}:=\exp\langle\varphi,\psi\rangle_{\mathfrak{h}}\,. \tag{3.4}$$

Finite linear combinations of such vectors are dense in the Fock space $\mathfrak{H}$ and are particularly useful vectors for defining certain types of operator. For example, the annihilation operator $A(\varphi)$ associated with $\varphi\in\mathfrak{h}\approx L^2(\Sigma\times W,d\vartheta\otimes dv)$ is defined on exponential vectors by

$$A(\varphi)\exp\psi=\langle\varphi,\psi\rangle_{\mathfrak{h}}\exp\psi, \tag{3.5}$$

in which context it is clear that the normalized "vacuum" state is $\exp\vec{0}$ where $\vec{0}$ denotes the null vector in $\mathfrak{h}$; for the sake of clarity this vacuum state will be denoted $|0\rangle$ hereafter. The unsmeared operators $A(x,\lambda)$ and $A^\dagger(x,\lambda)$ satisfy the usual commutation relations $[A(x,\lambda),A^\dagger(x',\lambda')]=\delta(x,x')\delta(\lambda,\lambda')$, so we do indeed have a conventional Fock representation.

If $\mathrm{u}(k)$ is a unitary representation of any group K on $\mathfrak{h}$, then an associated representation on $\mathfrak{H}=\exp\mathfrak{h}$ can be defined by its action on exponential vectors

$$\mathfrak{U}(k)\exp\psi:=\exp\mathrm{u}(k)\psi. \tag{3.6}$$

In particular, this applies to the representation of $C(\Sigma,W^*)\,\text{Ⓢ}\,C(\Sigma,G)$ defined by (3.1), (3.2), although, as is clear from the construction, this new representation is even more reducible than (3.1), (3.2) insofar as each subspace $\overset{n}{\mathrm{Sym}}\otimes\mathfrak{h}$ of $\mathfrak{H}$ is invariant under the action of the group. However the situation becomes potentially more interesting if we introduce *cocycles* for the action on $\mathfrak{h}$. A cocycle is a map β: $K\to\mathfrak{h}$ satisfying the condition

$$\beta(k_1k_2)=\beta(k_1)+\mathrm{u}(k_1)\beta(k_2)\quad\text{for all } k_1,k_2\in K. \tag{3.7}$$

Then the definition

$$\begin{aligned}\mathfrak{U}(k)\exp\psi:&\\ =\exp(&-(1/2)\langle\beta(k),\beta(k)\rangle_{\mathfrak{h}}-\langle\beta(k),\mathrm{u}(k)\psi\rangle_{\mathfrak{h}})\\ &\times\exp(\mathrm{u}(k)\psi+\beta(k))\end{aligned} \tag{3.8}$$

gives a projective representation of the group K:

$$\mathfrak{U}(k_1)\mathfrak{U}(k_2)=\exp-i\,\mathrm{Im}\langle\beta(k_1),\mathrm{u}(k_1)\beta(k_2)\rangle_{\mathfrak{h}}\mathfrak{U}(k_1k_2). \tag{3.9}$$

A special case arises when the cocycle is a *coboundary* of the form

$$\beta(k):=\mathrm{u}(k)c-c, \tag{3.10}$$

where c is a vector in $\mathfrak{h}$. In this case the multiplier in (3.9) can be removed by appending the phase factor $\exp-i\,\mathrm{Im}\langle\mathrm{u}(k)c,c\rangle_{\mathfrak{h}}$ to the transformation in (3.8) to give the following genuine representation of K:

$$\begin{aligned}\mathfrak{U}(k)\exp\psi&\\ :=\exp&\langle c,(\mathrm{u}(k)-1)(c+\psi)\rangle_{\mathfrak{h}}\\ &\times\exp(\mathrm{u}(k)(\psi+c)-c),\end{aligned} \tag{3.11}$$

whose vacuum expectation value is

$$\langle 0|\mathfrak{U}(k)|0\rangle=\exp\langle c,(\mathrm{u}(k)-1)c\rangle_{\mathfrak{h}}. \tag{3.12}$$

In particular, if w is any self-adjoint operator on $\mathfrak{h}$, an associated self-adjoint operator W on $\exp\mathfrak{h}$ is defined using Stone's theorem on the unitary operator

$$\begin{aligned}e^{it W}&\exp\psi\\ &:=\exp\langle c,(e^{it w}-1)(c+\psi)\rangle_{\mathfrak{h}}\exp(e^{it w}(\psi+c)-c),\end{aligned} \tag{3.13}$$

where $t\in\mathbb{R}$. The corresponding generating function is

$$\langle 0|e^{it W}|0\rangle=\exp\langle c,(e^{it w}-1)c\rangle_{\mathfrak{h}}. \tag{3.14}$$

At first nothing appears to be gained by this operation since the representation (3.11) of the group K is unitarily equivalent to that in (3.6) via the unitary map I from $\exp\mathfrak{h}$ to $\exp\mathfrak{h}$ defined by

$$I(\exp\psi):=\exp-(\tfrac{1}{2}\langle c,c\rangle_{\mathfrak{h}}-\langle c,\psi\rangle_{\mathfrak{h}})\exp(\psi+c). \tag{3.15}$$

However, one may consider using functions $c:\Sigma\times W\to\mathbb{C}$ which are *not* square integrable [i.e., c does not belong to $\mathfrak{h}\approx L^2(\Sigma\times W,d\vartheta\otimes dv)$] and yet for which (3.11), (3.12) are still well defined. The ensuing representation will typically *not* be equivalent to the simple (reducible) Fock representation (3.6), and in fact it is possible to construct *irreducible* representations in this way. It is these "c non-L^2" representations with which we will be mainly concerned for the remainder of the paper.

It is a standard result[11] that the above construction is equivalent to defining new operators

$$\begin{aligned}B(x,\lambda)&:=A(x,\lambda)+c(x,\lambda),\\ B^\dagger(x,\lambda)&:=A^\dagger(x,\lambda)+c^*(x,\lambda),\end{aligned} \tag{3.16}$$

which evidently satisfy the commutation relations $[B(x,\lambda),B^\dagger(x',\lambda')]=\delta(x,x')\delta(\lambda,\lambda')$. When applied to our affine group, the generators of the transformation (3.11) can be expressed (on an appropriate dense domain) as

$$u^a(x)=\int_W B^\dagger(x,\lambda)\lambda^a B(x,\lambda)dv(\lambda), \tag{3.17}$$

$$\begin{aligned}J_i(x)=(i/2)L_{iab}&\int_W B^\dagger(x,\lambda)\\ &\times(\lambda^a\partial^b-\lambda^b\partial^a)B(x,\lambda)dv(\lambda),\end{aligned} \tag{3.18}$$

in which λ^a are the components of the vector $\lambda\in W$ with respect to the basis $\{e_1,e_2,\ldots,e_N\}$, which is dual to the basis $\{e^1,e^2,\ldots,e^N\}$ of W^*. With respect to the G invariant metric on W dual to that on W^*, these basis vectors will satisfy $(e_a,e_b)=\delta_{ab}$ and the norm of λ is $\|\lambda\|=(\lambda^a\lambda^a)^{1/2}$. It is straightforward to verify that (3.17), (3.18) do indeed satisfy the affine algebra (1.11)–(1.13).

In order that the operators (3.17), (3.18) be well defined, it is sufficient[11] that a non-L^2 function c: $\Sigma\times W\to\mathbf{R}$ (it is no significant loss to take c as real valued) satisfies

$$\int c^2(x,\lambda)d\vartheta(x)dv(\lambda)=\infty, \tag{3.19}$$

$$\left|\int\lambda^a c^2(x,\lambda)d\vartheta(x)dv(\lambda)\right|<\infty, \tag{3.20}$$

$$\int c^2(x,\lambda)\|\lambda\|^2/(1+\|\lambda\|^2)d\vartheta(x)dv(\lambda)<\infty. \tag{3.21}$$

If c is a continuous function, and if the manifold Σ is compact, then divergences in the expressions above can only come from the integration over the noncompact space W. If Σ also is noncompact, a finite λ integral will produce a representation that is locally (spatially speaking) equivalent to that of (3.6). Hence it is the integration properties in λ that are of main interest for us. Note that the general spectral arguments applied previously to the Abelian subgroup $C(\Sigma,W^*)$ will be valid here too, and the representation above will be equivalent to one on $L^2(C(\Sigma,W^*)',d\mu)$ for some measure μ on the topological dual of the nuclear space $C(\Sigma,W^*)$. However, in the present case it is known[20] that μ and its translation by any nonzero element of $C(\Sigma,W^*)$ are mutually singular, and consequently, the fields in (3.17) do not possess any conjugate variables of the canonical kind. As remarked earlier, this gives added significance to the use of such representations of the affine commutation relations as the primary ingredient in a nonlinear quantum theory.

We see from (3.21) that a non-L^2 behavior for c can only arise from some singularity around $\lambda = 0$. Note that, if necessary, we can define (3.20) as a principal value integral and assume that c is such that the result is finite. In fact, in some earlier studies even the last integral has been taken divergent; however, such cases do not materially change things and are not discussed in this paper. To account for the singularity at $\lambda = 0$ consistent with (3.19)–(3.21) we could consider the family of functions

$$c(x,\lambda) = \kappa\|\lambda\|^{-\gamma}e^{-\rho(\lambda)}, \tag{3.22}$$

where $N/2<\gamma<1+N/2$ (recall that $\dim W=N$), κ is a constant, and ρ may be taken as a polynomial such that $\rho>0$. We will assume from now on that Σ is compact so that the effects of c are determined entirely by its integration properties in λ.

In order for a particular self-adjoint operator w on $\mathfrak{h}$ to truly generate an operator on $\mathfrak{H} = \exp\mathfrak{h}$, it is necessary and sufficient that [cf. (3.14)]

$$\int |(e^{itw}-1)c(x,\lambda)|^2 d\vartheta(x)\,d\nu(\lambda) \tag{3.23}$$

should be finite and continuous in $t\in\mathbb{R}$. We can now see at once why this "c-twisted" Fock representation is less reducible than the original one on $\mathfrak{h}$. First note that if P is any projection operator on a Hilbert space then the equation $P^2=P$ implies that

$$e^{itP}-1=(e^{it}-1)P. \tag{3.24}$$

When applied to the present case, this implies that the operator P_A on $\mathfrak{h}$ that projects onto any subset A of Σ with $\vartheta(A)>0$ does not produce a well-defined operator on $\exp\mathfrak{h}$ since the divergent λ integral in (3.19) will result in an infinite value for (3.23). Similarly, the projection operator Q_C onto a G-invariant subset C of W does not pass to an operator on $\exp\mathfrak{h}$ if C contains the point $\lambda = 0$. However, if this point is *not* in C then a genuine operator *will* be produced and this will commute with the generators of the representation of $C(\Sigma,W^*)\,\text{s}\,C(\Sigma,G)$. Thus even this c-twisted representation is not irreducible although, by this means, we have managed to remove the spatial decoupling of the original representation on $\mathfrak{h}$ and of the simple nontwisted Fock representation.

We are mainly interested in studying the properties of the Casimir operator in this new family of representations. Note that

$$u^a(x)u^a(y) = \delta(x,y)\int \|\lambda\|^2 B^\dagger(x,\lambda)B(x,\lambda)\,d\nu(\lambda) + \text{regular terms}, \tag{3.25}$$

where "regular terms" are less singular that the leading term as $x\to y$. The situation is similar to that arising from (2.12) and suggests that, as therein, a suitable regularization of the formal Casimir $u^a(x)u^a(x)$ would be to "divide" by $\delta(0)$. As in the previous case, we make this rigorous by using the eigenfunctions $w_1,w_2,\ldots$, of the Laplacian on Σ which generate $\delta(x,y)$ via (2.13). Once again, the Carleman asymptotic relation (2.16) shows that a well-defined operator can be constructed as

$$T_{\rm reg}(r) := \operatorname*{Lt}_{M\to\infty} M^{-s}\sum_{\lambda_n<M^2} u^a(rw_n)u^a(w_n). \tag{3.26}$$

This shows that (when suitably smeared),

$$T_{\rm reg}(x) := \int \|\lambda\|^2 B^\dagger(x,\lambda)B(x,\lambda)\,d\nu(\lambda) \tag{3.27}$$

is a well-defined local operator that commutes with the generators $u^a(x)$, $J_i(x)$ of the affine group. Since $T_{\rm reg}(x)$ is nonconstant (see below) it follows that the representation of this group is reducible.

Expressions like the vacuum generating functional in (3.12) follow from the general symbolic relation

$$\exp i(B^\dagger,wB) = :\!\exp(B^\dagger,(e^{iw}-1)B)\!: \tag{3.28}$$

where w is an operator in the Hilbert space $\mathfrak{h}$ and $:\;:$ denotes normal ordering of the $A^\dagger$ and A (hence $B^\dagger$ and B) operators. In particular, let $(w\psi)(x,\lambda) = \alpha_a(x)\lambda^a\psi(x,\lambda)$. Then (3.28) implies that [cf. (3.14)]

$$\langle 0|e^{iu(\alpha)}|0\rangle = \exp\int (e^{i\alpha_a(x)\lambda^a}-1)c^2(x,\lambda)\,d\vartheta(x)\,d\nu(\lambda), \tag{3.29}$$

while for $(w\psi)(x,\lambda) = r(x)\|\lambda\|^2\psi(x,\lambda)$,

$$\langle 0|\exp i\int r(x)T_{\rm reg}(x)\,d\vartheta(x)|0\rangle = \exp\int (e^{ir(x)\|\lambda\|^2}-1)c^2(x,\lambda)\,d\vartheta(x)\,d\nu(\lambda). \tag{3.30}$$

This expression shows clearly that $T_{\rm reg}$ is not simply a multiple of the unit operator, as otherwise its generating vacuum expectation value would have been given by $\exp i\int r(x)t(x)\,d\vartheta(x)$ for some $f(x)$, which is not the case. Moreover, like u, $T_{\rm reg}$ has an absolutely continuous spectrum since it follows from (3.30) that the function

$$C(t) := 0|\exp it\int r(x)T_{\rm reg}\,d\vartheta(x)|0\rangle$$

is an integrable function of $t\in\mathbb{R}$ (Ref. 25).

It might be thought that an irreducible set of field operators could arise by choosing the measure ν on W to be concentrated on one of the G orbits in order to remove the reducibility from the operators on $\exp \mathfrak{h}$ arising from the projection operators $Q_C, \lambda \neq 0$, on $\mathfrak{h}$; if ν is Lebesgue measure on W this would correspond formally to choosing a distributional form for c such that c^2 is a δ function concentrated on this orbit. With respect to the G-invariant metric on W, such an orbit would necessarily be a subspace of a sphere in W with some radius ρ; in particular, $\|\lambda\|$ would be equal to this constant ρ and would therefore drop out of the integral (3.21). But then the condition (3.21) implies that c is an L^2 function, and hence that the representation is equivalent to the simpler one with no cocycle at all. The question therefore is whether such a representation is interesting when the measure ν is concentrated on a G orbit. The motivation for such a step is that it removes part of the reducibility of the original representation on $\mathfrak{h}$ since there are no longer any nontrivial G-invariant subspaces of W other than the orbit itself. However, it does not lead to an irreducible representation of the affine fields on the Fock space $\exp \mathfrak{h}$. This is evident from the fact that

$$\langle 0|\exp i\int r(x)T_{\text{reg}}(x)d\vartheta(x)|\,0\rangle = \exp a\int (e^{i\rho r(x)}-1)d\vartheta(x) \tag{3.31}$$

for some constant a and, again, this is not of the form $\exp i\int r(x)t(x)d\vartheta(x)$. Note that, in the present case, the spectrum of the operator $\int r(x)T_{\text{reg}}(x)d\vartheta(x)$ with, for example, a smooth test function r of compact support, is not absolutely continuous. For if it were, the Riemann–Lebesgue lemma would ensure that $\langle 0|\exp it\int r(x)T_{\text{reg}}\,d\vartheta(x)|0\rangle$ vanished as $t\to\infty$, which is not the case.

Let us return now to the case of a non-L^2 cocycle function c and consider the implications of the detailed behavior near the singular point $\lambda=0$. There is a good reason for choosing the behavior of $c(x,\lambda)$ in (3.22) near $\lambda=0$ to be a function of the norm $\|\lambda\|$ only. In the case of the group $C(\Sigma,G)$, condition (3.23) becomes

$$\int |c(x,L(\Lambda(x))\lambda)-c(x,\lambda)|^2 d\vartheta(x)d\nu(\lambda)<\infty \quad \text{for all } \Lambda\in C(\Sigma,G), \tag{3.32}$$

whose "infinitesimal" form is precisely (3.20). Thus (for compact Σ) the existence of $J_i(x)$ as a local operator requires that the singular component of c should be G invariant.

There is in fact a subclass of such representations in which the function c is *fully* G invariant. In this case the representation of $C(\Sigma,G)$ on $\exp \mathfrak{h}$ simplifies to

$$V(\Lambda)\exp\psi=\exp\langle c,(\mathfrak{v}(\Lambda)-1)\psi\rangle_{\mathfrak{h}}\exp \mathfrak{v}(\Lambda)\psi \tag{3.33}$$

from which, in particular, it follows that $V(\Lambda)|0\rangle=|0\rangle$. Thus the vacuum $|\,0\rangle$ is annihilated by all the generators $J_i(x)$. Full G invariance of c also means that

$$\langle 0|u^a(x)|0\rangle=\int \lambda^a c^2(x,\|\lambda\|)d\nu(\lambda)=0. \tag{3.34}$$

In the general case, however, $c(x,\lambda)$ is not fully G invariant and consequently the rhs of (3.34) becomes some nonvanishing function $u^a_c(x)$. If Σ is just the flat space $\mathbf{R}^s$, and if we require the vacuum $|\,0\rangle$ to be translation invariant, then the function c must be chosen to be a function of λ only, in which case the rhs of (3.34) is a nonvanishing vector in W. Such cases are analogous to broken symmetry solutions in conventional quantum field theory. Unfortunately, ultralocal model theories do not themselves possess broken symmetry solutions as there can be no zero mass (Goldstone) excitations to account for such breaking.

An interesting property of the classical nonlinear σ model is the existence of topologically nontrivial configurations. These correspond classically to maps in $C(\Sigma,G/H)$ which are not deformable to the trivial map, and the topological sectors of the classical theory are labelled by the homotopy classes of these maps. It is therefore an interesting question to see if such topological effects also arise in the affine group representation theory. Thus the quantum field $u^a_{(c)}(x)$ corresponding to a specific choice of c could be said to belong to a particular homotopy class if

$$\langle 0|u^a_{(c)}(x)|0\rangle=\int \lambda^a c^2(x,\lambda)d\nu(\lambda)=\gamma^a_{(c)}(x), \tag{3.35}$$

where $\gamma_{(c)}$ is a map from Σ into W whose image lies in the G/H subspace of W and which, considered as a G/H valued map, belongs to the specified homotopy class.

To see that such quantum sectors can indeed be constructed let $c(\lambda)$ be a cocycle satisfying (3.19)–(3.21) and which is a function of λ only. Let $B\in C(\Sigma,G)$ and define a new cocycle by $c(x,\lambda):=c(L(B(x))\lambda)$. Then for the corresponding quantum field denoted $u^a_{(B)}(x)$, one finds

$$\langle 0|u^a_{(B)}(x)|0\rangle=\int \lambda^a c^2(L(B(x))\lambda)d\nu(\lambda) = L(B(x)^{-1})^a{}_b\gamma^b_{(c)}, \tag{3.36}$$

where $\gamma^b_{(c)}:=\int \lambda^b c^2(\lambda)d\nu(\lambda)$ is the vacuum expectation value associated with the original cocycle $c(\lambda)$. Now suppose $c(\lambda)$ is chosen such that the vector $\gamma_{(c)}$ lies on the particular orbit in W identified with G/H. For example, if $f(\|\lambda\|)$ is a suitably damped function of $\|\lambda\|$ then $\int f(\|\lambda\|)\lambda^a\lambda^b d\nu(\lambda)$ will be proportional to δ^{ab} and hence, for any vector $v\in W$, if

$$c(\lambda):=[\|\lambda\|f(\|\lambda\|)(1+\langle v,\lambda\rangle/2\|v\|\;\|\lambda\|)]^{1/2},$$

then $\int \lambda^a c^2(\lambda)d\nu(\lambda)$ defines the components of a vector pointing in the direction of v. With such a choice for c, it follows from (3.36) that $\gamma^a_{(B)}(x):=\langle 0|u^a_{(B)}(x)|0\rangle$ lies on this same orbit and so $\gamma_{(B)}$ does indeed define a map from Σ into the G/H subspace of W. Furthermore, $\gamma_{(B)}$ is obtained by acting with $L(B(x)^{-1})$ on the vector $\gamma_{(c)}$, which can be thought of as a constant function from Σ into G/H. Thus the topological sectors that can be reached in this way are precisely the set mentioned in the introduction, i.e., the homotopy classes of the maps $C_{(0)}(\Sigma,G/H)$ from Σ into G/H that can be lifted to G.

It is interesting to enquire when these different sectors are unitarily equivalent; i.e., does there exist a unitary operator $\mathfrak{J}$ on exp $\mathfrak{h}$ such that $\mathfrak{J}\, u^a_{(B)}(x)\, \mathfrak{J}^{-1} = u^a_{(c)}(x)$ where B and C are any pair of maps in $C(\Sigma,G)$? To answer this, note that (1.8) implies that

$$V(\Lambda)u^a(x)V(\Lambda)^{-1} = u^b(x)\widetilde{L}(\Lambda(x))_b{}^a. \qquad (3.37)$$

Thus, since it is always possible to obtain $B\in C(\Sigma,G)$ by multiplying $C\in C(\Sigma,G)$ with another group element in $C(\Sigma,G)$, a failure to construct such an operator would mean that the Hilbert space associated with the original cocycle c does not carry a represenation of the entire group but only the connected component of the identity [obtained by exponentiating the generators of the Lie algebra of $C(\Sigma,G)$] plus whatever disconnected transformations *can* be attached to a unitary operator $\mathfrak{J}$. A similar remark applies to the representations associated with any specific function B. However, the condition for a particular $\Lambda\in C(\Sigma,G)$ to be implementable in the representation with cocycle c is just (3.31), and similarly the necessary and sufficient condition for the B and C sectors to be unitarily equivalent is that

$$\int |c(L(B(x))\lambda) - c(L(C(x))\lambda)|^2 d\vartheta(x)d\nu(\lambda) < \infty. \qquad (3.38)$$

Since, as remarked above, the singular part of c is G invariant, it follows from (3.38) that if Σ is compact there will be no problem and all sectors are unitarily equivalent. The Hilbert space will then decompose into a direct sum of subspaces labelled by the topological sectors, each one of which carries a representation of the connected component of the canonical group. The generators of the disconnected transformations will then serve as intertwining operators between these sectors. This property is quite important in, for example, discussions of strings propagating on a torus.[26,27] Note that if Σ is noncompact then the fields $u^a_{(B)}(x)$ and $u^a_{(C)}(x)$ will be unitarily equivalent provided that B and C are asymptotically equivalent (as $x\to\infty$) in the sense of (3.38), whatever difference they may possess in between.

IV. REPRESENTATIONS AND DYNAMICS

A. The Hamiltonian for canonical fields

In the previous sections we have discussed various representations of the affine field commutation relations (1.11)–(1.13). In this final section we come to the important question of the relation of these representations to dynamics, i.e., the assignment to each such representation of a Hamiltonian generator of time translations. Our aim is to develop an analog in the affine case of Araki's well-known demonstration[28] of the way in which the choice of a particular representation of the canonical commutation relations in quantum field theory essentially specifies the Hamiltonian. Such an analysis is not totally straightforward since many of our representations of the affine relations are reducible. On the other hand, all our examples—reducible as well as irreducible—have in common the fact that the Hilbert space vector denoted by $|0\rangle$ (and which is typically chosen to define expectation functionals) is *cyclic* for the field operators $u(\alpha)$ alone. Thus, for each of the representations discussed above, vectors of the form $|\alpha\rangle := U(\alpha)|0\rangle = e^{-iu(\alpha)}|0\rangle$, for all test functions α, span the relevant Hilbert space.

It is pedagogically useful to start by reviewing briefly Araki's argument for an N vector of canonical fields in a form suitable for our own purposes. The expectation that any associated canonical Hamiltonian can be written heuristically as

$$H = \frac{1}{2}\int (\pi^a(x)\pi^a(x) + V(\varphi))d\vartheta(x) \qquad (4.1)$$

[where $V(\varphi)$ includes terms involving spatial derivatives of φ] is reflected more formally in the requirement that H should satisfy the basic identities

$$i[H,\varphi^a(x)] = \pi^a(x), \qquad (4.2)$$

$$i^2[[H,\varphi^a(x)],\varphi^b(y)] = i[\pi^a(x),\varphi^b(y)] = \delta^{ab}\delta(x,y), \qquad (4.3)$$

modulo domain technicalities that we will ignore. If we introduce vectors of the form $|f\rangle := e^{i\varphi(f)}|0\rangle$, then it follows that

$$\left(\frac{\delta}{\delta f_b(y)} + \frac{\delta}{\delta f'_b(y)}\right)\langle f|\pi^a(x)|f'\rangle = \langle f|i[\pi^a(x),\varphi^b(y)]|f'\rangle = \delta^{ab}\delta(x,y)E(f'-f), \qquad (4.4)$$

where the expectation functional $E(f'-f) := \langle f|f'\rangle = \langle 0|f'-f\rangle$ depends only on the difference $f'-f$. Consequently,

$$\langle f|\pi^a(x)|f'\rangle = \tfrac{1}{2}(f_a(x) + f'_a(x))E(f'-f) + F_a(f'-f;x) \qquad (4.5)$$

for some F_a. To fix F_a we investigate

$$F_a(2f;x) = \langle -f|\pi^a(x)|f\rangle = \langle 0|e^{i\varphi(f)}\pi^a(x)e^{i\varphi(f)}|0\rangle \qquad (4.6)$$

and observe that if the vector $|0\rangle$ is time reversal invariant, then $F_a\equiv 0$. This fact may readily be seen in the context of the Schrödinger representation of n-dimensional quantum mechanics where time-reversal invariance implies the reality of the wave function. In that case the quantum mechanical analog of F_a is given by

$$\int \psi(x)e^{ipx}\left(-i\frac{\partial}{\partial x^a}e^{ipx}\psi(x)\right)d^nx = \frac{1}{2}\int\left(-i\frac{\partial}{\partial x^a}\right)(e^{ipx}\psi(x))^2\, d^nx = 0, \qquad (4.7)$$

provided that ψ falls off sufficiently fast as $|x|\to\infty$.

Hamiltonians that lead to time reversal invariant eigenstates are even functions of the canonical momenta $\pi(x)$; in particular this is true of (4.1) which is consistent with (4.2), (4.3). Let us assume therefore that $|0\rangle$ is time reversal invariant so that $F_a = 0$ for the canonical case, i.e.,

$$\langle f|\pi^a(x)|f'\rangle = \tfrac{1}{2}(f_a(x) + f'_a(x))E(f'-f).$$

Then we observe that

$$\left(\frac{\delta}{\delta f_a(x)} + \frac{\delta}{\delta f'_a(x)}\right)\langle f|H|f'\rangle$$
$$= \langle f|i[H,\varphi^a(x)]|f'\rangle$$
$$= \tfrac{1}{2}(f_a(x) + f'_a(x))E(f'-f), \qquad (4.8)$$

from which we deduce that $\langle f|H|f'\rangle = \tfrac{1}{2}(ff')E(f'-f) + G(f'-f)$ for some G given by $G(f) = \langle -f|H|0\rangle$. If $|0\rangle$ is an eigenstate of H with eigenvalue Λ_0 then $G(f) = \Lambda_0\langle -f|0\rangle = \Lambda_0 E(f)$, and hence

$$\langle f|H|f'\rangle = (\tfrac{1}{2}(f,f') + \Lambda_0)E(f'-f). \qquad (4.9)$$

Under the stated assumptions, this is Araki's expression determining the matrix elements of the Hamiltonian as a symmetric operator on the dense set of all finite linear combinations of the vectors $|f\rangle$ for all test functions f. Since $H \geqslant \Lambda_0$ on this dense set, the domain of H may be uniquely extended to give a self-adjoint operator that preserves this lower bound (this is the familiar Friedrich's extension, e.g., Riesz and Nagy[29]). We draw the reader's attention to the fact that the discussion above did not require irreducibility of the canonical fields but only that the vectors $\{|f\rangle\}$ spanned the relevant Hilbert space.

B. The Hamiltonian for affine fields

We wish now to develop, as far as possible, an analogous argument for fields satisfying the affine commutation relations. We have argued that relations of this type are appropriate in the quantization of systems like the nonlinear σ model and, for convenience, we will restrict our attention to the SO(N) theory in which the fields take their values in the $(N-1)$ sphere S^{N-1}. The classical Hamiltonian for this model can be obtained by taking the free-field Hamiltonian

$$H = \frac{1}{2}\int(\partial_0 u^a(x)\partial_0 u^a(x) + \nabla_i u^a(x)\nabla_i u^a(x))d^s x \qquad (4.10)$$

and then imposing the constraint $\Sigma_{a=1}^N u^a(x)u^a(x) = \text{const}$. Note that ∇_i denotes the spatial derivation operator and, for simplicity, we have assumed that the spatial manifold Σ has a trivial metric (i.e., a torus T^s or the Euclidean space $\mathbb{R}^s$). The generators of the SO(N) transformation are given by (in the notation of Sec. II) $J_i(x) = -L_{iab}u^a(x)\partial_0 u^b(x)$ and hence, if C^{ij} is the quadratic Casimir for SO(N),

$$C^{ij}J_i(x)J_j(x) = k(u^a(x)u^a(x)\partial_0 u^b(x)\partial_0 u^b(x) - u^a(x)\partial_0 u^a(x)u^b(x)\partial_0 u^b(x)) \qquad (4.11)$$

for some constant k. If the constraint is imposed, (4.11) becomes proportional to the time derivative part of the constrained form of (4.10) leading to the well-known Sugawara form of the Hamiltonian of the nonlinear σ model.[30] This in turn suggests that in our case [where the constraint is "imposed" only insofar as $T(x) := u^a(x)u^a(x)$ is potentially a Casimir operator], a plausible formal model for the Hamiltonian is

$$H = \varkappa\int(T(x)^{-1}C^{ij}J_i(x)J_j(x) + W(u))d^s x, \qquad (4.12)$$

where $\varkappa$ is some constant and W is a local function of $u^a(x)$ and its spatial derivatives.

The starting point for our analysis is the remark that (4.12) motivates choosing as the affine analog of (4.3),

$$i^2[[H,u^a(x)],u^b(y)] = B^{ab}(x)\delta(x,y), \qquad (4.13)$$

where

$$[B^{ab}(x),u^c(x)] = 0. \qquad (4.14)$$

Note that, for a suitable choice of $\varkappa$, the formal expression for $B^{ab}(x)$ is

$$B^{ab}(x) = \delta^{ab} - u^a(x)u^b(x)/T(x), \qquad (4.15)$$

which has the projection property $B^{ab}(x)B^{bc}(x) = B^{ac}(x)$. A rigorous definition of (4.15) will involve some sort of local operator product renormalization. Each affine field operator representation implicity carries its own "rule book" for renormalization, which may not be easy to implement. Indeed the renormalized operator $B^{ab}(x)$ may no longer possess the projection property, although it will generally retain its nonnegative definite character, a fact that will be useful later. An example is afforded by the ultralocal representation of Sec. III in which the renormalized version of (4.15) [cf. (3.27)]

$$B^{ab}(x) = \delta^{ab} - \frac{\int B^\dagger(x,\lambda)\lambda^a\lambda^b B(x,\lambda)d\nu(\lambda)}{\int B^\dagger(x,\lambda)\lambda^c\lambda^c B(x,\lambda)d\nu(\lambda)}. \qquad (4.16)$$

This symbolic expression for the resulting quadratic form is no longer projection-valued.

In any event, we shall assume that the form $B^{ab}(x)$ is known and satisfies (4.14). It then follows from (4.13), (4.14) that

$$-\left(\frac{\delta}{\delta\alpha_b(y)} + \frac{\delta}{\delta\alpha'_b(y)}\right)\langle\alpha|i[H,u^a(x)]|\alpha'\rangle$$
$$= \delta(x,y)\langle\alpha|B^{ab}(x)|\alpha'\rangle$$
$$= \delta(x,y)X^{ab}(\alpha'-\alpha;x). \qquad (4.17)$$

Note that the functional X^{ab} is determined, at least in principle, by the expectation functional $\tilde{E}(\alpha) := \langle 0|\alpha\rangle = \langle 0|e^{-iu(\alpha)}|0\rangle$ for the fields; explicit expressions for specific examples are given in Secs. II and III. Even though, in practice, X^{ab} may be difficult to determine from $\tilde{E}$ we shall regard it as known in what follows.

From (4.17) we deduce that

$$\langle\alpha|i[H,u^a(x)]|\alpha'\rangle$$
$$= -\tfrac{1}{2}(\alpha_b(x) + \alpha'_b(x))X^{ab}(\alpha'-\alpha;x) + Y^a(\alpha'-\alpha;x). \qquad (4.18)$$

The analog of this relation with (4.5) in the canonical case should be clear. In particular, we shall invoke an appropriate analog of the time reversal argument to set $Y^a = 0$. Hence, we are led to the relation

$$\left(\frac{\delta}{\delta\alpha_a(x)}+\frac{\delta}{\delta\alpha'_a(x)}\right)\langle\alpha|H|\alpha'\rangle$$
$$=\tfrac{1}{2}(\alpha_b(x)+\alpha'_b(x))X^{ab}(\alpha'-\alpha;x), \tag{4.19}$$

which implies that

$$\langle\alpha|H|\alpha'\rangle=\frac{1}{2}\int\alpha_a(x)\alpha'_b(x)X^{ab}(\alpha'-\alpha;x)d^sx$$
$$+Z(\alpha'-\alpha) \tag{4.20}$$

for some Z defined by $Z(\alpha):=\langle-\alpha|H|0\rangle$. Thus (4.20) determines the matrix of the Hamiltonian on the dense set of vectors made up of finite sums of states $|\alpha\rangle$.

This is our most general result for the affine analog of Araki's analysis of the canonical theory. However, as in the canonical case, any further information regarding $|0\rangle$ can be used to give a further characterization of the matrix elements of the Hamiltonian. For example, if $|0\rangle$ is invariant under spatial translations in $\Sigma(\approx\mathbb{R}^s$ or $T^s)$ then, for all vectors $a\in\Sigma$, $\widetilde{E}(\alpha^{[a]})=\widetilde{E}(\alpha)$ where $\alpha^{[a]}(x):=\alpha(x+a)$. It follows that $X^{ab}(\alpha'-\alpha;x)$ is independent of x and hence the first term of (4.20) becomes $\frac{1}{2}X^{ab}(\alpha'-\alpha)$ $\int\alpha_a(x)\alpha'_b(x)d^sx$. Another and independent, example, is if $|0\rangle$ is chosen as an eigenstate of H with eigenvalue Λ_0. Then $Z(\alpha)=\Lambda_0\langle-\alpha|0\rangle=\Lambda_0\widetilde{E}(\alpha)$, which thus becomes a known function. In particular,

$$\langle\alpha|H|\alpha'\rangle=\frac{1}{2}\int\alpha_a(x)\alpha'_b(x)X^{ab}(\alpha'-\alpha;x)d^sx$$
$$+\Lambda_0\widetilde{E}(\alpha'-\alpha). \tag{4.21}$$

By the nature of its construction, $X^{ab}(\alpha'-\alpha;x)$ is a nonnegative-definite functional, and it follows that $H\geqslant\Lambda_0$ on the dense set of vectors introduced above. Thus the Friedrich extension procedure quoted earlier uniquely defines a self-adjoint Hamiltonian having the same lower bound.

Let us conclude by emphasizing that, as in the canonical theory, our results for the Hamiltonian in the affine case apply to a very large class of representations, both reducible and irreducible, and characterize the Hamiltonian whenever states of the form $|\alpha\rangle$ span the Hilbert space, i.e., when $|0\rangle$ is a cyclic vector for the field operators alone. (Even if $|0\rangle$ is *not* a cyclic vector for the field operators, the expression derived above for the matrix elements of the Hamiltonian still holds in the *sub*space spanned by such vectors.) Moreover, in the case of fields where uncountably many inequivalent representations exist, it is entirely natural to link the Hamiltonian to the representation space by adopting the eigenvalue relation $H|0\rangle=\Lambda_0|0\rangle$ for some Λ_0. As we have seen, this choice reduces the Hamiltonian matrix elements to expressions determined in principle from the expectation functional for the fields.

From the perspective of the affine field algebra representations presented in Secs. II and III, perhaps the most interesting application of the relation between Hamiltonians and operator representations is to the case of the topologically nontrivial and nontranslationally invariant case $|0\rangle$ associated with x-dependent c functions in the ultralocal representations [cf. (3.29)]. Applications of the general discussion for Hamiltonians to explicitly covariant models must await further advances in the development of affine field representations themselves.

ACKNOWLEDGMENTS

J. R. K. thanks Imperial College for hospitality.

During a period when the bulk of this paper was prepared, he received financial support through SERC Grant GR/E/74144.

[1]G. Vilkovisky, in *Quantum Gravity—A Colleciton of Essays in Honour of Bryce DeWitt's Sixtieth Birthday*, edited by S. Christensen (Adam Hilger, Bristol, 1984).
[2]C. J. Isham, in *Relativity, Groups and Topology II. Proceedings of the 1983 Les Houches Summer School*, edited by B. DeWitt and R. Stora (North-Holland, Amsterdam, 1984).
[3]R. Palais, J. Math. Mech. **6**, 673 (1957).
[4]G. D. Mostow, Ann. Math. **65**, 432 (1957).
[5]R. Percacci, *Geometry of Nonlinear Field Theories* (World Scientific, Singapore, 1986).
[6]J. R. Klauder, in *Relativity*, edited by L. Witten (Plenum, New York, 1970).
[7]M. Pilati, Phys. Rev. D **26**, 2645 (1982).
[8]C. J. Isham, and A. C. Kakas, Class. Quant. Grav. **1**, 621 (1984).
[9]C. J. Isham and A. C. Kakas, Class. Quant. Grav. **1**, 633 (1984).
[10]J. R. Klauder, J. Math. Phys. **6**, 1666 (1965).
[11]J. R. Klauder, in *Mathematical Methods in Theoretical Physics*, edited by W. B. Britten (Colorado Assoc. U.P., Boulder, CO, 1973).
[12]R. Streater, Rend. Sci. Int. Fis. E. Fermi **11**, 247 (1969).
[13]H. Araki, Publ. RIMS, Kyoto Univ. **5**, 361 (1969).
[14]I. M. Gel'fand, M. J. Graev, and A. M. Veshik, Russian Math. Surveys **30**, 1 (1975).
[15]I. M. Gel'fand, M. J. Graev, and A. M. Vershik, Composito Math. **35**, 299 (1977).
[16]I. M. Gel'fand, M. J. Graev, and A. M. Vershik, Composito Math. **42**, 217 (1981).
[17]G. Mackey, *Unitary Group Representations in Physics, Probability, and Number Theory* (Benjamin–Cummings, London, 1978).
[18]I. M. Gel'fand and N. Y. Vilenkin, *Generalized Functions IV* (Academic, New York, 1964).
[19]N. E. Hurt, *Geometric Quantization in Action* (Reidel, London, 1983).
[20]G. C. Hegerfeldt and J. R. Klauder, Nuovo Cimento **10A**, 723 (1972).
[21]J. R. Klauder, Commun. Math. Phys. **18**, 307 (1970).
[22]J. R. Klauder, Acta Phys. Austriaca (Suppl.) VIII, 227 (1971).
[23]A. Guichardet, *Symmetric Hilbert Spaces and Related Topics*, Lecture Notes in Mathematics, Vol. 261 (Springer-Verlag, New York, 1972).
[24]K. R. Parthasarathy and K. Schmidt, *Positive Definite Kernels, Continuous Tensor Products, and Central Limit Theorems of Probability Theory*, Lecture Notes in Mathematics, Vol. 272 (Springer-Verlag, New York, 1972).
[25]E. Lukacs, *Characteristic Functions* (Hafner, New York, 1970), p. 33.
[26]Z. F. Ezawa, S. Nakakmura, and A. Tezuka, Tokyo University preprint TU/86/298 (1986).
[27]C. J. Isham and N. Linden, Class. Quant. Grav. **5**, 71 (1988).
[28]H. Araki, J. Math. Phys. **1**, 492 (1960).
[29]F. Riesz and B. Sz. Nagy, *Functional Analysis* (Ungar, New York, 1955).
[30]H. Sugarawa, Phys. Rev. **170**, 1659 (1968).

On the compatibility of relativistic wave equations in Riemann–Cartan spaces. II

H. A. Buchdahl
Department of Physics and Theoretical Physics, Faculty of Science, Australian National University, GPO Box 4, Canberra 2601, Australia

(Received 26 July 1989; accepted for publication 8 November 1989)

In a Riemann–Cartan space-time U_4 the minimally coupled massive spin-S Dirac equations are subject to constraints for all $S > \frac{1}{2}$. One may seek to modify these equations by the inclusion of appropriate supplementary terms $\widehat{\alpha}, \widehat{\beta}$ (indices suppressed) in such a way that the modified equations are unconstrained. This is done here explicitly in the case where the field spinors are $\xi^{\nu_1 \cdots \nu_n}$ and $\eta^{\dot{\mu}\nu_1 \cdots \nu_{n-1}}$, with $n = 2S$. The spin-1 tensor equations, i.e., the minimally coupled Proca equations, are transcribed into spinorial form. The explicit expressions for $\widehat{\alpha}$ and $\widehat{\beta}$ that so appear are in harmony with those already obtained. The freedom to choose $\widehat{\alpha}$ to be a zero spinor is shown to be circumscribed. Finally, it is pointed out that whether the unconstrained (spin-1) equations are minimally coupled or not depends on whether one chooses to write them as tensor or spinor equations.

I. INTRODUCTION

In an earlier paper,[1] after first clarifying what is to be understood by the "incompatibility" of relativistic wave equations, I established the following result: the minimally coupled massive spin-S ($S > \frac{1}{2}$) first-order Dirac equations

$$\nabla^{\dot{\mu}_{s+1}}{}_{\nu_t} \xi^{\dot{M}_s N_t} = \kappa \eta^{\dot{M}_{s+1} N_{t-1}}, \qquad (1.1a)$$

$$\nabla_{\dot{\mu}_{s+1}}{}^{\nu_t} \eta^{\dot{M}_{s+1} N_{t-1}} = \kappa \xi^{\dot{M}_s N_t} \qquad (1.1b)$$

are incompatible in any nontrivial Riemann–Cartan space-time U_4. The equations of constraint (or simply "'the constraints") (4.3) of Ref. 1 formally represent this incompatibility, for in an arbitrary U_4 they limit the freedom of choice of initial data; see Secs. V B and VI of Ref. 1. In the corresponding Riemannian case[2] one may alternatively take the constraints to impose generic limitations upon the V_4 rather than upon the choice of initial data, but to do so in the present context would be an empty exercise since the constraints then require the absence of torsion.

Equations (1.1) have their origin in the corresponding equations in a space S_4, free of curvature and torsion, from which they arise by the prescription of minimal coupling, i.e., by a modification that consists solely in the replacement of partial by covariant derivatives. Therefore, neither the curvature tensor $R^{n}{}_{klm}$ nor the torsion tensor $S_{kl}{}^{m}$ will appear explicitly in Eqs. (1.1). While this prescription has the charm of simplicity, it is flawed because in general it does not lead to unconstrained equations and one would appear to have no option but to abandon it. Following the Riemannian case,[3] it suggests itself to modify Eqs. (1.1) by the addition of spinors $\widehat{\alpha}^{\dot{M}_{s+1} N_{t-1}}$ and $\widehat{\beta}^{\dot{M}_s N_t}$, both symmetric in $\dot{M}_s$ and N_{t-1}, to the rhs of (1.1a) and (1.1b), respectively. Where these spinors are so chosen that the modified equations are unconstrained, while both spinors must vanish identically in the absence of curvature and torsion.

The problem of finding explicit expressions for[4] $\widehat{\alpha}$ and $\widehat{\beta}$ for general values of s and t is formidable and its solution is extant: I shall therefore confine myself here to the simplest case, i.e., $s = 0$, $t = 2S (= : n)$ since this is tractable, yet has some interesting features absent from the corresponding work contained in Ref. 2. In Sec. II, following the precedent set in Sec. V of Ref. 2, $\widehat{\alpha}$ is taken to be a zero spinor and an explicit expression for $\widehat{\beta}$ is then found. The remainder of the paper is devoted to a detailed examination of the special case of spin-1. In Sec. III the transcription of the minimally coupled Proca equations into spinor form directly provides particular expressions for $\widehat{\alpha}^{\dot{\mu}\rho}$ and $\widehat{\beta}^{\nu\rho}$, with $\widehat{\alpha}^{\dot{\mu}\rho} \neq 0$: Not unexpectedly, they are fully in harmony with the results of Sec. II. A possible pitfall in nonchalantly choosing $\widehat{\alpha}^{\dot{\mu}\rho}$ to be a zero spinor is examined in Sec. IV. Finally, Sec. V concerns a peculiar ambiguity inherent in the minimal coupling prescription, namely, whether the (first order) equations are minimally coupled or not depends on whether one chooses to write them as tensor or spinor equations.

II. $t=n$: UNCONSTRAINED EQUATIONS

When $s = 0$ and $t = n$ only one dotted index occurs in Eqs. (1.1), so that one may simply write $\dot{\mu}$ in place of $\dot{\mu}_1$. Then the equations to be dealt with are

$$\nabla^{\dot{\mu}}{}_{\nu_n} \xi^{N_n} = \kappa \eta^{\dot{\mu} N_{n-1}} + \widehat{\alpha}^{\dot{\mu} N_{n-1}}, \qquad (2.1a)$$

$$\nabla_{\dot{\mu}}{}^{\nu_n} \eta^{\dot{\mu} N_{n-1}} = \kappa \xi^{N_n} + \widehat{\beta}^{N_n}. \qquad (2.1b)$$

Alternatively, eliminating ξ and η between Eqs. (2.1) one has

$$(\tfrac{1}{2}\Box + \kappa^2) \xi^{N_n} = S^{kl\nu_n}{}_{\lambda} \xi^{N_{n-1}\lambda}{}_{;[kl]} - \nabla_{\dot{\mu}}{}^{\nu_n} \widehat{\alpha}^{\dot{\mu} N_{n-1}} - \kappa \widehat{\beta}^{N_n}, \qquad (2.2a)$$

$$(\tfrac{1}{2}\Box + \kappa^2) \eta^{\dot{\mu} N_{n-1}} = S^{kl\dot{\mu}}{}_{\dot{\lambda}} \eta^{\dot{\lambda} N_{n-1}}{}_{;[kl]} - \kappa \widehat{\alpha}^{\dot{\mu} N_{n-1}} - \nabla^{\dot{\lambda}}{}_{\nu_n} \widehat{\beta}^{N_n}. \qquad (2.2b)$$

From (2.2) there arises only a single condition upon $\widehat{\alpha}$ and $\widehat{\beta}$, namely, by transvection of (2.2a) with $\gamma_{\nu_{n-1}\nu_n}$,

$$\nabla_{\dot{\mu}\nu_{n-1}} \widehat{\alpha}^{\dot{\mu} N_{n-1}} - \kappa \gamma_{\alpha\beta} \widehat{\beta}^{N_{n-2}\alpha\beta} = S^{kl}{}_{\alpha\beta} \xi^{N_{n-2}\alpha\beta}{}_{;[kl]}. \qquad (2.3)$$

If one sets $\widehat{\alpha} = 0, \widehat{\beta} = 0$, one is left with the single equation of

 0022-2488/90/030712-03$03.00

constraint (4.3a) of Ref. 1, while the constraint (4.3b) of Ref. 1 is nugatory.] By inspection, if $\widehat{\alpha}$, $\widehat{\beta}$ satisfy condition (2.3), so do $\widehat{\alpha}+\widetilde{\alpha}$, $\widehat{\beta}+\widetilde{\beta}$, provided only that

$$\kappa\gamma_{\alpha\beta}\widetilde{\beta}^{N_{n-2}\alpha\beta}=\nabla_{\dot{\mu}\nu_{n-1}}\widetilde{\alpha}^{\dot{\mu}N_{n-1}}, \tag{2.4}$$

where $\widetilde{\alpha}$ can be chosen arbitrarily. In the first instance it therefore suffices to make any convenient ansatz for $\widehat{\alpha}$ and $\widehat{\beta}$. The simple choice

$$\widehat{\alpha}^{\dot{\mu}N_{n-1}}=0, \tag{2.5a}$$

$$\widehat{\beta}^{N_{n-2}\alpha\beta}=[(n-1)/n]\beta^{(N_{n-2}\gamma\alpha)\beta} \tag{2.5b}$$

suggests itself, where β is a symmetric spinor. [The insertion of the numerical factor $n/(n-1)$ ensures that $\beta^{N_{n-2}}=\gamma_{\alpha\beta}\widehat{\beta}^{N_{n-2}\alpha\beta}$.] Then (2.3) gives

$$\kappa\beta^{N_{n-2}}=-S^{kl}{}_{\nu_{n-1}\nu_n}\xi^{N_n}{}_{;[kl]}$$

$$\equiv\widehat{D}_{\nu_{n-1}\nu_n}\xi^{N_n}; \tag{2.6}$$

cf. (4.3a) of Ref. 1. Bearing (2.9) of Ref. 1 in mind, (2.6) becomes

$$\kappa\beta^{N_{n-2}}=\tfrac{1}{2}(n+1)S^{kl}{}_{\nu_{n-1}\nu_n}P^{(\lambda}{}_{\lambda kl}\xi^{N_n)}$$

$$+S^{m}{}_{\nu_{n-1}\nu_n}\nabla_m\xi^{N_n}. \tag{2.7}$$

The expression on the rhs of (2.7) may be simplified along the lines pursued in Sec. V of Ref. 2, but it is not necessary to spell this out in detail. At any rate, granted (2.5) and (2.7), Eqs. (2.1) are unconstrained.

Of course, one still cannot be sure that one is on the right track with the somewhat *ad hoc* device for "removing constraints" adopted above. In fact, it does gain some independent support from the results of a detailed investigation of the spin-1 equations. However, at the same time this throws up a possible pitfall in the ansatz (2.5a) which is peculiar to the Riemann–Cartan space-time, i.e., there does not exist a corresponding difficulty in a Riemannian space-time.

III. THE CASE OF SPIN-1

When $S=1$, i.e., $s=0$, $t=2$, Eqs. (1.1) are

$$\nabla^{\dot{\mu}}{}_{\rho}\xi^{\nu\rho}=\kappa n^{\dot{\mu}\nu}, \tag{3.1a}$$

$$\nabla_{\dot{\mu}}{}^{\rho}\eta^{\dot{\mu}\nu}=\kappa\xi^{\nu\rho}. \tag{3.1b}$$

In a Riemann–Cartan space-time U_4 (3.1) are constrained (see Sec. V A of Ref. 1), whereas in a Riemannian space-time they are not. Now, instead of pursuing the *ad hoc* addition of supplementary terms to Eqs. (3.1) one may argue as follows. When space-time is Minkowskian one may take as the starting point the usual flat space Proca equations[5]

$$\mu f_{kl}=\phi_{l;k}-\phi_{k;l}, \tag{3.2a}$$

$$f^{kl}{}_{;l}=\mu\phi^k, \tag{3.2b}$$

where subscripts following a semicolon temporarily denote partial derivatives. The corresponding equations in a U_4 are to be obtained from Eqs. (3.2) through the minimal coupling prescription: No formal change occurs in them, but subscripts following a semicolon now denote covariant derivatives.

Equations (3.2) are manifestly unconstrained; consequently, one can obtain from them unconstrained spinor equations by mere transcription. To begin with, any bivector f_{kl} has as its spinor equivalent a pair of symmetric spinors $\xi^{\mu\nu}$ and $\zeta^{\dot{\mu}\dot{\nu}}$ such that

$$f_{kl}=\tfrac{1}{2}(S_{kl\mu\nu}\xi^{\mu\nu}+S_{kl\dot{\mu}\dot{\nu}}\zeta^{\dot{\mu}\dot{\nu}}). \tag{3.3}$$

Also, because of (2.8) in Ref. 6, the dual of f_{kl} is

$$f^{\dagger kl}:=-\tfrac{1}{2}ie^{klmn}f_{mn}=\tfrac{1}{2}(S_{kl\mu\nu}\xi^{\mu\nu}-S_{kl\dot{\mu}\dot{\nu}}\zeta^{\dot{\mu}\dot{\nu}}). \tag{3.4}$$

Therefore,

$$\xi^{\mu\nu}=S^{kl\mu\nu}f_{kl},\quad \zeta^{\dot{\mu}\dot{\nu}}=S^{kl\dot{\mu}\dot{\nu}}f_{kl}. \tag{3.5}$$

If

$$F_{\pm}{}^{kl}:=f^{kl}\pm f^{\dagger kl}, \tag{3.6}$$

$F_{+}{}^{kl}$ and $F_{-}{}^{kl}$ are self-dual and antiself-dual, respectively.

Again, the spinor equivalent of ϕ_k is a mixed spinor $\eta^{\dot{\mu}\nu}$ such that

$$\phi_k=(1/\sqrt{2})\sigma_{k\dot{\mu}\nu}\eta^{\dot{\mu}\nu}. \tag{3.7}$$

Now, because of (3.5) and (3.7),

$$\nabla^{\dot{\mu}}{}_{\rho}\xi^{\nu\rho}=\sigma^{m\dot{\mu}}{}_{\rho}S^{kl\nu\rho}f_{kl;m}=-\sigma_k{}^{\dot{\mu}\nu}F_{+}{}^{kl}{}_{;l}, \tag{3.8}$$

where (2.17) of Ref. 6 has been used. Setting $\mu=:-\kappa\sqrt{2}$, (3.8) finally becomes

$$\nabla^{\dot{\mu}}{}_{\rho}\xi^{\nu\rho}=\kappa\eta^{\dot{\mu}\nu}-\sigma_k{}^{\dot{\mu}\nu}f^{\dagger kl}{}_{;l}. \tag{3.9a}$$

Next, by transvection of (3.2a) throughout with $S^{kl\mu\nu}$ one obtains

$$\kappa\xi^{\mu\nu}=-S^{kl\mu\nu}\sigma_{l\dot{\alpha}\beta}\eta^{\dot{\alpha}\beta}{}_{;k},$$

bearing (3.5) and (3.7) in mind. Using (2.16) of Ref. 6, this becomes

$$\nabla_{\dot{\lambda}}{}^{(\mu}\eta^{\dot{\lambda}\nu)}=\kappa\xi^{\mu\nu}.$$

Adding $\nabla_{\dot{\lambda}}{}^{[\nu}\eta^{\dot{\lambda}\mu]}$ to both sides of this equation,

$$\nabla_{\dot{\lambda}}{}^{\nu}\eta^{\dot{\lambda}\mu}=\kappa\xi^{\mu\nu}-\tfrac{1}{2}\gamma^{\mu\nu}\nabla_{\dot{\alpha}\beta}\eta^{\dot{\alpha}\beta}$$

in view of (4.5) of Ref. 6. Equivalently, because of (3.2b) and (3.7),

$$\nabla_{\dot{\lambda}}{}^{\nu}\eta^{\dot{\lambda}\mu}=\kappa\xi^{\mu\nu}+(1/2\kappa)\gamma^{\mu\nu}f^{kl}{}_{;lk}. \tag{3.9b}$$

Now, when $n=2$ the generic, unconstrained equations (2.1) are

$$\nabla^{\dot{\mu}}{}_{\nu}\xi^{\mu\rho}=\kappa\eta^{\dot{\mu}\rho}+\hat{\alpha}^{\dot{\mu}\rho}, \tag{3.10a}$$

$$\nabla_{\dot{\mu}}{}^{\nu}\eta^{\dot{\mu}\rho}=\kappa\xi^{\nu\rho}+\hat{\beta}^{\nu\rho}; \tag{3.10b}$$

Eqs. (3.10) remain unconstrained when $\tilde{\alpha}^{\dot{\mu}\rho}$, $\tilde{\beta}^{\mu\rho}$ are added to $\hat{\alpha}^{\dot{\mu}\rho}$, $\hat{\beta}^{\nu\rho}$, respectively, provided only that condition (2.4), i.e.,

$$\kappa\gamma_{\nu\rho}\tilde{\beta}^{\nu\rho}=\nabla_{\dot{\mu}\rho}\tilde{\alpha}^{\dot{\mu}\rho}, \tag{3.11}$$

is satisfied. Specifically, according to Eqs. (3.9),

$$\hat{\alpha}^{\dot{\mu}\rho}=-\sigma_k{}^{\dot{\mu}\rho}f^{\dagger kl}{}_{;l}+\tilde{\alpha}^{\dot{\mu}\rho}, \tag{3.12a}$$

$$\hat{\beta}^{\nu\rho}=(2\kappa)^{-1}\gamma^{\nu\rho}f^{kl}{}_{;lk}+\tilde{\beta}^{\nu\rho}; \tag{3.12b}$$

thus the particular choice

$$\tilde{\alpha}^{\dot{\mu}\rho}=\sigma_k{}^{\dot{\mu}\rho}f^{\dagger kl}{}_{;l} \tag{3.13}$$

reduces $\hat{\alpha}^{\dot{\mu}\rho}$ to zero. In regards to $\tilde{\beta}^{\nu\rho}$, it suffices to set

$$\tilde{\beta}^{\nu\rho}=:\tfrac{1}{2}\gamma^{\nu\rho}\tilde{\beta}, \tag{3.14}$$

where the scalar $\tilde{\beta}$ is then determined by (3.11), i.e.,

$$\kappa\tilde{\beta} = \nabla_{\dot{\mu}\rho}\tilde{\alpha}^{\dot{\mu}\rho} = f^{\dagger kl}{}_{;l}.$$

In effect, one thus has altogether

$$\hat{\alpha}^{\dot{\mu}\rho} = 0, \quad \hat{\beta}^{\mu\rho} = (2\kappa)^{-1}\gamma^{\nu\rho}F_{+}{}^{kl}{}_{;lk}, \tag{3.15}$$

It is easy to convince oneself by inspection that, with $n = 2$, Eqs. (2.5), together with Eq. (2.6), are exactly reproduced by Eqs. (3.15).

IV. THE ASSUMPTION $\hat{\alpha}^{\dot{\mu}\rho}=0$ REVISITED

In making the choice $\hat{\alpha}^{\dot{\mu}\rho} = 0$ in Sec. II, I simply followed the precedent set for a Riemannian space-time in Sec. V of Ref. 2: While this appears to lead again to an acceptable solution of the problem, the special case $S = 1$ indicates that it contains a possible pitfall which should not be overlooked.

Hitherto, attention has been confined exclusively to Eqs. (1.1). However, if one requires invariance under the full Lorentz group of equations in the flat space-time from which they originated, then one has to adjoin to Eqs. (1.1) the equations

$$\nabla_{\dot{\mu}_t}{}^{\nu_{s+1}}\zeta^{\dot{M}_tN_s} = \kappa\eta^{\dot{M}_{t-1}N_{s+1}}, \tag{4.1a}$$

$$\nabla^{\dot{\mu}_t}{}_{\nu_{s+1}}\eta^{\dot{M}_{t-1}N_{s+1}} = \kappa\zeta^{\dot{M}_tN_s}. \tag{4.1b}$$

When $S = 1$, (4.1) become, with the addition of supplementary terms,

$$\nabla_{\dot{\rho}}{}^{\nu}\zeta^{\dot{\mu}\dot{\rho}} = \kappa\eta^{\dot{\mu}\nu} + \check{\alpha}^{\dot{\mu}\nu}, \tag{4.2a}$$

$$\nabla^{\dot{\rho}}{}_{\nu}\eta^{\dot{\mu}\nu} = \kappa\zeta^{\dot{\mu}\dot{\rho}} + \check{\beta}^{\dot{\mu}\dot{\rho}}. \tag{4.2b}$$

Bearing Sec. II in mind, it is obvious that the absence of $\check{\alpha}^{\dot{\mu}\nu}$ and $\check{\beta}^{\dot{\mu}\dot{\rho}}$ would entail the constraint

$$S^{kl}{}_{\dot{\alpha}\dot{\beta}}\zeta^{\dot{\alpha}\dot{\beta}}{}_{;kl} = 0. \tag{4.3}$$

It suggests itself to follow precedent with the choice

$$\check{\alpha}^{\dot{\mu}\nu} = 0. \tag{4.4}$$

Granted, then, that $\hat{\alpha}^{\dot{\mu}\nu}$ and $\check{\alpha}^{\dot{\mu}\nu}$ are both absent, Eqs. (3.1a) and (4.2a) jointly imply

$$\nabla^{\dot{\mu}}{}_{\rho}\xi^{\nu\rho} - \nabla_{\dot{\rho}}{}^{\nu}\zeta^{\dot{\mu}\dot{\rho}} = 0. \tag{4.5}$$

One now arrives at the desired conclusion most easily by applying the operator $\nabla_{\dot{\mu}\nu}$ to the lhs of (4.5), which then takes the generic form $h^{kl}{}_{;lk}$, where h^{kl} is a bivector. However,

$$h^{kl}{}_{;lk} = \hat{K}_{kl}h^{kl} \tag{4.6}$$

(recall Sec. V A of Ref. 1). In the nontrivial U_4 (4.6) fails to vanish in general. It follows that in Eqs. (3.10a) and (4.2a) on cannot take $\hat{\alpha}^{\dot{\mu}\nu}$ and $\check{\alpha}^{\dot{\mu}\nu}$ to be zero spinors simultaneously. This conclusion will evidently not be affected if one prescribes minimal Hermitian coupling in place of minimal coupling. On the other hand, the conclusion in question cannot be drawn in a V_4.

V. CONCLUDING REMARK

The results of Sec. III point to a certain weakness inherent in the minimal coupling prescription which stands apart from the remarks of Sec. IV of Ref. 2. Normally, minimal coupling is understood to be a prescription which is to be applied to first-order equations. (The iterated equations are then usually not minimally coupled.) Here, however, one has a different state of affairs. The Proca equations (3.2) are minimally coupled, whereas their spinorial equivalents (3.9), or (3.10) together with (3.15), are not: Mere transcription has vitiated the prescription. In other words, whether the unconstrained equations are minimally coupled or not depends on whether one chooses to write them as tensor or as spinor equations.

[1] H. A. Buchdahl, J. Math. Phys. **30**, 700 (1989); where not otherwise indicated the terminology and notation of this paper are adhered to here.
[2] H. A. Buchdahl, J. Phys. A: Math. Gen. **15**, 1 (1982).
[3] H. A. Buchdahl, J. Phys. A: Math. Gen. **15**, 1057 (1982).
[4] Occasionaly, when there is no risk of confusion, all indices of given spinors are suppressed.
[5] E. M. Corson, *Tensors, Spinors and Relativistic Wave Equations* (Blackie, London, 1953), Sec. 27b.
[6] H. A. Buchdahl, J. Aust. Math. Soc. **2**, 369 (1962); knowledge of the definitions and notation in this paper is presupposed.

Separation of variables and exact solution to Dirac and Weyl equations in Robertson–Walker space-times

V. M. Villalba[a] and U. Percoco
Centro de Física, Instituto Venezolano de Investigaciones Científicas (IVIC), Apartado 21827, Caracas 1020-A, Venezuela

(Received 16 August 1989; accepted for publication 27 September 1989)

In this paper the separation of variables is presented in the Dirac equation in open, flat, and closed expanding cosmological Robertson–Walker universes. The equations governing the radial variable and the evolution of the time-dependent factor are obtained. An exact solution to the Weyl equation is derived for an arbitrary expansion factor of the Robertson–Walker metrics. An exact solution to Dirac equation in a universe filled with radiation is also presented.

I. INTRODUCTION

Recently there has been an increasing interest in quantum mechanical properties of particles in curved space-times. Audretsch and Schäfer[1,2] have presented a detailed analysis of the energy spectrum of the hydrogen atom in static Robertson–Walker universes. The study of a one-electron atom in a general curved space-time as well as the Hamiltonian of the Dirac equation in Fermi coordinates was found by Parker.[3,4] The Hawking radiation[5,6] is an appropriate example of the importance of the effects of strong gravitational fields in quantum mechanical processes. The construction of a quantum field theory in curved space-times and the definition of quantum vacuum is impossible without a careful study of one-particle states, i.e., without a detailed investigation of the exact solutions of relativisitic wave equations in curved backgrounds.

Some exact solutions to the Dirac equation in curved space-times have been reported,[7,8] and considerable attention has recently been paid to the study of de Sitter cosmological models and spatially flat Robertson–Walker universes.[9,10] Separation of variables in the Dirac equation was possible because of the simple form taken by the metric in these models, and an appropriate selection of the local frame. A useful method to carry out separation of variables in the previous cases is based on the obtainment of a complete set of first-order differential operators.[11] Nevertheless, exact solutions to the Dirac equation in spatially closed or open expanding Robertson–Walker universes can not be achieved using this technique. In the present article we exhibit a second-order method of separation of variables in the Dirac equation for the three types of Robertson–Walker metrics. This allows us to find the explicit structure of the Dirac spinor. Time and radial dependence of the solution are determined by two systems of two coupled ordinary differential equations. In the massless case, we succeed in finding exact solutions to the Weyl equation in a Robertson–Walker background for an arbitrary expansion factor and arbitrary space curvature.

The paper is organized as follows: In Sec. II, we write down the covariant generalization of the Dirac equation in cosmological Robertson–Walker universes and a complete separation of variables is achieved. In the same section we present the explicit structure of the spinor solution. In Sec. III the solution to the Weyl equation for each Robertson–Walker metric and arbitrary expansion factor is obtained. We present in Sec. IV an exact solution to the Dirac equation in spatially closed and open Robertson–Walker space-times with a metric representing a universe filled with radiation $(p=\rho/3)$.

II. DIRAC EQUATION IN ROBERTSON–WALKER SPACE-TIMES

The generalization of the Dirac equation in curved space-time is

$$(\gamma^\nu(x)\nabla_\nu + m)\Psi(x) = 0, \tag{2.1}$$

where the generalized Dirac matrices satisfy the anticommutation relations

$$\gamma^\nu(x)\gamma^\mu(x) + \gamma^\mu(x)\gamma^\nu(x) = 2g^{\nu\mu}, \tag{2.2}$$

and

$$\Gamma_\lambda = \frac{1}{4} g_{\mu\alpha}\left(\frac{\partial b^{\beta}_{\nu}}{\partial^{\lambda}_{x}} a^{\alpha}_{\beta} - \Gamma^{\alpha}_{\nu\lambda}\right) s^{\mu\nu}, \tag{2.3}$$

where

$$s^{\mu\nu} = \tfrac{1}{2}(\gamma^\mu\gamma^\nu - \gamma^\nu\gamma^\mu), \tag{2.4}$$

and the matrices b^{β}_{α}, a^{α}_{β} establish the connection between the Dirac matrices (γ) on a curved space-time and the Minkowski space $(\tilde{\gamma})$ Dirac matrices as follows:

$$\gamma_\mu = b^{\alpha}_{\mu}\tilde{\gamma}_\alpha, \quad \gamma^\mu = a^{\mu}_{\beta}\tilde{\gamma}^\beta. \tag{2.5}$$

The Robertson–Walker metric in spherical coordinates reads[12]

$$ds^2 = R^2(t)\left(\frac{dr^2}{1-\epsilon r^2} + r^2\, d\Omega^2\right) - dt^2, \tag{2.6}$$

with

[a] Postal address: Carmelitas 4282, Caracas 1010-A, Venezuela.

 0022-2488/90/030715-06$03.00

$$d\Omega^2 = d\theta^2 + \sin^2\theta\, d\varphi^2, \tag{2.7}$$

where

$$\epsilon = 0,\ +1,\ -1. \tag{2.8}$$

It is convenient to write the interval (2.6) in chronometrical coordinates expressed in the comoving frame,[13] where the expansion factor depends only on the time parameter τ

$$ds^2 = e^{\alpha(\tau)}\{dr^2 + \xi^2(r)d\Omega^2 - d\tau^2\}, \tag{2.9}$$

and

$$\xi(r) = \begin{cases} r, & \epsilon = 0, \\ \sinh r, & \epsilon = -1, \\ \sin r, & \epsilon = -1. \end{cases} \tag{2.10}$$

Using the relations (2.3) and (2.4) and choosing a diagonal tetrad a^α_β,

$$a^\alpha_\beta = \mathrm{diag}[e^{-\alpha/2}, e^{-\alpha/2}, e^{-\alpha/2}\xi(r)^{-1}, e^{-\alpha/2}(\xi(r)\sin\theta)^{-1}], \tag{2.11}$$

we obtain the spinor connections Γ_μ for the metrics (2.9)

(a) $\epsilon = -1$,

$$\Gamma_0 = 0, \quad \Gamma_1 = \tfrac{1}{4}e^\alpha\dot\alpha\gamma^0\gamma^1, \tag{2.12}$$

$$\Gamma_2 = \tfrac{1}{4}e^\alpha(\dot\alpha\sin^2(r)\gamma^0\gamma^2 + 2\sin(r)\cos(r)\gamma^1\gamma^2), \tag{2.13}$$

$$\Gamma_3 = \tfrac{1}{4}e^\alpha(\dot\alpha\sin^2(r)\sin^2(\theta)\gamma^0\gamma^3 + 2\cos(r)\sin(r) \times\sin^2(\theta)\gamma^1\gamma^3 + 2\sin^2(r)\sin(\theta)\cos(\theta)\gamma^2\gamma^3). \tag{2.14}$$

(b) $\epsilon = 1$,

$$\Gamma_0 = 0, \quad \Gamma_1 = \tfrac{1}{4}e^\alpha\dot\alpha\gamma^0\gamma^1, \tag{2.15}$$

$$\Gamma_2 = \tfrac{1}{4}e^\alpha(\dot\alpha\sinh^2(r)\gamma^0\gamma^2 + 2\sinh(r)\cosh(r)\gamma^1\gamma^2), \tag{2.16}$$

$$\Gamma_3 = \tfrac{1}{4}e^\alpha(\dot\alpha\sinh^2(r)\sin^2(\theta)\gamma^0\gamma^3 + 2\cosh(r) \times\sinh(r)\sin^2(\theta)\gamma^1\gamma^3 + 2\sinh^2(r)\sin(\theta) \times\cos(\theta)\gamma^2\gamma^3). \tag{2.17}$$

(c) $\epsilon = 0$,

$$\Gamma_0 = 0, \quad \Gamma_1 = \tfrac{1}{4}e^\alpha\dot\alpha\gamma^0\gamma^1, \tag{2.18}$$

$$\Gamma_2 = \tfrac{1}{4}e^\alpha(\dot\alpha r^2\gamma^0\gamma^2 + 2r\gamma^1\gamma^2), \tag{2.19}$$

$$\Gamma_3 = \tfrac{1}{4}e^\alpha(\dot\alpha r^2\sin^2(\theta)\gamma^0\gamma^3 + 2r\sin^2(\theta)\gamma^1\gamma^3 + 2r^2\sin(\theta)\cos(\theta)\gamma^2\gamma^3), \tag{2.20}$$

where 0, 1, 2, and 3 denote τ, r, θ, and φ, respectively; and where the dot denotes $d/d\tau$. Substituting (2.12)–(2.20) in (2.1), the Dirac equation in the Robertson–Walker background reads

$$\left\{\frac{\tilde\gamma^0}{e^{\alpha/2}}\partial_0 + \frac{\tilde\gamma^1}{e^{\alpha/2}}\partial_1 + \frac{\tilde\gamma^2}{e^{\alpha/2}\xi(r)}\partial_2 + \frac{\tilde\gamma^3}{e^{\alpha/2}\xi(r)\sin\theta}\partial_3 + m\right\}\Phi = 0, \tag{2.21}$$

where Φ is related to Ψ by

$$\Psi = \xi(r)^{-1}(\sin\theta)^{-1/2}e^{-3\alpha/4}\Phi. \tag{2.22}$$

Applying the method of separation of variables proposed in Ref. 11, it is possible to write Eq. (2.21) as a sum of two first-order differential operators $\hat{\mathbf{K}}_1$, $\hat{\mathbf{K}}_2$ satisfying the relation

$$[\hat{\mathbf{K}}_1, \hat{\mathbf{K}}_2] = 0, \quad \{\hat{\mathbf{K}}_1 + \hat{\mathbf{K}}_2\}\Sigma = 0, \tag{2.23}$$

$$\hat{\mathbf{K}}_1\Sigma = k\Sigma = -\hat{\mathbf{K}}_2\Sigma, \tag{2.24}$$

where

$$\hat{\mathbf{K}}_1 = -i\{\tilde\gamma^2\partial_2 + (\tilde\gamma^3/\sin\theta)\partial_3\}\tilde\gamma^1\tilde\gamma^0, \tag{2.25}$$

$$\hat{\mathbf{K}}_2 = -i\xi(r)\{(\tilde\gamma^0\partial_0 + \tilde\gamma^1\partial_1) + me^{\alpha/2}\}\tilde\gamma^1\tilde\gamma^0, \tag{2.26}$$

$$\Sigma = \tilde\gamma^1\tilde\gamma^0\Phi. \tag{2.27}$$

The operator $\hat{\mathbf{K}}_1$ corresponds to the "momentum" obtained by Brill and Wheeler[14] in the problem of separation of variables of the Dirac equation in the Schwarzschild metric. The physical angular momentum[15] $\hat{\mathbf{K}}$ is related to $\hat{\mathbf{K}}_1$ by the unitary transformation S, connecting Dirac matrices and the spinor Ψ in the diagonal tetradic gauge to Dirac matrices and the spinor Ψ_c in the Cartesian gauge, i.e.,

$$S\gamma^\nu S^{-1} = h^\nu_\beta\tilde\gamma^\beta = \gamma^\nu_c, \tag{2.28}$$

$$S\Psi = \Psi_c, \tag{2.29}$$

where γ^ν_c are Dirac matrices in the Cartesian tetrad gauge. The matrix h^ν_β and the operator S are given by the relations

$$h^\alpha_\beta = e^{-\alpha/2}\begin{bmatrix} \sin\theta\cos\varphi & \sin\theta\sin\varphi & \cos\theta & 0 \\ \xi^{-1}\cos\theta\cos\varphi & \xi^{-1}\cos\theta\sin\varphi & -\xi^{-1}\sin\theta & 0 \\ -\xi^{-1}\sin^{-1}\theta\cos\varphi & \xi^{-1}\sin^{-1}\theta\cos\varphi & 0 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}, \tag{2.30}$$

$$S = \exp(-(\varphi/2)\tilde\gamma^1\tilde\gamma^2)\exp(-(\theta/2)\tilde\gamma^3\tilde\gamma^1)\,\mathfrak{U}, \tag{2.31}$$

where

$$\mathfrak{U} = \tfrac{1}{2}(\tilde\gamma^1\tilde\gamma^2 + \tilde\gamma^2\tilde\gamma^3 + \tilde\gamma^3\tilde\gamma^1 + 1). \tag{2.32}$$

Using a suitable representation of Dirac matrices[16]

$$\tilde\gamma^0 = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}, \quad \tilde\gamma^k = \begin{pmatrix} 0 & \sigma^k \\ \sigma^k & 0 \end{pmatrix}, \quad k = 1,2,3, \tag{2.33}$$

and making the transformation (2.32) to Dirac matrices (2.33)

$$\mathfrak{U}\tilde\gamma^1\mathfrak{U}^{-1} = \tilde\gamma^3, \quad \mathfrak{U}\tilde\gamma^2\mathfrak{U}^{-1} = \tilde\gamma^1, \quad \mathfrak{U}\tilde\gamma^3\mathfrak{U}^{-1} = \tilde\gamma^2. \tag{2.34}$$

Equation (2.21) can be expressed as follows:

$$(-\sigma^2\partial_\theta + im(\sigma^1/\sin\theta) - ik)\tilde\Phi_1 = 0, \tag{2.35}$$

$$(\sigma^2\partial_\theta - im(\sigma^1/\sin\theta) - ik)\tilde\Phi_2 = 0, \tag{2.36}$$

with

$$\mathfrak{U}\Sigma = \tilde\Sigma = \begin{pmatrix} \tilde\Phi_1 \\ \tilde\Phi_2 \end{pmatrix}. \tag{2.37}$$

Using the algebra of Pauli matrices and taking into account the form of Eqs. (2.35) and (2.36), it is easy to see that the spinor $\tilde\Sigma$ has the following structure:

$$\tilde{\Sigma} = \begin{bmatrix} q(r,\tau)\chi_1(\theta) \\ q(r,\tau)\chi_2(\theta) \\ p(r,\tau)\chi_1(\theta) \\ -p(r,\tau)\chi_2(\theta) \end{bmatrix} \exp(im\varphi) . \tag{2.38}$$

The constant m in Eqs. (2.35) and (2.36) is the eigenvalue of the operator $-i\partial_\varphi$ which commutes with the operator $\hat{\mathbf{K}}_1$. We would expect m to take half-integer values because the spinor wave function Ψ_c has to be continuous everywhere, i.e.,

$$\Psi_c(\varphi = \varphi_0) = \Psi_c(\varphi = \varphi_0 + 2\pi) , \tag{2.39}$$

and since

$$S(\varphi_0) = -S(\varphi_0 + 2\pi) , \tag{2.40}$$

it becomes clear that

$$\Phi(\varphi_0) = -\Phi(\varphi_0 + 2\pi) , \tag{2.41}$$

with the above boundary conditions, $-i\partial_\varphi$ is a Hermitian operator, with eigenvalues

$$m = \pm\tfrac{1}{2}, \pm\tfrac{3}{2}, \pm\tfrac{5}{2}, \ldots . \tag{2.42}$$

Using the standard representation for the Pauli matrices, the spinor equation (2.35) splits in two equations

$$\left(\frac{d}{d\theta} + \frac{m}{\sin\theta}\right)\chi_2 - k\chi_1 = 0 , \tag{2.43}$$

$$\left(\frac{d}{d\theta} - \frac{m}{\sin\theta}\right)\chi_1 + k\chi_2 = 0 , \tag{2.44}$$

where

$$\tilde{\Phi}_1 = \begin{pmatrix} \chi_1 \\ \chi_2 \end{pmatrix} . \tag{2.45}$$

The ansatz

$$\chi_1 = \sin^m(\theta)\cos(\theta/2)\, g(\theta) , \tag{2.46}$$

$$\chi_2 = \sin^m(\theta)\sin(\theta/2)\, f(\theta) , \tag{2.47}$$

leads to

$$(x+1)\,\frac{dg}{dx} + \left(\frac{1}{2} + m\right) g = kf, \tag{2.48}$$

$$(x-1)\,\frac{df}{dx} + \left(\frac{1}{2} + m\right) f = kg , \tag{2.49}$$

where we have made the substitution

$$x = \cos\theta. \tag{2.50}$$

The solution of Eqs. (2.48), (2.49) is given in terms of the Jacobi polynomials[17] by the expressions

$$f(x) = c_0\, P_n^{(m+1/2,m-1/2)}(x) , \tag{2.51}$$

$$g(x) = c_0\, P_n^{(m-1/2,m+1/2)}(x) , \tag{2.52}$$

where c_0 is a constant of normalization and n is given by

$$n = |k| - |m| - \tfrac{1}{2} . \tag{2.53}$$

Therefore, the functions χ_1, χ_2 are

$$\chi_1 = c_0 \sin^m(\theta)\cos(\theta/2) P_n^{(m+1/2,m-1/2)}(\cos\theta) , \tag{2.54}$$

$$\chi_2 = c_0 \sin^m(\theta)\sin(\theta/2) P_n^{(m-1/2,m+1/2)}(\cos\theta) . \tag{2.55}$$

The separation of variables in Eq. (2.26) cannot be written in terms of first-order commuting differential operators. It is then convenient to rewrite Eq. (2.24) as

$$[(\tilde{\gamma}^0\partial_0 + me^{\alpha/2})\tilde{\gamma}^3\tilde{\gamma}^0 + (\tilde{\gamma}^0\partial_1 + i(k/\xi))]\,\tilde{\Sigma} = 0 . \tag{2.56}$$

Introducing the auxiliary function η defined by

$$\tilde{\Sigma} = [(\tilde{\gamma}^0\partial_0 + me^{\alpha/2})\tilde{\gamma}^3\tilde{\gamma}^0 + (\tilde{\gamma}^0\partial_1 - i(k/\xi))]\,\eta , \tag{2.57}$$

Eq. (2.56) reduces to

$$\{\hat{N}_1 + \hat{N}_2\}\,\eta = 0, \quad [\hat{N}_1,\hat{N}_2] = 0 , \tag{2.58}$$

where

$$\hat{N}_1 = (\tilde{\gamma}^0\partial_0 + me^{\alpha/2})(-\tilde{\gamma}^0\partial_0 + me^{\alpha/2}) , \tag{2.59}$$

$$\hat{N}_2 = (\tilde{\gamma}^0\partial_1 + i(k/\xi))(\tilde{\gamma}^0\partial_1 - i(k/\xi)) , \tag{2.60}$$

$$\hat{N}_2\eta = -\hat{N}_1\eta = \lambda^2\eta . \tag{2.61}$$

Using the representation (2.33) for $\tilde{\gamma}^0$ and (2.59)–(2.61) we can write η as follows:

$$\eta = \begin{pmatrix} \alpha(\tau)a(r)C_1(\vartheta,\phi) \\ \alpha(\tau)a(r)C_2(\vartheta,\phi) \\ \beta(\tau)b(r)C_3(\vartheta,\phi) \\ \beta(\tau)\,b(r)C_4(\vartheta,\phi) \end{pmatrix} . \tag{2.62}$$

Applying the auxiliary condition (2.57) on η, $\tilde{\Sigma}$ can be written as

$$\tilde{\Sigma} = \begin{pmatrix} \alpha(\tau)b(r)\chi_1(\vartheta) \\ \alpha(\tau)b(r)\chi_2(\vartheta) \\ \beta(\tau)a(r)\chi_1(\vartheta) \\ -\beta(\tau)a(r)\chi_2(\vartheta) \end{pmatrix} \exp(im\phi) , \tag{2.63}$$

where $\alpha(\tau)$, $\beta(\tau)$, $a(r)$, and $b(r)$ satisfy

$$\left(\frac{d}{d\tau} - ime^{\alpha/2}\right)\alpha = -i\lambda\beta , \tag{2.64}$$

$$\left(\frac{d}{d\tau} + ime^{\alpha/2}\right)\beta = -i\lambda\alpha , \tag{2.65}$$

$$\left(\frac{d}{dr} + \frac{k}{\xi}\right) a = \lambda b , \tag{2.66}$$

$$\left(\frac{d}{dr} - \frac{k}{\xi}\right) b = -\lambda a . \tag{2.67}$$

Solution of the first two equations requires the knowledge of the expansion factor $e^{\alpha(\tau)}$. From the other two equations three cases emerge.

(i) $\epsilon = 0$. Then $\xi = r$ and the solution is given by

$$a(r) = a_0 r^{1/2} J_{k+1/2}(\lambda r) , \tag{2.68}$$

$$b(r) = a_0 r^{1/2} J_{k-1/2}(\lambda r) , \tag{2.69}$$

where a_0 is an arbitrary constant.

(ii) $\epsilon = 1$. In this case $\xi = \sin r$. Since the equations have the same structure as the system (2.43)–(2.44), we can use the ansatz (2.46) and (2.47). Hence, we obtain

$$a(r) = c_0 \sin^k(r)\sin(r/2) F(\tfrac{1}{2} - \lambda + k, \tfrac{1}{2} + \lambda + k, \tfrac{3}{2} + k, (1-\cos r)/2) , \tag{2.70}$$

$$b(r) = c_0[(\tfrac{1}{2} + k)/\lambda]\sin^k(r)\cos(r/2) F(\tfrac{1}{2} - \lambda + k, \tfrac{1}{2} + \lambda + k, \tfrac{1}{2} + k, (1-\cos r)/2) , \tag{2.71}$$

where c_0 is an arbitrary constant and $F(a,b,c,z)$ are the Gauss hypergeometric functions.[17]

(iii) $\epsilon = -1$. In this case $\xi = \sinh r$ and the solution to (2.66), (2.67) is

$$a = a_0 \sinh^k(r)\cosh(r/2)F(\tfrac{1}{2} - i\lambda + k, \tfrac{1}{2} + i\lambda + k, \tfrac{3}{2} + k, (1-\cosh r)/2), \qquad (2.72)$$

$$b = a_0(\tfrac{1}{2} + k)/\lambda \sinh^k(r)\sinh(r/2)F(\tfrac{1}{2} - i\lambda + k, \tfrac{1}{2} + i\lambda + k, \tfrac{1}{2} + k, (1-\cosh r)/2). \qquad (2.73)$$

The solution to the Dirac equation in the Cartesian gauge is obtained by performing the transformation (2.29), which gives the familiar form of the solution in terms of the spherical harmonics[18] (the details of this calculation can be found in the Appendix):

$$\Psi_0 = \xi(r)^{-1}e^{-3\alpha/4} \times \begin{pmatrix} (k-m+\frac{1}{2})^{1/2}\beta(\tau)a(r)Y_k^{m-1/2}(\theta,\varphi) \\ (k+m+\frac{1}{2})^{1/2}\beta(\tau)a(r)Y_k^{m+1/2}(\theta,\varphi) \\ -(k+m+\frac{1}{2})^{1/2}\alpha(\tau)b(r)Y_k^{m-1/2}(\theta,\varphi) \\ (k-m+\frac{1}{2})^{1/2}\alpha(\tau)b(r)Y_k^{m+1/2}(\theta,\varphi) \end{pmatrix}. \qquad (2.74)$$

The expansion factor $e^{\alpha(\tau)}$ is determined by the equation of state.[13]

III. SOLUTION TO THE WEYL EQUATION

Neutrinos in a Robertson–Walker background are described by the Weyl equation, which corresponds to the massless limit of the Dirac equation plus the chirality condition. The latter, in the representation (2.33), reads

$$(1 - i\gamma_5)\Psi_c = 0, \qquad (3.1)$$

where

$$\gamma_5 = \tilde{\gamma}_0\tilde{\gamma}_1\tilde{\gamma}_2\tilde{\gamma}_3. \qquad (3.2)$$

Then, the Weyl spinor can be expressed in terms of the solution of the Dirac equation as follows:

$$(1 + i\gamma_5)\Psi_c = \Psi_w. \qquad (3.3)$$

The matrix γ_5 commutes with $\mathfrak{U}$ and with the transformation matrix S, therefore we can substitute Ψ_c by Φ in (3.3). Then, we can use the results of the preceding section. Indeed, Eqs. (2.43) and (2.44) apply with no changes to the present case, while solutions to Eqs. (2.64) and (2.65) take the form

$$\alpha \propto e^{i\lambda\tau}, \qquad (3.4)$$

$$\beta \propto e^{i\lambda\tau}. \qquad (3.5)$$

Using, now, the explicit representation of the matrix $\mathfrak{U}$, the chirality condition, and (2.63), the solution of the Weyl equation in the diagonal gauge takes the form

$$\Psi = \xi^{-1}(r)\sin^{-1}\theta \exp(-\tfrac{3}{4}\alpha(\tau))e^{im\phi}e^{i\lambda t}\begin{pmatrix} q \\ iq \end{pmatrix}, \qquad (3.6)$$

where

$$q = \begin{pmatrix} (1-i)[(a+ib)\chi_1 - (1+i)(a-ib)\chi_2] \\ (1-i)[(a+ib)\chi_1 + (1+i)(a-ib)\chi_2] \end{pmatrix}. \qquad (3.7)$$

The solution in the Cartesian gauge can be obtained applying the transformation (2.31) to the spinor (3.6).

The above result represents the general solution to the Weyl equation for each Robertson–Walker universe and for an arbitrary expansion factor $\alpha(\tau)$.

IV. SOLUTION TO DIRAC EQUATION IN A RADIATION UNIVERSE

In this section we study the Dirac equation in closed and open Robertson–Walker radiation universes. In the chronometrical coordinates the line element for the closed universe takes the form

$$ds^2 = \sin^2\tau[-d\tau^2 + dr^2 + \sin^2 r(d\theta^2 + \sin^2\theta\, d\varphi^2)]. \qquad (4.1)$$

The system of equations (2.64), (2.65) with the expansion factor $a = \sin\tau$, reads

$$\left(\frac{d}{d\tau} + im\sin\tau\right)\beta = -i\lambda\alpha, \qquad (4.2)$$

$$\left(\frac{d}{d\tau} - im\sin\tau\right)\alpha = -i\lambda\beta, \qquad (4.3)$$

from which we obtain

$$\left(\frac{d^2}{d\tau^2} - im\cos\tau + m^2\sin^2\tau + \lambda^2\right)\alpha = 0, \qquad (4.4)$$

$$\left(\frac{d^2}{d\tau^2} + im\cos\tau + m^2\sin^2\tau + \lambda^2\right)\beta = 0. \qquad (4.5)$$

The solution to these equations can be obtained, in a way analogous to the Mathieu equation, as a series of periodic functions. However, this solution requires a five terms recurrence relation for the coefficients.

A simpler solution can be found by realizing that Eqs. (4.4) and (4.5) are of the form of the Whittaker–Hill equation.[19] Indeed after the change

$$\alpha = e^{-im\cos\tau}Y_1 + e^{im\cos\tau}Y_2, \qquad (4.6)$$

$$\beta = e^{-im\cos\tau}\chi_1 + e^{im\cos\tau}\chi_2, \qquad (4.7)$$

can be written as a system of equations of the Ince type.[19] That is,

$$\left(\frac{d^2}{d\theta^2} + i4m\sin 2\theta\frac{d}{d\theta} + 4\lambda^2 + i8m\cos\theta\right)\chi_1 = 0, \qquad (4.8)$$

$$\left(\frac{d^2}{d\theta^2} - i4m\sin 2\theta\frac{d}{d\theta} + 4\lambda^2\right)\chi_2 = 0, \qquad (4.9)$$

$$\left(\frac{d^2}{d\theta^2} + i4m\sin 2\theta\frac{d}{d\theta} + 4\lambda^2\right)Y_1 = 0, \qquad (4.10)$$

$$\left(\frac{d^2}{d\theta^2} - i4m\sin 2\theta\frac{d}{d\theta} + 4\lambda^2 - i8m\cos 2\theta\right)Y_2 = 0, \qquad (4.11)$$

where $\tau = 2\theta$. The solution to the system (4.8)–(4.10) is

$$\alpha = e^{-im\cos\tau}\left[\sum_{r=0}^{\infty} C_r \sin r\tau\right] + e^{im\cos\tau}\left[\frac{i}{2\lambda}\sum_{r=1}^{\infty} rG_r \cos r\tau\right], \tag{4.12}$$

$$\beta = e^{-im\cos\tau}\left[\frac{i}{2\lambda}\sum_{r=1}^{\infty} rC_r \sin r\tau\right] + e^{im\cos\tau}\left[\sum_{r=0}^{\infty} G_r \cos r\tau\right], \tag{4.13}$$

where the coefficients C_r and G_r satisfy the three terms recurrence relations

$$2(1-\lambda^2)C_1 + C_2 = 0, \tag{4.14}$$

$$-(i/2)(r-1)mC_{r-1} + (r^2 - 4\lambda^2)C_r + (i/2)m(r+1)C_{r+1} = 0, \tag{4.15}$$

$$2(1-\lambda^2)G_1 + G_2 = 0, \tag{4.16}$$

$$(i/2)(r-1)mG_{r-1} + (r^2 - 4\lambda^2)G_r - (i/2)m(r+1)G_{r+1} = 0. \tag{4.17}$$

Equations (4.15) and (4.17) are valid for $r \geqslant 2$.

The solution corresponding to an open universe filled with radiation, can be obtained making the substitution $\tau \to i\tau$ in the equations for α and β.

The above solutions are absolutely and uniformly convergent for any value of τ. A detailed analysis of this point has been carried out by Urwin and Arscott.[20]

These results indicate that interesting solutions could be obtained if different matter distributions, in presence of electromagnetic radiation, are considered. This will be done in a forthcoming paper.

ACKNOWLEDGMENTS

We express our indebtedness to L. A. Núñez for his careful reading of the manuscript and interesting discussions.

APPENDIX

We give the explicit form of the transformation matrix S in the representation (2.33) of Dirac matrices

$$S = \frac{1}{2}\begin{pmatrix} Z & 0 \\ 0 & Z \end{pmatrix}, \tag{A1}$$

where

$$Z = \exp(-(i\varphi/2)\sigma^3)\exp(-(i\theta/2)\sigma^2) \times [1 + i(\sigma^1 + \sigma^2 + \sigma^3)], \tag{A2}$$

or

$$Z = \begin{pmatrix} e^{-i\varphi/2} & 0 \\ 0 & e^{i\varphi/2} \end{pmatrix}\begin{pmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{pmatrix} \times \begin{pmatrix} 1+i & 1+i \\ i-1 & 1-i \end{pmatrix}. \tag{A3}$$

In order to obtain the form of the spinor, solution of the Dirac equation in the Cartesian gauge, we let S act on $\widetilde{\Sigma}$ given in (2.63), and with the help of Eqs. (2.45)–(2.47) we obtain for the lower components of Ψ

$$\Psi_2 = \sin^{m-1/2}\theta\begin{pmatrix} e^{-i\varphi/2} & 0 \\ 0 & e^{i\varphi/2} \end{pmatrix} \times \begin{pmatrix} \cos(\theta/2) & -\sin(\theta/2) \\ \sin(\theta/2) & \cos(\theta/2) \end{pmatrix} \times \begin{pmatrix} -\cos(\theta/2)P_n^{(m-1/2,m+1/2)(x)} \\ \sin(\theta/2)P_n^{(m+1/2,m-1/2)(x)} \end{pmatrix}, \tag{A4}$$

where $x = \cos\theta$. Taking into account the recurrence relations for the Jacobi polynomials

$$(1-x)P_n^{(a+1,b)}(x) + (1+x)P_n^{(a,b+1)}(x) = 2P_n^{(a,b)}, \tag{A5}$$

$$P_n^{(a,b-1)}(x) - P_n^{(a-1,b)}(x) = P_{n-1}^{(a,b)}(x), \tag{A6}$$

we obtain

$$\Psi_2 = \sin^{m-1}\theta\begin{pmatrix} -\cos(\theta/2)P_n^{(m-1/2,m+1/2)}(x) \\ \sin(\theta/2)P_n^{(m+1/2,m-1/2)}(x) \end{pmatrix}. \tag{A7}$$

Using

$$P_{k-a}^{(a,a)}(x) = \frac{2^a k!}{(k+a)!}\frac{d^a}{dx^a}P_k(x), \tag{A8}$$

Ψ_2 can be written as

$$\Psi_2 = \frac{2^{m-1/2}k!}{(k+m-1/2)!}\sqrt{\frac{2}{2k+1}} \times \begin{pmatrix} -Y_{k,m-1/2}(x) \\ \dfrac{1}{(k+m+\frac{1}{2})^{1/2}}(k-m+\tfrac{1}{2})^{1/2}Y_{k,m+1/2}(x) \end{pmatrix}. \tag{A9}$$

Finally the explicit form of Ψ_1 is obtained in a similar fashion.

[1] J. Audretsch and G. Schäfer, Gen. Relativ. Gravit. **9**, 243 (1978).
[2] J. Audretsch and G. Schäfer, Gen. Relativ. Gravit. **9**, 489 (1978).
[3] L. Parker, Phys. Rev. D **22**, 1922 (1980).
[4] L. Parker, Phys. Rev. Lett. **44**, 1559 (1980).
[5] S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975); G. W. Gibbons and S. W. Hawking, Phys. Rev. D **15**, 2738 (1977).
[6] V. S. Otchik, Class. Quant. Grav. **2**, 539 (1985).
[7] L. P. Chimento and M. S. Mollerach, Phys. Rev. D **34**, 3689 (1986); Phys. Lett. A **121**, 7 (1987).
[8] M. A. Castagnino, C. D. El Hasi, F. D. Mazzitelli, and J. P. Paz, Phys. Lett. A **128**, 25 (1988).
[9] A. O. Barut and I. H. Duru, Phys. Rev. D **36**, 3705 (1987).
[10] G. V. Shishkin and I. E. Andrushkevich, Vestsi. Akad. Nauk BSSR Fiz.-Mat. Navuk **4**, 109 (1986).
[11] I. E. Andrushkevich and G. V. Shishkin, Theoret. Math. Phys. **70**, 204 (1987).
[12] S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).
[13] L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Pergamon, Oxford, 1971).
[14] D. Brill and J. A. Wheeler, Rev. Mod. Phys. **29**, 465 (1957).

[15]A. S. Davydov, *Quantum Mechanics* (Pergamon, Oxford, 1965).
[16]J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Springer, New York, 1976).
[17]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).
[18]P. Berestestskii, E. M. Lifshitz, and L. P. Pitaevskii, *Relativistic Quantum Mechanics* (Pergamon, Oxford, 1971).
[19]F. M. Arscott, Proc. Roy. Soc. Edinburgh A **67**, 265 (1967).
[20]K. M. Urwin and F. M. Arscott, Proc. Roy. Soc. Edinburgh Sect. A **69**, 28 (1969).

Recursion relation of character expansion coefficient in SU(3) lattice gauge theory

N. Tanimura
Justus-Liebig Universität Giessen, 6300 Giessen, West Germany
O. Tanimura
Araki 1106, Toyama, Japan

(Received 18 January 1989; accepted for publication 25 October 1989)

In the SU(3) lattice gauge theory the recursion relations of coefficients of the character expansion are derived through the Schwinger–Dyson equation for the ensemble average containing multilink variables. The relations form two kinds of recursion relations which correspond to two independent integers λ and μ of the SU(3) $(\lambda\mu)$ representation of the Young tableau. The property of the recursion relations in comparison with those of the U(1) and SU(2) groups is discussed.

I. INTRODUCTION

In the Wilson formulation of lattice gauge theory[1] in the quantum chromodynamics (QCD) the Schwinger–Dyson equation[2] plays an important role in the inspection of the properties of the QCD—confinement and asymptotic freedom. Brower and Nauenberg[3] have presented an exact solution in the differential form of the Schwinger–Dyson equation in the limit of $N\to\infty$ at fixed N/g^2 (g denotes the coupling constant) in SU(N) and shown the occurrence of the third-order phase transition in any dimension. Xue *et al.*[4] have developed the method of the evaluation of the Wilson loop by use of the Schwinger–Dyson variational equation. This method is expected to be able to treat those quantities that the Monte Carlo method fails to simulate, such as Wilson loops in large size and the loops in the scaling region.[5] In these approaches, the Schwinger–Dyson equation is derived from the variation of one link variable, e.g., $\delta_U\,\langle\mathrm{tr}(UA)\rangle$, where U denotes the link variable and A is an arbitrary matrix.

In this paper we derive the Schwinger–Dyson equation from the variation of the average on the trial action $S_0(=\mathrm{tr}(UJ+J^\dagger U^\dagger))$ for multilink variables in the SU(3) group, $\delta_U\langle\mathrm{tr}(UA_1)\mathrm{tr}(UA_2)\cdots\mathrm{tr}(UA_n)\rangle$, where A_i's are again arbitrary matrices. We then express the Schwinger–Dyson equation in terms of the coefficients of the character expansion[6,7] for the single-plaquette action. From a series of these equations we derive the two kinds of recursion relations of the coefficients functions.

II. CHARACTER EXPANSION

We start with the Wilson action[1]

$$S=\beta\sum_P\mathrm{tr}(U_P+U_P^\dagger),\tag{1}$$

where $U_P(=U_1U_2U_3^\dagger U_4^\dagger)$ represents the single-plaquette variable. In SU(3) lattice gauge theory the exponentiated single-plaquette action is expanded by the character, $\chi_{\lambda\mu}(U_p)$, as follows[6,7]:

$$\exp[\beta\,\mathrm{tr}(U_P+U_P^\dagger)]=N(\beta)\sum_{\lambda\mu}d_{\lambda\mu}\,A_{\lambda\mu}(\beta)\chi_{\lambda\mu}(U_P),\tag{2}$$

with

$$d_{\lambda\mu}=\tfrac{1}{2}(\lambda+1)(\mu+1)(\lambda+\mu+2),\tag{3}$$

where $N(\beta)$ denotes the normalization factor and $d_{\lambda\mu}$ is the dimension of the representation (λ,μ) for the SU(3) Young tableau. From the properties of the group integration we find that $A_{\mu\lambda}=A_{\lambda\mu}$ and $\chi_{\lambda\mu}=\chi^*_{\mu\lambda}$, with zero or positive integer values of λ and μ. Specifically, $A_{00}=1$ and $\chi_{00}=1$. With the use of the orthogonality relation of the character,

$$\int\chi^*_{\lambda\mu}(U)\chi_{\lambda'\mu'}(U)dU=\delta_{\lambda\lambda'}\,\delta_{\mu\mu'}.\tag{4}$$

We write $A_{\lambda\mu}(\beta)$ as

$$A_{\lambda\mu}(\beta)=\frac{1}{N(\beta)d_{\lambda\mu}}\int dU_P\,\chi_{\lambda\mu}(U_P)\times\exp[\beta\,\mathrm{tr}(U_P+U_P^\dagger)].\tag{5}$$

As simple examples let us take $(\lambda,\mu)=(1,0)$ and $(1,1)$. The corresponding characters are given by

$$\begin{aligned}\chi_{10}(U)&=\mathrm{tr}(U),\\ \chi_{11}(U)&=\mathrm{tr}(U)\mathrm{tr}(U^\dagger)-1.\end{aligned}\tag{6}$$

Substituting Eqs. (6) into Eq. (5) and performing some group integrations[8,9] we find

$$A_{10}(\beta)=\frac{1}{N(\beta)}\beta\sum_{k=0}^{1}\beta^{1-k}\sum_{r=0}^{2k}\binom{2k}{r}\beta^{2r}\Gamma^{(k)}_{1+r-k}\tag{7}$$

and

$$A_{11}(\beta)=\frac{1}{N(\beta)}\beta^2\sum_{s=0}^{1}\sum_{k=0}^{2-2s}\binom{2-2s}{k}\beta^{2-2s-k}\times\sum_{r=0}^{2k+s}\binom{2k+s}{r}\beta^{2r}\Gamma^{(k+s)}_{2+r-k-s}(\beta),\tag{8}$$

with

$$N(\beta)=\Gamma^{(0)}_0(\beta)$$

and

 0022-2488/90/030721-04$03.00

$$\Gamma_r^{(k)}(\beta) = \sum_{n=0}^{\infty} \sum_{m=0}^{n} \frac{2^{m+1}(3n+3r+3)!\beta^{2n+m}}{(n+r+1)!(n+r+k+2)!(2n+m+3r+3)!(n-m)!m!}. \tag{9}$$

Here we note that $N(\beta)$ is nothing but the one-link partition function $Z_0(\beta)$.[9] With the use of characters with larger values of λ and μ, one can calculate the corresponding $A_{\lambda\mu}$ from Eq. (5). From a series of the functions of $A_{\lambda\mu}$'s we find a general expression for $A_{\lambda\mu}$ with arbitrary integers of λ and μ ($\lambda \geqslant \mu \geqslant 0$): By setting $n = \lambda + \mu$ it can be described by

$$A_{\lambda\mu}(\beta) = \frac{1}{N(\beta)} \beta^n \sum_{s=0}^{\mu} s! \binom{\lambda}{s}\binom{\mu}{s} \times \sum_{k=0}^{n-2s} \binom{n-2s}{k} \beta^{n-2s-k} \times \sum_{r=0}^{2k+s} \binom{2k+s}{r} \beta^{2r} \Gamma^{(k+s)}_{n+r-k-s}(\beta). \tag{10}$$

We note that the β expansion of $A_{\lambda\mu}$ starts with the power of $\lambda + \mu$ [see, also, Eq. (22)], which is consistent with the case of the modified Bessel function $I_\mu(\beta)$ whose β series starts with the power of μ.

III. THE SCHWINGER–DYSON EQUATION

We begin with the definition of the average on the generating function of the one-link integral:

$$\langle f(U)\rangle_0 = \frac{\int dU f(U) \exp[\mathrm{tr}(UJ + J^\dagger U^\dagger)]|_{J=\beta 1}}{\int dU \exp[\mathrm{tr}(UJ + J^\dagger U^\dagger)]|_{J=\beta 1}},$$

where $f(U)$ is a given function of the link variable U. Here J stands for an arbitrary 3×3 matrix of the group $\mathrm{GL}(3,C)$. The index $J = \beta\mathbb{1}$ means that after the group integration the matrix J is set to β times the unit matrix $\mathbb{1}$.

In order to express the plaquette integral dU_P, we set $U_P = U$ and make a transformation $(U_1, U_2, U_3, U_4) \to (U, W)$, where $dU_P = dU\, dW$. Then Eq. (5) is reduced to

$$A_{\lambda\mu}(\beta) = (1/d_{\lambda\mu})\langle \chi_{\lambda\mu}\rangle_0.$$

Let us consider the following one-link average for the purpose of the action variation of the Schwinger–Dyson type[3]:

$$\langle L_{km}(\lambda^\alpha)\rangle_0 \equiv \langle \mathrm{tr}(\lambda^\alpha U V_1)\mathrm{tr}(UV_2)\cdots \mathrm{tr}(UV_k) \times \mathrm{tr}(V_{k+1}^\dagger U^\dagger)\cdots \mathrm{tr}(V_{k+m}^\dagger U^\dagger)\rangle_0, \tag{11}$$

where λ^α denotes the SU(3) Gell-Mann matrix with an integer suffix $\alpha (1 \leqslant \alpha \leqslant 8)$ and V_r is the 3×3 arbitrary matrix of constant parameter. Performing the action variation $U \to \exp(i\epsilon_a \lambda^\alpha) U$ (ϵ_a is an infinitesimal parameter) in Eq. (11) and using the orthonormal property of the Gell-Mann matrix we find

$$[3 - \tfrac{1}{3}(k-m)]\langle L_{km}(1)\rangle_0 + \sum_{r=2}^{k} \langle \mathrm{tr}(UV_1UV_r)L_{km;1r}\rangle_0 - \sum_{r=k+1}^{m} \langle \mathrm{tr}(V_1V_r^\dagger)L_{km;1r}\rangle_0 + \langle L_{km;1}\, \mathrm{tr}(UV_1UJ)\rangle_0 - \langle L_{km;1}\, \mathrm{tr}(V_1^\dagger J^\dagger)\rangle_0 - \tfrac{1}{3}\langle L_{km}(1)\mathrm{tr}(UJ)\rangle_0 + \tfrac{1}{3}\langle L_{km}(1)\mathrm{tr}(J^\dagger U^\dagger)\rangle_0 = 0, \tag{12}$$

with

$$L_{km;1} = L_{km}(1)/\mathrm{tr}(UV_1),$$
$$L_{km;1r} = \begin{cases} L_{km}(1)/[\mathrm{tr}(UV_1)\mathrm{tr}(UV_r)] & (r \leqslant k), \\ L_{km}(1)/[\mathrm{tr}(UV_1)\mathrm{tr}(V_r^\dagger U^\dagger)] & (r > k), \end{cases} \tag{13}$$

where the denominators are assumed to be nonzero c numbers that cancel the same-terms in the $L_{km}(1)$. Equation (12) is the Schwinger–Dyson equation for a function containing multilink variables.

Let us compare Eq. (12) with the Schwinger–Dyson equations for the groups U(1) and SU(2). Here we set $m = 0$ for L_{km} in Eq. (11), since the term $\Pi_r\, \mathrm{tr}(U^\dagger V_r^\dagger)$ does not lead to independent relations. For U(1) we drop the symbol λ and the suffix α and take the action variation $U \to \exp(i\epsilon)U$ in $\langle L_{k0}\rangle_0$. Then we find

$$k\left\langle \prod_{r=1}^{k} UV_r \right\rangle_0 + \left\langle UJ \prod_{r=1}^{k} UV_r \right\rangle_0 - \left\langle J^* V_1 \prod_{r=2}^{k} UV_r \right\rangle_0 = 0. \tag{14}$$

The Schwinger–Dyson equation for SU(2) is reduced to

$$\left(2 - \frac{k}{2}\right)\langle L_{k0}(1)\rangle_0 + \sum_{r=2}^{k} \langle \mathrm{tr}(UV_1 u V_r)L_{k0;1r}\rangle_0 + \langle L_{k0;1}\, \mathrm{tr}(UV_1UJ)\rangle_0 - \langle L_{k0;1}\, \mathrm{tr}(V_1^\dagger J^\dagger)\rangle_0 = 0. \tag{15}$$

Here we note that the terms corresponding to the last two terms in Eq. (12) are canceled out because of $\mathrm{tr}(U^\dagger) = \mathrm{tr}(U)$ in SU(2).

IV. RECURSION RELATION

We first consider the Schwinger–Dyson equation for the groups U(1) and SU(2). The Young tableaux of the groups U(1) and SU(2) are represented with the single integer k instead of (λ,μ) and the corresponding dimensions d_k are equal to 1 and $k+1$, respectively. Then the character expansion of the exponentiated single-plaquette action is given by Eq. (2) modified with the replacement of $(\lambda,\mu) \to k$. We notice that the index k does not represent the number of the dimension, contrary to the conventional character expansion.[7] Simple group integrations of Eq. (5) for U(1) and SU(2) yield, for $N(\beta)$ and $A_k(\beta)$:

$$N(\beta)=\begin{cases}I_0(2\beta), & \text{for U(1)},\\ I_1(4\beta)/2\beta, & \text{for SU(2)},\end{cases}$$
$$A_k(\beta)=\begin{cases}I_k(2\beta)/I_0(2\beta), & \text{for U(1)},\\ I_{k+1}(4\beta)/I_1(4\beta), & \text{for SU(2)},\end{cases} \tag{16}$$

where I_k represents the modified Bessel function.

We perform some group integrations of the Schwinger–Dyson equation for $\langle L_{k0}\rangle_0$ [Eq. (11)] and then set all $V_r(r=1,...,k)$ to be unit matrices. This leads to the following recursion relations of the modified Bessel function, respectively:

$$\begin{aligned} &kI_k(2\beta)/\beta=I_{k-1}(2\beta)-I_{k+1}(2\beta), && \text{for U(1)},\\ &(k+1)I_{k+1}(4\beta)/2\beta=I_k(4\beta)-I_{k+2}(4\beta), && \text{for SU(2)}.\end{aligned} \tag{17}$$

In contrast to the cases of U(1) and SU(2) the Young tableau of SU(3) is specified with two independent integers λ and μ. This implies that the Schwinger–Dyson equation (12) leads to two independent recursion relations. We practically carry out the integration in Eq. (12) and again set all V_r's and $V_r^\dagger$'s to be unit matrices. We then find a relation among the $A_{\lambda\mu}$'s. For simple examples of $(k,m)=(1,0)$ and $(1,1)$ Eq. (12) yields, respectively,

$$3(A_{20}-A_{10})+2(A_{11}-1)+(6/\beta)A_{10}=0 \tag{18}$$

and

$$5A_{21}-2A_{20}-3A_{10}+(8/\beta)A_{11}=0. \tag{19}$$

With the successive variations of k and m we obtain a series of equations among the $A_{\lambda\mu}$'s. From these equations we find the following two kinds of recursion relations.

(i) The recursion relation for which the largest value of $\lambda+\mu$ is kept constant ($=n\geqslant 2$) is

$$\begin{aligned}(n+1)A_{n-r,r}&+[(n-r)/(r+1)]A_{n-r-1,r+1}\\ =&-[r(n+1)/(r+1)]A_{n-r,r-1}\\ &+(n+1)A_{n-r-2,r+1}+nA_{n-r-2,r}\\ &-[(n+1)(n-r)/\beta]A_{n-r-1,r},\end{aligned} \tag{20}$$

where $A_{\lambda\mu}$ with $\lambda<0$ or $\mu<0$ is assumed zero. We note that the $A_{\lambda\mu}$'s on the lhs take the largest values ($=n$) of $\lambda+\mu$, while those on the rhs take smaller values $\lambda+\mu<n$. Hence the recursion relation (20) can be used to obtain $A_{n-r,r}$ from $A_{n-r-1,r+1}$ or vice versa with the use of $A_{\lambda\mu}(\lambda+\mu<n)$.

(ii) The recursion relation for the determination of A_{n1} ($n\geqslant 2$) is

$$\begin{aligned}(n+3)&A_{n1}\\ =&-(n+2)A_{n0}-(n/2)A_{n-1,0}+\tfrac{3}{2}(n+2)A_{n-2,2}\\ &+(n+2)A_{n-2,1}-[n(n+2)/\beta]A_{n-1,1}.\end{aligned} \tag{21}$$

We again note that the $\lambda+\mu$ values of $A_{\lambda\mu}$ on the rhs are smaller than that of A_{n1}.

By using recursion relations (20) and (21) we can find $A_{\lambda\mu}$ with any integer value of λ and μ. We begin with the known quantities of A_{10} and A_{11} at a given value β. From (i) with $n=2$ and $r=1$, i.e., Eq. (18), we obtain A_{20}. Starting with $k=2$ we repeat the following three processes by incrementing k by 1 until the required $A_{\lambda\mu}$ is reached.

(a) We fix A_{k1} from (ii) with $n=k$.

(b) We fix $A_{k+1,0}$ from (i) with $n=k+1$, $r=0$.

(c) We fix $A_{k-r,r+1}$ ($1\leqslant r\leqslant[(k+1)/2]$) from (i) with $n=k+1$.

where [ ] denotes the Gauss symbol. Here we have used the symmetry property of the function $A_{\lambda\mu}=A_{\mu\lambda}$.

It is quite difficult to prove the recursion relations analytically. However, we have found that the $A_{\lambda\mu}$ obtained from the recursion relations agree numerically with $A_{\lambda\mu}$ of Eq. (10), which is derived from the direct group integration of Eq. (5). Since the recursive and direct integration methods are independent, we obtain a confirmation to the recursion relations. (In a previous paper[10] we used the present recursion relations [Eq. (9) in Ref. 10 is nothing but Eq. (18) in the present text].)

Let us check the behavior of $A_{\lambda\mu}$ at $\beta\to 0$. From Eq. (10) one sees that at $\beta\to 0$,

$$\begin{aligned}A_{\lambda\mu}(\beta)&\sim\beta^{\lambda+\mu}\sum_{s=0}^{\mu}s!\binom{\lambda}{s}\binom{\mu}{s}\frac{2}{(s+1)!(\lambda+\mu+2)!}\\ &\sim\frac{2}{(\lambda+\mu+2)(\mu+1)!(\lambda+1)!}\beta^{\lambda+\mu}.\end{aligned} \tag{22}$$

Namely, the β series starts with $\beta^{\lambda+\mu}$. On the other hand, the recursion relations start with the inputs of $A_{10}\sim\beta/3$ and $A_{11}\sim\beta^2/8$. From the iterative procedure (a)–(c) we can easily find $A_{\lambda\mu}\sim O(\beta^{\lambda+\mu})$ at $\beta\to 0$. The recursion relation (i) then generates, at $\beta\to 0$,

$$\begin{aligned}A_{\lambda\mu}&\sim[(\lambda+\mu+1)/(\lambda+\mu+2)(\lambda+1)]\beta A_{\lambda-1,\mu}\\ &\sim[(\mu+2)/(\lambda+\mu+2)(\lambda+1)!]\beta^\lambda A_{0\mu}\\ &\sim[2/(\lambda+\mu+2)(\lambda+1)!(\mu+1)!]\beta^{\lambda+\mu},\end{aligned} \tag{23}$$

where the last line of this equation is given by setting $\mu=0$ in the equation of the second line and then by replacing λ by μ with the use of $A_{\mu 0}=A_{0\mu}$. Here we see that the recursive and direct integration methods yield the same behaviors at $\beta\to 0$. We have another confirmation to the recursion relations.

V. RELATION CONTAINING THE DERIVATIVE

The derivative of the general one-link partition function with respect to β can be written as follows:

$$\begin{aligned}\frac{d}{d\beta}&\int dU\,L_{km}(1)\exp[\mathrm{tr}(UJ+J^\dagger U^\dagger)]|_{J=\beta 1}\\ &=\int dU\,[L_{km}(1)\mathrm{tr}(UJ)+L_{km}(1)\mathrm{tr}(J^\dagger U^\dagger)]\\ &\quad\times\exp[\mathrm{tr}(UJ)+J^\dagger U^\dagger)]|_{J=\beta 1}\end{aligned} \tag{24}$$

where L_{km} is defined by Eq. (11). Setting all V_r unit matrices we find

$$\begin{aligned}\frac{d}{d\beta}B_{n-r,r}=&B_{n-r,r-1}+B_{n-r+1,r-1}+B_{n-r+1,r}\\ &+B_{n-r-1,r}+B_{n-r-1,r+1}+B_{n-r,r+1},\end{aligned} \tag{25}$$

with

$$B_{\lambda\mu}=d_{\lambda\mu}\,N(\beta)A_{\lambda\mu}(\beta), \tag{26}$$

where $B_{\lambda\mu}$ with $\lambda<0$ or $\mu<0$ is again assumed to be zero.

For U(1) and SU(2) the well-known relations for the modified Bessel function are deduced for $m = 0$ in Eq. (24):

$$\frac{d}{d\beta} I_k(2\beta) = I_{k-1}(2\beta) + I_{k+1}(2\beta), \qquad \text{for U(1)},$$
$$\frac{1}{2}\frac{d}{d\beta} I_{k+1}(4\beta) = I_k(4\beta) + I_{k+2}(4\beta), \qquad \text{for SU(2)}. \tag{27}$$

We see from these results that the derivatives for the U(1), and SU(2) functions cause the linear combination of two adjacent functions, whereas the derivatives for the SU(3) function causes the linear combination of six adjacent functions.

VI. DISCUSSIONS

We have given two independent recursion relations for the coefficient functions $A_{\lambda\mu}$ in the SU(3) character expansion. This reflects the fact that the SU(3) Young tableau is specified by two integers. It is quite advantageious to calculate the coefficients $A_{\lambda\mu}(\beta)$ with large values of λ and μ from the present recursion relations. We have also shown that the function $A_{\lambda\mu}(\beta)$ has large similarities to the modified Bessel function in that the characteristic relations are derived from the same Schwinger–Dyson equation of the multilink variable.

For the function $A_{\lambda\mu}(\beta)$, by combining the recursion relations with the derivative relation one may derive a linear differential equation as one derives the second-order differential equation for the Bessel function. This problem is quite interesting since the function $A_{\lambda\mu}$ is not only the coefficient of the character expansion, but it also may show a closed set of functions such as the modified Bessel functions.

[1]K. Wilson, Phys. Rev. D **10**, 2445 (1974).

[2]T. Eguchi, Phys. Lett. B **87**, 91 (1979); D. Foerster, Phys. Lett. B **87**, 87 (1979).

[3]R. C. Brower and M. Nauenberg, Nucl. Phys. B **180**, 221 (1981); R. C. Brower, P. Rossi, and C.-I. Tan, Phys. Rev. D **23**, 992, 953 (1981).

[4]S. S. Xue, T. C. Hsien, and C. W. Wu, Phys. Lett. B **175**, 341 (1986).

[5]W. Kerler and L. Schülke, Phys. Lett. B **201**, 123 (1988).

[6]I. Bars, J. Math. Phys. **21**, 2678 (1980).

[7]M. Creutz, *Quarks, Gluons and Lattice* (Cambridge U. P., Cambridge, 1983) p. 63; G. Münster and P. Weisz, Phys. Lett. B **96**, 119 (1980).

[8]K. E. Eriksson, N. Svartholm, and B. S. Skagerstam, J. Math. Phys. **22**, 2276 (1981).

[9]C. B. Lang, P. Salomonson, and B. S. Skagerstam, Phys. Lett. B **100**, 29 (1981).

[10]N. Tanimura, Phys. Lett. B **221**, 55 (1989).

Local symmetries and constraints

Joohan Lee and Robert M. Wald
Enrico Fermi Institute and Department of Physics, University of Chicago, 5640 S. Ellis Ave., Chicago, Illinois 60637

(Received 5 September 1989; accepted for publication 1 November 1989)

The general relationship between local symmetries occurring in a Lagrangian formulation of a field theory and the corresponding constraints present in a phase space formulation are studied. First, a prescription—applicable to an arbitrary Lagrangian field theory—for the construction of phase space from the manifold of field configurations on space-time is given. Next, a general definition of the notion of local symmetries on the manifold of field configurations is given that encompasses, as special cases, the usual gauge transformations of Yang–Mills theory and general relativity. Local symmetries on phase space are then defined via projection from field configuration space. It is proved that associated to each local symmetry which suitably projects to phase space is a corresponding equivalence class of constraint functions on phase space. Moreover, the constraints thereby obtained are always first class, and the Poisson bracket algebra of the constraint functions is isomorphic to the Lie bracket algebra of the local symmetries on the constraint submanifold of phase space. The differences that occur in the structure of constraints in Yang–Mills theory and general relativity are fully accounted for by the manner in which the local symmetries project to phase space: In Yang–Mills theory all the "field-independent" local symmetries project to all of phase space, whereas in general relativity the nonspatial diffeomorphisms do not project to all of phase space and the ones that suitably project to the constraint submanifold are "field dependent." As by-products of the present work, definitions are given of the symplectic potential current density and the symplectic current density in the context of an arbitrary Lagrangian field theory, and the Noether current density associated with an arbitrary local symmetry. A number of properties of these currents are established and some relationships between them are obtained.

I. INTRODUCTION

The gauge structures of Yang–Mills theory and general relativity are very similar when viewed from the perspective of Lagrangian field theory. In both cases, there is a group of local symmetries of the Lagrangian that acts on the manifold of field configurations, $\mathscr{F}$, on space-time M. In Yang–Mills theory, the field variable is a connection on a principal fiber bundle with group G over M, which may be represented locally as a Lie-algebra-valued one-form A_μ on M. The group of local symmetries, $\mathscr{G}_{\rm YM}$, consists of the usual gauge transformations, i.e., the set of maps from M into G. In general relativity, the field variable is a metric $g_{\mu\nu}$ on M, and the group of local symmetries $\mathscr{G}_{\rm GR}$ is the diffeomorphism group of M. In both cases, the action of the group of local symmetries gives $\mathscr{F}$ the natural structure of a principal fiber bundle.

Both Yang–Mills theory and general relativity also can be given a Hamiltonian formulation on a phase space. In both cases, there are constraints on phase space associated with the local symmetries on the field configuration manifold $\mathscr{F}$. However, when the structure of the constraints is examined carefully, the close analogy between Yang–Mills theory and general relativity appears to end.

In Yang–Mills theory one can define constraint functions on phase space (i.e., functions whose simultaneous vanishing defines the constraint submanifold) by

$$C_\Lambda = \int_\Sigma \mathrm{tr}(\Lambda \cdot \mathscr{D}_\mu E^\mu), \tag{1.1}$$

where Σ denotes the initial data surface, $\mathscr{D}_\mu$ denotes the gauge covariant derivative operator, E^μ is the electric field of A_μ [see Eq. (2.43)], and Λ is a map from Σ into the Lie algebra of G. Note that the set of such maps, Λ, is isomorphic to the Lie algebra of the factor group $\mathscr{G}'_{\rm YM} = \mathscr{G}_{\rm YM}/\mathscr{H}$, where $\mathscr{H}$ is the normal subgroup of $\mathscr{G}_{\rm YM}$ composed of the gauge transformations that act trivially on the initial data. The Poisson bracket algebra of the C_Λ's is naturally isomorphic to the Lie algebra of $\mathscr{G}'_{\rm YM}$, i.e., have

$$\{C_{\Lambda_1}, C_{\Lambda_2}\} = C_{[\Lambda_1,\Lambda_2]}. \tag{1.2}$$

In general relativity, the structure with respect to the spatial diffeomorphisms (i.e., the diffeomorphisms that map the initial data surface Σ into itself) is very similar to that occurring in Yang–Mills theory. Constraint functions for the spatial diffeomorphisms can be defined by

$$C_{\beta^\mu} = -2\int_\Sigma \beta_\mu h^{1/2} D_\nu (h^{-1/2}\pi^{\mu\nu}), \tag{1.3}$$

where β^μ is an arbitrary vector field on Σ, the tensor density $\pi^{\mu\nu}$ is given by Eq. (2.49), and D_μ is the covariant derivative operator associated with the spatial metric $h_{\mu\nu}$ on Σ. The Poisson bracket algebra of these constraint functions is naturally isomorphic to the Lie algebra of the diffeomorphism group of Σ, i.e., we have

$$\{C_{\beta_1^\mu}, C_{\beta_2^\mu}\} = C_{[\beta_1,\beta_2]^\mu}. \tag{1.4}$$

However, the situation for the nonspatial (i.e., "time translation") diffeomorphisms is quite different. Constraint func-

 0022-2488/90/030725-19$03.00

tions associated with these diffeomorphisms can be defined by

$$C_\alpha = \int_\Sigma \alpha h^{1/2}\Big[-{}^3R + h^{-1}\Big(\pi^{\mu\nu}\pi_{\mu\nu} - \frac{1}{2}\pi^2\Big)\Big], \qquad (1.5)$$

where α is an arbitrary function on Σ. However, although the Poisson bracket of any pair of constraint functions (1.3) and (1.5) is proportional to a constraint function (so that the constraints are "first class"), we now find that the proportionality factors are not constant on phase space. Rather, we obtain

$$\{C_{\beta^\mu}, C_\alpha\} = C_\gamma, \qquad (1.6)$$

$$\{C_{\alpha_1}, C_{\alpha_2}\} = C_{\gamma^\mu}, \qquad (1.7)$$

where

$$\gamma = \beta^\mu \partial_\mu \alpha, \qquad (1.8a)$$

$$\gamma^\mu = \alpha_1 h^{\mu\nu}\partial_\nu \alpha_2 - \alpha_2 h^{\mu\nu}\partial_\nu \alpha_1. \qquad (1.8b)$$

The spatial metric explicitly appears on the right-hand side of Eq. (1.8b). Hence γ^μ varies from point to point of phase space, and the Poisson bracket (1.7) yields different constraint functions at different points of phase space. Thus the canonical transformations generated by the constraints (1.3) and (1.5) do not correspond to a group action on phase space, and the constraints do not appear to reflect the structure of the local symmetries (i.e., the space-time diffeomorphisms) on field configuration space.

The above situation has been noted and studied by many authors (see, e.g., Isham and Kuchař,[1] and the references cited therein), particularly with respect to the difficulties that arise on account of the lack of Lie algebra structure when one attempts to apply the Dirac procedure for imposing the constraints (1.3) and (1.5) in the canonical quantization of general relativity. In this paper, however, we shall not be concerned primarily with these difficulties or their remedies, although a step toward a possible remedy will be suggested near the end of Sec. IV. Rather, our primary focus will be on developing the general theory of local symmetries and constraints in order to enable us to understand how such differences can arise. Specifically, we seek to answer the following questions: For an arbitrary Lagrangian field theory, under precisely which circumstances and in precisely what manner does the presence of local symmetries on the manifold of field configurations give rise to the presence of constraints on phase space? When such constraints do arise, what is the general relationship between the Poisson bracket algebra of the constraint functions and the Lie algebra of the local symmetries? We shall give complete answers to these questions in this paper, and these answers will enable us to account fully for the above differences that occur in Yang–Mils theory and general relativity.

The first major obstacle encountered in our analysis is caused by the fact that the local symmetries are defined on the manifold of field configurations, $\mathscr{F}$, whereas the constraints are defined on phase space Γ. Hence, in order to relate constraints to local symmetries, we must first relate Γ to $\mathscr{F}$. We overcome this obstacle in Sec. II by giving a general prescription—valid for an arbitrary Lagrangian field theory—for constructing Γ from $\mathscr{F}$. To do so, we give general definitions of a "symplectic potential current density" θ^μ and a "symplectic current density" ω^μ on space-time, and we establish a number of their properties. A "presymplectic form" ω_{AB} on $\mathscr{F}$ then is defined by integrating ω^μ over a Cauchy hypersurface. Phase space Γ with symplectic form Ω_{AB} is then obtained from $(\mathscr{F}, \omega_{AB})$ by a reduction procedure. For Yang–Mills theory and general relativity, this construction yields the usual phase space of these theories. For a parametrized scalar field theory, the construction yields a phase space equivalent to that of the "deparametrized" theory.

Section III is devoted to the study of local symmetries on the manifold of field configurations, $\mathscr{F}$. We give a general definition of the notion of local symmetries for an arbitrary Lagrangian field theory that encompasses the usual notions of local symmetries for Yang-Mills theory and general relativity as special cases. The Noether current density J^μ and Noether charge Q of a local symmetry are then introduced. The main result of this section is a theorem relating the variation of the Noether charge to the local symmetry field variation and the presymplectic form. An immediate corollary of the theorem is that the presymplectic form always is "gauge invariant" in a suitable sense. We thereby obtain a completely general proof of a result previously obtained for the particular cases of general relativity[2,3] and Yang-Mills theory.[3]

In Sec. IV, we combine the results and constructions of the previous sections to obtain our general relations between local symmetries and constraints. It is proved that to each local symmetry which suitably "projects" from solutions in $\mathscr{F}$ to phase space there exists a corresponding constraint. Furthermore, the constraints thereby obtained are always first class, and the Poisson bracket algebra of these constraints is always isomorphic to the Lie bracket algebra of the local symmetries. The differences occurring in the structure of the constraints between Yang–Mills theory and general relativity arise mainly from the following fact: In Yang-Mills theory all the "field-independent" local symmetries (i.e., the gauge transformations $\Lambda: M \mapsto G$, with Λ chosen to be independent of A_μ) suitably project to phase space, whereas in general relativity one must choose the nonspatial diffeomorphisms to be "field dependent" (i.e., dependent upon $g_{\mu\nu}$) in order to obtain a well defined projection. Consequently, although the nonspatial diffeomorphisms are fully represented on the constraint submanifold of the phase space of general relativity, the principal bundle structure of $\mathscr{F}$ arising from the "field-independent" local symmetries does not "project" to phase space. The relevance of considering "field-dependent" diffeomorphisms in analyzing the gauge structure of general relativity previously has played a prominent role in the work of Bergmann,[4] Bergmann and Komar,[5] and Salisbury and Sundermeyer.[6]

Our paper concludes with an appendix giving a brief discussion of Hamiltonian formulations of the general class of Lagrangian field theories considered here.

Finally, we comment briefly on the nature of the results of this paper. Essentially all of our analysis divides cleanly into one of the following two categories: (i) local constructions of quantities on space-time—such as the current densities θ^μ, ω^μ, and J^μ—and derivations of relationships between them; and (ii) constructions involving the manifold of field configurations, $\mathscr{F}$, and/or phase space Γ and properties of tensor fields defined on these infinite-dimensional manifolds. The results falling into category (i) (comprised by the first half of Sec. II and all of Sec. III) are completely rigorous. In particular, formulas involving the varied fields $\delta\phi^a$ are rigorous statements about partial derivatives of appropriate one-parameter families of field configurations. On the other hand, the results falling into category (ii) assume that a Banach manifold structure has been given to $\mathscr{F}$ and Γ. There is no difficulty in doing this (at least in typical theories), although the appropriate choice of manifold structure will depend upon the degree of differentiability and asymptotic conditions one wishes to impose on the fields. However, we have not shown that a manifold structure can be defined so that appropriate continuity and other properties are satisfied by the quantities obtained by our constructions, so that they rigorously define tensor fields of the indicated types. For example, we define by Eq. (2.24) below a functional θ on $\mathscr{F}$ that is linear in the varied field $\delta\phi$. However, in order that θ define a one-form at each point of $\mathscr{F}$ as we assume, the manifold structure must be chosen so that θ is a continuous (i.e., bounded) functional of $\delta\phi$. Since we have not attempted to treat technical issues of this nature, the results of this paper falling into category (ii) must be viewed as heuristic. Nevertheless, it should be noted that the tensor calculus we use involving the Lie derivative and exterior derivative is well defined on infinite-dimensional Banach manifolds.[7]

II. PHASE SPACE OF LAGRANGIAN FIELD THEORIES

In this section, we describe in detail a geometrical construction of phase space for Lagrangian field theories formulated on an n-dimensional space-time M of topology $\mathbb{R}\times\Sigma$. For simplicity, we shall assume that Σ (and hence M) is orientable. If the space-time metric $g_{\mu\nu}$ is part of the "background structure" (as in special relativistic theories), we assume that $(M, g_{\mu\nu})$ is globally hyperbolic with each Σ_t in the foliation of $\mathbb{R}\times\Sigma$ being a spacelike Cauchy surface. If $g_{\mu\nu}$ itself is a dynamical variable (as in general relativity), we simply restrict $g_{\mu\nu}$ to be such that each Σ_t is a spacelike Cauchy surface. In various places below, we will integrate total divergences over M. We then shall assume either that Σ is compact or that the fields satisfy asymptotic boundary conditions appropriate to ensure that no "spatial boundary terms" arise from applying Gauss' law to such integrations.

We shall assume that the field (or collection of fields) ϕ of our theory can be described as a map from space-time M into another finite-dimensional manifold M', i.e., $\phi: M \mapsto M'$. In some theories (e.g., for a real- or complex-valued scalar field), the field ϕ is initially presented in this manner. In other cases, it may be necessary to introduce some additional structure in order to so describe the field. For example, in Yang–Mills theory the field is a connection in a principal fiber bundle over space-time. However, by choosing a cross section of the bundle as well as a basis field of the cotangent space of M, we can locally express the Yang–Mills field as a map from M into $M' = L(G)\times\mathbb{R}^n$, where $L(G)$ denotes the Lie algebra of the Yang–Mills group G and $n = \dim(M)$. Thus, by describing the field as a map from space-time into M', we assume that any such additional structure has been introduced. We will verify below that for the Yang–Mills theory and general relativity, our construction of phase space is independent of the choices of cross sections and/or bases needed to so describe the field as a map between manifolds. Note that, locally on M, there should be no loss of generality in assuming that the field can be expressed as a map of space-time into M', i.e., more precisely, we would take this as the *definition* of a field theory on space-time. However, globally on M it may not be possible to express the field in this manner, as occurs, for example, in Yang–Mills theory based on a nontrivial principal bundle. Nevertheless, since our fundamental constructions of the current densities θ^μ and ω^μ given below are entirely local in nature, there are no problems with globalizing our results in such cases provided only that the integrands appearing on the right side of Eqs. (2.23) and (2.24) are independent of the allowed choices, as they are in Yang–Mills theory and general relativity. Thus there should be no essential loss of generality in assuming that the fields are described *globally* as a map ϕ: $M \mapsto M'$; for simplicity we shall assume this is the case.

For simplicity, we assume, further, that M' has been chosen so that *all* sufficiently smooth maps ϕ: $M \mapsto M'$ satisfying appropriate asymptotic conditions are "kinematically allowed" field configurations. Again, there should be no essential loss of generality in making this assumption. We denote by $\mathscr{F}$ the collection of all allowed field configurations on space-time.

In a sufficiently small neighborhood U' of any point $\phi_0 \in M'$, we may choose coordinates for M' such that the map ϕ can be represented locally as a collection of scalar functions, ϕ^a, of the space-time point x. (Here we use lowercase greek letters for indices referring to M and lowercase roman letters for indices on M'. On account of a shortage of alphabets, abstract index notation will *not* be used.) Note that a change of coordinates in U' corresponds to an x-independent field redefinition $\psi^a = f^a(\phi^b)$. We also choose a fixed derivative operator ∇_μ globally on M. (Below, we will introduce a fixed volume element, $\epsilon_{\alpha_1\cdots\alpha_n}$, on M and will then restrict ∇_μ to satisfy $\nabla_\mu \epsilon_{\alpha_1\cdots\alpha_n} = 0$.) We act with ∇_μ on ϕ^a by treating ϕ^a as a scalar function. (Indeed, the purpose of introducing coordinates on M' is to enable us to define a notion of second and higher derivatives of ϕ.) Our constructions below will make use of our choice of coordinates in M' and derivative operator on M. However, we will point out explicitly when the quantities we define are independent of any such additional structure we have introduced. As we shall see, the key quantities θ^μ and ω^μ obtained below are independent of the choice of coordinates in M', but may depend upon ∇_μ for sufficiently high derivative theories.

We assume that the field equations satisfied by ϕ are derived (in the manner to be specified below) by variation of an action S: $\mathscr{F} \mapsto \mathbb{R}$ of the form

$$S[\phi] = \int_M \mathscr{L}, \qquad (2.1)$$

where the Lagrangian $\mathscr{L}$ is a scalar density of weight 1 (see below) which locally has the form of a function of ϕ^a, its first k symmetrized derivatives, and, in addition, may also depend upon nondynamical background fields γ^b (such as the space-time metric in special relativistic theories),

$$\mathscr{L} = \mathscr{L}(\phi^a, \nabla_\mu \phi^a, ..., \nabla_{(\mu_1} \cdots \nabla_{\mu_k)} \phi^a; \gamma^b). \tag{2.2}$$

In other words, the value of the density $\mathscr{L}$ at a point $x \in M$ depends only on the value of the quantities appearing on the right side of Eq. (2.2) evaluated at point x. (There is no loss of generality in our assumption that $\mathscr{L}$ is a function only of totally symmetrized derivatives of ϕ^a, since any antisymmetric part of $\nabla_{\mu_1} \cdots \nabla_{\mu_j} \phi^a$ can be reexpressed in terms of the curvature tensor associated with ∇_μ and lower derivatives of ϕ^a.) Note that the statement that $\mathscr{L}$ depends only upon ϕ^a and its first k symmetrized derivatives is independent of the choice of derivative operator on M and coordinates in M'.

Since our analysis below will heavily involve unit weight scalar and vector densities, we take this opportunity to review their definition and properties and to explain our notational conventions. On an orientable manifold M of dimension n, a tensor density of type (k,l) and weight $s \in \mathbb{R}$ may be defined as an equivalence class of pairs $(T^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l}, \epsilon_{\alpha_1 \cdots \alpha_n})$, where $T^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l}$ is a tensor field of type (k,l), $\epsilon_{\alpha_1 \cdots \alpha_n} = \epsilon_{[\alpha_1 \cdots \alpha_n]}$ is a nonvanishing n-form (i.e., a volume element), and two such pairs (T,ϵ) and $(\tilde{T},\tilde{\epsilon})$ are said to be equivalent if

$$\tilde{T}^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l} = f^{-s} T^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l}, \tag{2.3}$$

where the function f is defined by

$$\tilde{\epsilon}_{\alpha_1 \cdots \alpha_n} = f \epsilon_{\alpha_1 \cdots \alpha_n}. \tag{2.4}$$

Normally, one proceeds by introducing a fixed volume element $\epsilon_{\alpha_1 \cdots \alpha_n}$ on space-time and representing a tensor density by the tensor field $T^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l}$, which is paired with $\epsilon_{\alpha_1 \cdots \alpha_n}$ in the equivalence class. However, for unit weight tensor densities (which is all that will be considered here) a much simpler description is available: For $s = 1$, we can represent a tensor density by the tensor field $T^{\mu_1 \cdots \mu_k}{}_{\nu_1 \cdots \nu_l} \epsilon_{\alpha_1 \cdots \alpha_n}$, of type $(k, l+n)$, which is antisymmetric in its last n lower indices. [This tensor field is independent of the choice of representative (T,ϵ) in the equivalence class.] Thus a unit weight scalar density, such as the Lagrangian density above, is equivalent to an n-form $\mathscr{L}_{\alpha_1 \cdots \alpha_n} = \mathscr{L}_{[\alpha_1 \cdots \alpha_n]}$. From this remark it is easily seen that the integral of the Lagrangian density over an oriented space-time M is well defined, without the need to specify additional structure on M. Similarly, a vector density of weight 1 may be represented by a tensor field $v^\mu{}_{\alpha_1 \cdots \alpha_n} = v^\mu{}_{[\alpha_1 \cdots \alpha_n]}$. By contracting the vector index with the first lowered index, we produce an $(n-1)$-form $v^\mu{}_{\mu \alpha_2 \cdots \alpha_n}$. Thus the integral of $v^\mu{}_{\mu\alpha_2 \cdots \alpha_n}$ over an oriented hypersurface Σ in M is well defined, without the need to specify any additional structure on Σ. Note that by Stokes' theorem, for any region $D \subset M$ that comprises a compact manifold with boundary, we have

$$n \int_D \nabla_{[\alpha_1} v^\mu{}_{|\mu| \alpha_2 \cdots \alpha_n]} = \int_{\partial D} v^\mu{}_{\mu \alpha_2 \cdots \alpha_n}, \tag{2.5}$$

where ∇_α is any derivative operator and $n = \dim M$.

Although the above viewpoint on unit weight tensor densities has considerable formulational advantages, it is notationally quite cumbersome to keep the n antisymmetric indices in formulas. Hence, mainly for notational convenience, we shall follow the usual practice of introducing a fixed volume element $\epsilon_{\alpha_1 \cdots \alpha_n}$ on space-time and representing a tensor density by the tensor in the equivalence class paired with $\epsilon_{\alpha_1 \cdots \alpha_n}$. (An additional reason for doing so is that to treat nonorientable space-times, this type of representation of tensor densities is necessary. [On a nonorientable space-time, a tensor density may be defined as an equivalence class of pairs (T, ϵ), where T again is a tensor field but now ϵ is a nonvanishing "n-form modulo sign." In the equivalence relation (2.3), f is taken to be the magnitude of the factor relating the two "n-forms modulo sign."] For simplicity, however, we consider only the orientable case here.) In this manner, the Lagrangian density will be represented by a scalar function $\mathscr{L}$, as was already done above. Similarly, a vector density will be represented by a vector field v^μ. Since our notation does not distinguish between tensors and tensor densities, we shall frequently remind the reader which quantities are densities (i.e., which tensors depend in the manner indicated above upon a choice of volume element). Note that in terms of the vector field representative v^μ of a vector density, Stokes' theorem (2.5) takes the Gauss' law form

$$\int_D \nabla_\mu v^\mu = \int_{\partial D} v^\mu n_\mu, \tag{2.6}$$

where ∇_μ is any derivative operator satisfying $\nabla_\mu \epsilon_{\alpha_1 \cdots \alpha_n} = 0$, and n_μ satisfies $n_\mu t^\mu = 1$ for an "outward pointing" vector field t^μ, whereas $n_\mu s^\mu = 0$ if s^μ is tangential to ∂D. [The volume element $\epsilon_{\alpha_1 \cdots \alpha_n}$ on M is understood on the left-hand side of (2.6), whereas the volume element $\epsilon_{\alpha_1 \cdots \alpha_n} t^{\alpha_1}$ on ∂D is understood on the right-hand side.] For convenience, we shall assume that the derivative operator ∇_μ introduced above [see Eq. (2.2)] has been chosen so that $\nabla_\mu \epsilon_{\alpha_1 \cdots \alpha_n} = 0$. Note that each side of Eq. (2.6) is equal to the corresponding side of Eq. (2.5). Thus we emphasize that both sides of Eq. (2.6) are well defined for vector densities, without the need to specify any additional structure on D or ∂D.

In the following, we will frequently encounter "local functions" (such as $\mathscr{L}$) of the fields, i.e., quantities whose value at x can be expressed as an ordinary function of the coordinates, $\phi^a(x)$, of the image of ϕ at x and of the symmetrized derivatives of ϕ^a evaluated at x. We also shall encounter "functionals" of the fields, i.e., quantities defined on field configuration space $\mathscr{F}$ whose value may depend nonlocally on ϕ. To help distinguish notationally between local functions of ϕ and functionals of ϕ, we will use parentheses to denote the arguments of local functions [see, e.g., Eq. (2.2) above] and brackets to denote the arguments of functionals [see, e.g., Eq. (2.23) below].

Finally, we introduce the following notation for the first partial derivatives of the function $\mathscr{L}$:

$$L_a \equiv \frac{\partial \mathscr{L}}{\partial \phi^a},$$
$$L_a{}^{\mu} \equiv \frac{\partial \mathscr{L}}{\partial (\nabla_\mu \phi^a)},$$
$$L_a{}^{\mu_1\mu_2} \equiv \frac{\partial \mathscr{L}}{\partial (\nabla_{(\mu_1}\nabla_{\mu_2)} \phi^a)}, \tag{2.7}$$
$$\vdots$$
$$L_a{}^{\mu_1\cdots\mu_k} \equiv \frac{\partial \mathscr{L}}{\partial (\nabla_{(\mu_1}\cdots\nabla_{\mu_k)} \phi^a)}.$$

In taking these partial derivatives (as well as in all field variations considered below), it is understood that any nondynamical background fields γ^a that may be present [see Eq. (2.2)] are held fixed. Note that each of the M tensor densities, M' covectors, $L_a{}^{\mu_1\cdots\mu_j}$, defined by Eq. (2.7) is totally symmetric in its space-time indices, $L_a{}^{\mu_1\cdots\mu_j} = L_a{}^{(\mu_1\cdots\mu_j)}$. However, with the exception of the highest derivative quantity $L_a{}^{\mu_1\cdots\mu_k}$, these tensor densities depend both upon the choice of derivative operator ∇_μ on M and coordinates on M'.

Consider, now, a smooth, one-parameter family, $\phi(\lambda)$: $M \mapsto M'$, of field configurations on space-time. We assume, initially, that $\phi(\lambda;x)$ is fixed (i.e., λ independent) for x outside of a compact set in M. The first variation of the Lagrangian density about the field configuration $\phi_0 = \phi(0)$ then takes the form

$$\delta\mathscr{L} \equiv \frac{d}{d\lambda}\mathscr{L}\Big|_{\lambda=0} = \sum_{j=0}^{k} L_a{}^{\mu_1\cdots\mu_j}\nabla_{(\mu_1}\cdots\nabla_{\mu_j)}\,\delta\phi^a, \tag{2.8}$$

where evaluation of $L_a{}^{\mu_1\ldots\mu_j}$ at ϕ_0 is understood, and where

$$\delta\phi^a(x) \equiv \frac{\partial\phi^a(\lambda;x)}{\partial\lambda}\Big|_{\lambda=0} \tag{2.9}$$

has compact support. Note that $\delta\phi^a(x)$ is the tangent to the curve $c(\lambda) = \phi(\lambda;x)$ (with x fixed) in M' at $\lambda = 0$, so $\delta\phi^a(x)$ may naturally be viewed as a vector in the tangent space to M' at the point $\phi_0(x)$. Thus $\delta\phi^a$ is an M'-vector-valued scalar field on M.

We may rewrite Eq. (2.8) as

$$\delta\mathscr{L} = E_a\,\delta\phi^a + \nabla_\mu\theta^\mu, \tag{2.10}$$

where

$$E_a = \sum_{j=0}^{k}(-1)^j\nabla_{\mu_1}\cdots\nabla_{\mu_j}L_a{}^{\mu_1\ldots\mu_j} \tag{2.11}$$

and

$$\theta^\mu = \sum_{j=1}^{k}\sum_{i=1}^{j}(-1)^{i+1}(\nabla_{\mu_2}\cdots\nabla_{\mu_i}L_a{}^{\mu\mu_2\ldots\mu_j}) \times (\nabla_{\mu_{i+1}}\cdots\nabla_{\mu_j}\,\delta\phi^a). \tag{2.12}$$

[In Eq. (2.12) it is to be understood that when $i = 1$, no derivatives act on $L_a{}^{\mu\mu_2\cdots\mu_j}$, and when $i = j$, no derivatives act on $\delta\phi^a$. Note that if we had chosen a derivative operator ∇_μ for which $\nabla_\mu\epsilon_{\alpha_1\cdots\alpha_n} \neq 0$, then Eqs. (2.10)–(2.12) would be modified.] We refer to θ^μ as the *symplectic potential current density*. Note that the dependence of θ^μ on $\delta\phi^a$ and its derivatives is linear.

Since $\mathscr{L}$—and hence $\delta\mathscr{L}$—is a scalar density on space-time and since $\delta\phi^a$—and hence θ^μ—has compact support in M, it follows that

$$\int_M E_a\,\delta\phi^a = \int_M \delta\mathscr{L}$$

is well defined, i.e., independent of the choices of volume element, $\epsilon_{\alpha_1\cdots\alpha_n}$, derivative operator ∇_μ in M satisfying $\nabla_\mu\epsilon_{\alpha_1\cdots\alpha_n} = 0$, and coordinates in M'. Since $\delta\phi^a$ is an arbitrary M'-vector-valued scalar field on M (subject only to being of compact support in M), it follows that the quantity E_a, which is a scalar density with respect to M and a dual vector with respect to M', must similarly be independent of these choices. It then follows immediately from Eq. (2.10) that the scalar density $\nabla_\mu\theta^\mu$ also is well defined. From the form of Eq. (2.12) and the fact that each $L_a{}^{\mu_1\cdots\mu_j}$ is totally symmetric in its space-time indices, it can be shown further than the vector density θ^μ is independent of the choice of coordinates in M'. However, in general, θ^μ will depend upon the choice of derivative operator ∇_μ on M; a different choice of derivative operator $\tilde{\nabla}_\mu$ (satisfying $\tilde{\nabla}_\mu\tilde{\epsilon}_{\alpha_1\cdots\alpha_n} = 0$ for some volume element $\tilde{\epsilon}_{\alpha_1\cdots\alpha_n}$ on M) will yield a vector density $\tilde{\theta}^\mu$ which, in general, differs from θ^μ by an identically conserved current density of the form $\nabla_\nu H^{\mu\nu}$, where $H^{\mu\nu} = H^{[\mu\nu]}$ is locally constructed from the derivative operators ∇_μ and $\tilde{\nabla}_\mu$, the background structure γ^a, and from ϕ^a, $\delta\phi^a$, and their derivatives. [Note that formula (2.12) for $\tilde{\theta}^\mu$ applies only when the volume element $\tilde{\epsilon}_{\alpha_1\cdots\alpha_n}$ compatible with $\tilde{\nabla}_\mu$ is used to express the Lagrangian density as a scalar function $\tilde{\mathscr{L}}$. To compare $\tilde{\theta}^\mu$ with θ^μ, we must multiply this expression for $\tilde{\theta}^\mu$ in terms of $\tilde{\mathscr{L}}$ and $\tilde{\nabla}_\mu$ by the factor f^{-1}, where f is given by Eq. (2.4).] Nevertheless, from the form of Eq. (2.12), it can be shown that θ^μ is independent of the choice of ∇_μ when $k < 3$. Thus, for theories in which the Lagrangian density does not contain derivatives higher than second order of the field variable, the symplectic potential current density θ^μ is independent of all extraneous structure introduced in our constructions above.

Note that the M-scalar-density, M'-convector E_a is a local function of ϕ^a and its derivatives up to order $2k$. Since E_a depends locally on ϕ^a and is independent of the choice of derivative operator on M, we may calculate it at any point $x \in M$ by choosing a local coordinate system in a neighborhood of x, taking $\epsilon_{\alpha_1\cdots\alpha_n}$ to be the coordinate volume element, ∇_μ to be the coordinate derivative operator ∂_μ, and then applying Eq. (2.11). Similarly, the M-vector-density, M'-scalar θ^μ is a local function of ϕ^a and its derivatives up to order $(2k-1)$ and of $\delta\phi^a$ and its derivatives up to order $(k-1)$. When $k < 3$ we also may calculate θ^μ by using Eq. (2.12) in a local coordinate patch. Finally, since both E_a and θ^μ depend locally on ϕ and $\delta\phi^a$, we now may drop the restriction that $\delta\phi^a$ have compact support in M, i.e., the quantities E_a and θ^μ given by Eqs. (2.11) and (2.12) continue to satisfy Eq. (2.10) and all the properties listed above when $\phi(\lambda)$: $M \mapsto M'$ is an arbitrary smooth one-parameter family.

From Eq. (2.10), it follows immediately that the action

S at field configuration ϕ_0: $M \mapsto M'$ will be stationary (i.e., $dS/d\lambda = 0$) for all variations, $\delta\phi^a$ of compact support if and only if

$$E_a = 0, \tag{2.13}$$

at ϕ_0. We take Eq. (2.13) as the equation of motion for the field ϕ.

It should be noted that our expression (2.12) for θ^μ changes when we modify the Lagrangian density $\mathscr{L}$ by the addition of a term of the form $\nabla_\mu v^\mu$, where vector density v^μ is a function of ϕ^a and finitely many of its derivatives. (Such a modification of $\mathscr{L}$, of course, has no effect upon E_a.) If v^μ is a function of ϕ^a only (i.e., if it does not depend upon derivatives of ϕ^a), then this change in $\mathscr{L}$ simply induces the change δv^μ in θ^μ. However, in general, the change in θ^μ determined by Eq. (2.12) will differ from δv^μ by an identically conserved vector density which is a local function of ϕ^a, $\delta\phi^a$, and finitely many of their derivatives. In particular, if v^μ is a local function of only ϕ^a and its first derivative $\nabla_\mu\phi^a$, i.e., if

$$v^\mu = v^\mu(\phi^a, \nabla_\nu\phi^a), \tag{2.14}$$

then the change in Lagrangian density, $\mathscr{L} \mapsto \mathscr{L} + \nabla_\mu v^\mu$, induces the following change in θ^μ:

$$\theta^\mu \mapsto \theta^\mu + \delta v^\mu + \frac{1}{2}\nabla_\nu\left\{\left[\frac{\partial v^\nu}{\partial(\nabla_\mu\phi^a)} - \frac{\partial v^\mu}{\partial(\nabla_\nu\phi^a)}\right]\delta\phi^a\right\}. \tag{2.15}$$

An important relation can be derived by considering the second-order variations in $\mathscr{L}$ resulting from an arbitrary, smooth two-parameter family $\phi(\lambda_1, \lambda_2)$ of field configurations. By Eq. (2.10) we have

$$\delta_2\mathscr{L} \equiv \frac{\partial\mathscr{L}}{\partial\lambda_2} = E_a\,\delta_2\phi^a + \nabla_\mu\theta_2^\mu, \tag{2.16}$$

where $\delta_2\phi^a = \partial\phi^a/\partial\lambda_2$ and θ_2^μ is given by Eq. (2.12) with $\delta\phi^a = \delta_2\phi^a$. Taking the derivative of Eq. (2.16) with respect to λ_1, we obtain

$$\delta_1\delta_2\mathscr{L} \equiv \frac{\partial^2\mathscr{L}}{\partial\lambda_1\,\partial\lambda_2} = (\delta_1 E_a)(\delta_2\phi^a) + E_a(\delta_1\delta_2\phi^a) + \nabla_\mu\,\delta_1\theta_2^\mu. \tag{2.17}$$

A similar expression holds, of course, for $\delta_2\delta_1\mathscr{L}$. Subtracting these expressions and using equality of mixed partial derivatives, we obtain

$$0 = (\delta_1 E_a)(\delta_2\phi^a) - (\delta_2 E_a)(\delta_1\phi^a) + \nabla_\mu\omega^\mu, \tag{2.18}$$

where

$$\omega^\mu = \delta_1\theta_2^\mu - \delta_2\theta_1^\mu = \sum_{j=1}^{k}\sum_{i=1}^{j}(-1)^{i+1}\{(\nabla_{\mu_2}\cdots\nabla_{\mu_i}\,\delta_1 L_a{}^{\mu\mu_2\cdots\mu_j}) \times (\nabla_{\mu_{i+1}}\cdots\nabla_{\mu_j}\,\delta_2\phi^a) - (1\leftrightarrow 2)\}. \tag{2.19}$$

(The variations $\delta L_a{}^{\mu_1\cdots\mu_j}$ can, of course, be expressed in terms of the second partial derivations of $\mathscr{L}$ and the field variation $\delta\phi^a$ and its space-time derivatives.) We refer to ω^μ as the *symplectic current density* associated with $\mathscr{L}$. It is a vector density with respect to M and a scalar with respect to M', and is a local function of ϕ^a, the varied fields $\delta_1\phi^a$, $\delta_2\phi^a$, and the space-time derivatives of these quantities. Note that the possible terms in ω^μ involving $\delta_1\delta_2\phi^a = \delta_2\delta_1\phi^a$ cancel, so ω^μ depends linearly both upon $\delta_1\phi^a$ and its derivatives, and upon $\delta_2\phi^a$ and its derivatives. It also is manifestly antisymmetric in $\delta_1\phi^a$ and $\delta_2\phi^a$. From the properties of θ^μ, it follows that ω^μ is independent of the choice of coordinates in M', but, in general, a change in the choice of derivative operator on M will change ω^μ by the addition of an identically conserved vector density locally constructed from ϕ^a, $\delta_1\phi^a$, $\delta_2\phi^a$, and finitely many of their derivatives. However, if $k < 3$, then ω^μ will remain unchanged. Note further that under the change in Lagrangian density $\mathscr{L} \mapsto \mathscr{L} + \nabla_\mu v^\mu$, ω^μ also will, in general, change by addition of an identically conserved vector density locally constructed from ϕ^a, $\delta_1\phi^a$, $\delta_2\phi^a$, and finitely many of their derivatives. For the case where v^μ is a function of only ϕ^a and $\nabla_\mu\phi^a$, the change in ω^μ is easily computed from Eq. (2.15). Note that if v^μ depends only on ϕ^a, then ω^μ remains unchanged. Finally, we point out that in the simple case where $\mathscr{L}$ depends only on ϕ^a and its first derivative (i.e., when $k = 1$), we have

$$\theta^\mu = L_a{}^\mu(\delta\phi^a) = \frac{\partial\mathscr{L}}{\partial(\nabla_\mu\phi^a)}\,\delta\phi^a \tag{2.20}$$

and hence ω^μ is explicitly given by

$$\omega^\mu = \frac{\partial^2\mathscr{L}}{\partial\phi^a\,\partial(\nabla_\mu\phi^b)}\,[\delta_1\phi^a\,\delta_2\phi^b - \delta_2\phi^a\,\delta_1\phi^b] + \frac{\partial^2\mathscr{L}}{\partial(\nabla_\nu\phi^a)\partial(\nabla_\mu\phi^b)} \times [(\nabla_\nu\,\delta_1\phi^a)\delta_2\phi^b - (\nabla_\nu\,\delta_2\phi^a)\delta_1\phi^b]. \tag{2.21}$$

Suppose, now, that $\phi(\lambda_1, \lambda_2)$ is a two-parameter family of solutions of the equations of motion. Then $E_a = 0$, for all λ_1, λ_2, so, in particular, $\delta_1 E_a = \delta_2 E_a = 0$. Hence, by Eq. (2.18), we obtain

$$\nabla_\mu\omega^\mu = 0. \tag{2.22}$$

In summary, we have shown that for any Lagrangian field theory, a symplectic current density ω^μ can be constructed from $\mathscr{L}$ via Eq. (2.19). This current ω^μ is a local function of a "background field" ϕ^a, two "linearized perturbations" $\delta_1\phi^a$ and $\delta_2\phi^a$, and a finite number of their derivatives. It is independent of the choice of coordinates in M' and, for $k < 3$, also is independent of the choice of derivative operator on M. It satisfies the property that when ϕ is a solution to the field equations and $\delta_1\phi^a$ and $\delta_2\phi^a$ solve the linearized field equations, then ω^μ is conserved.

We can construct from ω^μ a real-valued functional of the field variables, denoted $\omega[\phi^a, \delta_1\phi^a, \delta_2\phi^a]$, as follows. Choose a Cauchy surface Σ in the foliation of M (see the beginning of this section) and define

$$\omega[\phi^a, \delta_1\phi^a, \delta_2\phi^a] = \int_\Sigma \omega^\mu n_\mu. \tag{2.23}$$

(As discussed above [see Eq. (2.6)] this integral on the right side is naturally defined, without the need to specify a volume element on Σ.) If Σ is compact, then ω is independent of the choice of derivative operator on M (even when $k\geqslant 3$) and also remains unchanged under the change in Lagrangian density $\mathscr{L}\mapsto\mathscr{L}\mapsto\nabla_\mu v^\mu$. In the noncompact case, such changes in ∇_μ (when $k\geqslant 3$) or $\mathscr{L}$ may change ω by a "surface term at infinity." In general, ω will depend upon the choice of Σ. However, if ϕ^a, $\delta_1\phi^a$, and $\delta_2\phi^a$ are solutions, then ω will be independent of this choice provided that either Σ is compact or the fields satisfy asymptotic conditions appropriate to ensure that no "spatial boundary terms" arise from applying Gauss' law to Eq. (2.22). In the following, we shall assume that such asymptotic conditions have been imposed upon the fields in the case where Σ is noncompact.

The above definition of ω can be described in a much more geometrical manner, in terms of which our construction of phase space will be given. We assume that the collection $\mathscr{F}$ of field configurations on space-time (i.e., the set of all sufficiently smooth maps ϕ: $M\mapsto M'$ satisfying appropriate asymptotic conditions) has been given the structure of an infinite-dimensional Banach manifold. There is no difficulty in doing this (at least in typical cases), but we have not attempted to show that it can be done in such a manner that the continuity and other properties assumed below for various quantities will hold. Thus, as already mentioned at the end of the Introduction, the discussion we are about to give must be viewed as heuristic. (This contrasts with the preceding discussion, which was entirely rigorous.)

A field configuration ϕ on space-time is represented as a point of $\mathscr{F}$. Similarly, a field variation $\delta\phi^a$ on space-time about the field configuration ϕ [see Eq. (2.9)] may be viewed as a vector in the tangent space to $\mathscr{F}$ at point ϕ. To emphasize this viewpoint in our notation, we will adopt an abstract index notation for tensor fields on $\mathscr{F}$, using capital roman letters. Thus we will write $(\delta\phi)^A$ when we view the field variation in this manner. [By contrast, we shall continue to write $\delta\phi^a(x)$ to denote the tangent vector at point $\phi(x)\in M'$ defined by Eq. (2.9) above.] From the vector density θ^μ on space-time defined by Eq. (2.12) we can define a functional $\theta\,[\phi^a,\delta\phi^a]$ by integration over a Cauchy surface Σ,

$$\theta\,[\phi^a,\delta\phi^a]\equiv\int_\Sigma \theta^\mu n_\mu. \tag{2.24}$$

(This functional depends, of course, upon the choice of Σ.) We may view θ as a function on the tangent bundle of $\mathscr{F}$. However, since θ^μ is linear in $\delta\phi^a$ and its derivatives, it follows that θ is a linear function of $(\delta\phi)^A$. We shall assume that the manifold structure of $\mathscr{F}$ is such that θ is continuous in $(\delta\phi)^A$ (which is *not* a consequence of linearity in infinite dimensions) so that θ defines a dual vector field on $\mathscr{F}$, which we denote as θ_A, given by

$$\theta_A(\delta\phi)^A=\theta\,[\phi^a,\delta\phi^a], \tag{2.25}$$

for all $(\delta\phi)^A$. Similarly, since the functional ω defined by Eqs. (2.23) is linear and antisymmetric in $(\delta_1\phi)^A$ and $(\delta_2\phi)^A$, we assume that it defines a two-form $\omega_{AB}=\omega_{[AB]}$ on $\mathscr{F}$. Furthermore, it is easily seen that Eq. (2.19) corresponds to the relation

$$\omega_{AB}=(d\theta)_{AB}, \tag{2.26}$$

where d denotes the exterior derivative on $\mathscr{F}$. Thus our construction of ω given above may be viewed as producing as exact (and hence closed) two-form ω_{AB} on $\mathscr{F}$, which we refer to as a "presymplectic form."

The two-form ω_{AB} fails to be a symplectic form on $\mathscr{F}$ because it is degenerate; equivalently, $\mathscr{F}$ itself is unsuitable to serve as phase space because it is "too large." In particular, note that any field variation $\delta\phi^a$ with support away from Σ gives rise to a degeneracy direction $(\delta\phi)^A$ for ω_{AB}. However, these difficulties can be cured by the "reduction procedure" of taking the "symplectic quotient" of $(\mathscr{F},\omega)$ (see, e.g., Ref. 8), thereby producing a manifold Γ on which there is defined a nondegenerate closed two-form Ω_{AB}. This symplectic manifold will serve as our phase space. We outline, now, the steps (and assumptions) needed to construct this phase space.

By a standard identity (which holds on infinite-dimensional Banach manifolds[7]), for any vector field ψ^A on $\mathscr{F}$, we have

$$\pounds_\psi\omega_{AB}=\psi^C(d\omega)_{CAB}+(d\lambda)_{AB}, \tag{2.27}$$

where

$$\lambda_A=\omega_{BA}\psi^B. \tag{2.28}$$

The first term of Eq. (2.27) vanishes since ω_{AB} is closed. If ψ^A is a degeneracy vector field, i.e., if $\omega_{AB}\psi^A=0$ at all points of $\mathscr{F}$, we have $\lambda_A=0$ and thus we find that

$$\pounds_\psi\omega_{AB}=0. \tag{2.29}$$

Furthermore, if ψ^A is the commutator of two degeneracy vector fields, i.e., if

$$\psi^A=[\psi_1,\psi_2]^A=(\pounds_{\psi_1}\psi_2)^A, \tag{2.30}$$

with $\omega_{AB}\psi_1^A=\omega_{AB}\psi_2^A=0$, we have

$$\omega_{AB}\psi^A=\omega_{AB}\pounds_{\psi_1}\psi_2^A=\pounds_{\psi_1}(\omega_{AB}\psi_2^A)-\psi_2^A\pounds_{\psi_1}\omega_{AB}=0, \tag{2.31}$$

so ψ^A also is a degeneracy vector field.

Consider, now, the distribution of degeneracy subspaces, i.e., the collection of subspaces of the tangent space to $\mathscr{F}$ consisting of the degeneracy vectors of ω_{AB}. We assume that this distribution comprises a sub-bundle of the tangent bundle of $\mathscr{F}$. Since by Eq. (2.31) the commutator of two vector fields lying in this distribution also lies in this distribution, by Frobenius' theorem[7] the distribution admits integral submanifolds. Hence we may define an equivalence relation on $\mathscr{F}$ by setting $\phi_1\cong\phi_2$ if and only if ϕ_1 and ϕ_2 lie on the same integral submanifold. Let Γ denote the set of equivalence classes of $\mathscr{F}$ and let π: $\mathscr{F}\mapsto\Gamma$ denote the map that assigns each element of $\mathscr{F}$ to its equivalence class. We assume that a manifold structure can be defined on Γ such that $\mathscr{F}$ has the structure of a fiber bundle over Γ with projection map π. (This need not automatically be the case since, in particular, the integral surfaces need not be embedded submanifolds, i.e., they could "wind around in $\mathscr{F}$" like lines with irrational slope on the unit torus). We will use the same index notation (with capital roman letters) for tensors on Γ as for tensors on $\mathscr{F}$; no confusion should result since it should be clear on which manifold the various tensors are defined. We define a two-form Ω_{AB} on Γ by the condition that, for all vectors R^A, and S^A in the tangent space to each point $\xi \in \Gamma$, we have

$$\Omega_{AB} R^A S^B = \omega_{AB} \rho^A \sigma^B, \tag{2.32}$$

where ρ^A and σ^A are any tangent vectors at any point $\phi \in \mathscr{F}$ in the fiber over ξ such that $\pi^* \rho^A = R^A$, $\pi^* \sigma^A = S^A$, where π^* denotes the natural map from the tangent space V_ϕ of ϕ to the tangent space V_ξ of ξ induced by the map π. (We also shall use the same notation, π^*, to denote the natural pullback map taking differential forms on Γ to differential forms on $\mathscr{F}$.) On account of the degeneracy of ω_{AB} in the fiber directions, the right side of Eq. (2.32) does not depend upon the choice of "representative vectors" ρ^A and σ^A at ϕ. Furthermore, on account of Eq. (2.29), the right side of Eq. (2.32) also does not depend upon the choice of point $\phi \in \mathscr{F}$ in the fiber over ξ. Thus Ω_{AB} is a well defined two-form on Γ, which is related to ω_{AB} by

$$\omega_{AB} = \pi^* \Omega_{AB}, \tag{2.33}$$

i.e., ω_{AB} is the pullback of Ω_{AB} under the projection map π. Furthermore, Ω_{AB} is closed (as a direct consequence of the fact that ω_{AB} is closed) and, by construction, is nondegenerate. (Note, however, that Ω_{AB} need not be exact even though ω_{AB} satisfies this property). Thus (Γ, Ω_{AB}) is a symplectic manifold, which we take to be the phase space of our theory. Note that we have not provided Γ with a "polarization," i.e., we have not distinguished between configuration and momentum variables, e.g., by expressing Γ as a cotangent bundle of a configuration space. However, we will have no need for such a polarization in the analysis given below. Note also that the definition of ω depends upon a choice of Cauchy surface Σ [see Eq. (2.23)], and hence the construction of (Γ, Ω_{AB}) also depends upon such a choice.

Let $\bar{\mathscr{F}}$ denote the subset of $\mathscr{F}$ consisting of solutions to the equations of motion (2.13). We shall assume that $\bar{\mathscr{F}}$ is a submanifold of $\mathscr{F}$. Similarly, the image $\bar{\Gamma} \equiv \pi[\bar{\mathscr{F}}]$ of $\bar{\mathscr{F}}$ under the projection map π will be assumed to be a submanifold of Γ, which we shall refer to as the *constraint submanifold*. If we interpret Γ as representing the "kinematically possible" instantaneous states of the system, then $\bar{\Gamma}$ consists of those states that are "dynamically possible." If $\bar{\Gamma}$ is a proper subset of Γ, then not all kinematically possible states are dynamically possible, i.e., constraints are present. We denote by $\bar{\pi}$ the restriction of π to $\bar{\mathscr{F}}$. We assume, further, that $\bar{\mathscr{F}}$ has the structure of a fiber bundle over $\bar{\Gamma}$, with projection map $\bar{\pi}: \bar{\mathscr{F}} \mapsto \bar{\Gamma}$. Note that we then have

$$\bar{\omega}_{AB} = \bar{\pi}^* \bar{\Omega}_{AB}, \tag{2.34}$$

where $\bar{\omega}_{AB}$ denotes the pullback of ω_{AB} to $\bar{\mathscr{F}}$, and $\bar{\Omega}_{AB}$ denotes the pullback of Ω_{AB} to $\bar{\Gamma}$. Note also that although, as mentioned above, the construction of (Γ, Ω_{AB}) depends upon a choice of Cauchy surface Σ on space-time, the construction of $(\bar{\Gamma}, \bar{\Omega}_{AB})$ does not by virtue of the remark below Eq. (2.23). The relationships between $\mathscr{F}$, Γ, $\bar{\mathscr{F}}$, and $\bar{\Gamma}$ are illustrated in Fig. 1.

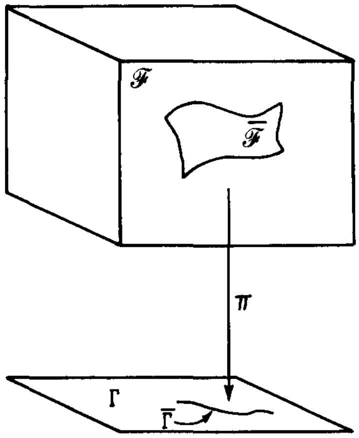

FIG. 1. The relationship between the manifold of field configurations $\mathscr{F}$ and phase space Γ. The reduction procedure described in the text produces from the "presymplectic" manifold $(\mathscr{F}, \omega_{AB})$ the symplectic manifold (Γ, Ω_{AB}), together with a projection map π: $\mathscr{F} \to \Gamma$. The submanifold of $\mathscr{F}$ comprised by the solutions to the field equations is denoted as $\bar{\mathscr{F}}$, and its image under π defines the constraint submanifold of phase space, denoted $\bar{\Gamma}$. The restriction of π to $\bar{\mathscr{F}}$ defines the projection map $\bar{\pi}$: $\bar{\mathscr{F}} \to \bar{\Gamma}$.

Several examples should help elucidate the above construction of phase space. Since we are primarily interested in theories with local symmetries, we focus attention on the three primary examples of such theories: Yang–Mills theory, general relativity, and a parametrized scalar field theory.

In Yang–Mills theory in a fixed, curved background space-time $(M, g_{\mu\nu})$ based on a semisimple Lie group G, the field variable is a connection on a principle fiber bundle over M with structure group G. As already mentioned at the beginning of the section, by choosing a cross section of this bundle and a basis of the cotangent space of M, we can (at least locally) describe this connection as a map $A^i{}_\mu$ from M into $M' = L(G) \times \mathbb{R}^n$, where $L(G)$ is the Lie algebra of G. Note that *both* the Lie algebra index i and the cotangent basis index μ correspond to the index "a" in our general treatment above. Since our general construction above calls for ϕ^a to be treated as a scalar function when acted upon by ∇_μ, we will avoid some potential confusion by introducing a local coordinate system in M, choosing the basis of the cotangent space of M to coincide with the coordinate basis of this system, and choosing the fixed volume element and derivative operator to be the coordinate volume element $\epsilon_{\alpha_1 \cdots \alpha_n}$ and coordinate derivative operator ∂_μ. [By doing so, the usual meaning of "$\partial_\mu A^i{}_\nu$" coincides with our usage; by contrast, for, e.g., the metric compatible derivative operator, the usual meaning of $\nabla_\mu A^i{}_\nu$ differs from what one would obtain by treating each component $A^i{}_\nu(x)$ as a scalar function.] The Lagrangian density $\mathscr{L}$ is taken to be

$$\mathscr{L} = -\tfrac{1}{4}\sqrt{-g}\, F^i{}_{\mu\nu} F_i{}^{\mu\nu}, \tag{2.35}$$

where

$$F^i{}_{\mu\nu} = 2\, \partial_{[\mu} A^i{}_{\nu]} + c^i{}_{jk} A^j{}_\mu A^k{}_\nu, \tag{2.36}$$

where $c^i{}_{jk}$ is the structure tensor of the Lie algebra and the Lie algebra indices are lowered and raised with the Killing metric $-c^k{}_{il} c^l{}_{jk}$. Since $\mathscr{L}$ does not depend upon second and higher derivatives of $A^i{}_\mu$, the quantities θ^μ and ω^μ do not depend upon our choice of derivative operator ∂_μ, and the simplified formulas (2.20) and (2.21) apply. Thus we obtain

$$\theta^\mu = -\sqrt{-g}\, F_i{}^{\mu\nu}\, \delta A^i{}_\nu, \tag{2.37}$$

$$\omega^\mu = \sqrt{-g}\, [\, -(\delta_1 F_i{}^{\mu\nu})(\delta_2 A^i{}_\nu) + (\delta_2 F_i{}^{\mu\nu})(\delta_1 A^i{}_\nu)\,]. \tag{2.38}$$

Now, under an infinitesimal change in the choice of local cross section (used to express the connection as a map between manifolds), we have

$$A^i{}_\mu \to A^i{}_\mu + \partial_\mu \xi^i + c^i_{jk} A^j{}_\mu \xi^k, \tag{2.39}$$

$$F^i{}_{\mu\nu} \to F^i{}_{\mu\nu} + c^i_{jk} F^j{}_{\mu\nu} \xi^k. \tag{2.40}$$

Hence the variation of $A^i{}_\mu$ is changed by

$$\delta A^i{}_\mu \to \delta A^i{}_\mu + c^i_{jk}(\delta A^i{}_\mu)\xi^k. \tag{2.41}$$

It follows from Eqs. (2.40) and (2.41) that both θ^μ and ω^μ remain unchanged under a change of local cross section. (Note, however, that θ^μ and ω^μ are *not* gauge invariant in the sense that if $\delta A^i{}_\mu$ takes the form of an infinitesimal gauge transformation, then θ^μ and ω^μ do not vanish. Nevertheless, in the next section, we will show, quite generally, that the pullback $\bar{\omega}_{AB}$ of ω_{AB} to the solution submanifold $\bar{\mathscr{F}}$ always is gauge invariant in this sense.) It also is manifest from the expressions (2.37) and (2.38) that the vector densities θ^μ and ω^μ do not depend upon the choice of coordinate basis of the contagent space in M. Thus θ^μ and ω^μ are well defined, i.e., they do not depend upon the additional structure introduced to express the Yang–Mills field as a map between manifolds.

By Eqs. (2.23) and (2.38), we have

$$\omega_{AB}(\delta_1 A)^A(\delta_2 A)^B = \int_\Sigma \sqrt{-g}\,[\,-(\delta_1 F_i{}^{\mu\nu})(\delta_2 A^i{}_\nu) + (\delta_2 F_i{}^{\mu\nu})(\delta_1 A^i{}_\nu)\,]\, n_\mu. \tag{2.42}$$

From this equation, it is manifest that the degeneracy directions of ω_{AB} consist of those field variations for which the spatial projection (i.e., pullback) of $\delta A^i{}_\mu$ to Σ and

$$\delta E_i{}^\mu \equiv -\sqrt{-g}\,(\delta F_i{}^{\mu\nu}) n_\nu \tag{2.43}$$

both vanish on Σ. Hence each integral submanifold of degeneracy directions consists of all field configurations that have on Σ the same $E_i{}^\mu$ and pullback of $A^i{}_\mu$. Thus the phase space Γ can be identified with such $(A^i{}_\mu, E_i{}^\mu)$ pairs, with symplectic form Ω_{AB} determined by Eq. (2.42). Thus (Γ, Ω_{AB}) corresponds precisely to the usual phase space of Yang–Mills theory. Note that the constraint submanifold $\bar{\Gamma}$ consists of the $(A^i{}_\mu, E_i{}^\mu)$ pairs that satisfy

$$\partial_\mu E^{i\mu} + c^i_{jk} A^j{}_\mu E^{k\mu} = 0. \tag{2.44}$$

As a second example, consider general relativity. The field variable here is the space-time metric. By introducing a basis of the cotangent space, we can (at least locally) view the metric as a map $g_{\mu\nu}$ from M into the vector space of symmetric $n \times n$ matrices. (Here the index pair "$\mu\nu$" corresponds to the index "a" in the general discussion above.) Again we introduce a local coordinate system and choose the cotangent space basis to coincide with the coordinate basis of our local coordinate system. We also again employ the coordinate volume element and derivative operator ∂_μ. For convenience, we choose one of our coordinates to be "t" (i.e., the time function appearing in our foliation of M by Cauchy surfaces Σ_t) and by a field redefinition (i.e., a coordinate transformation in M') take our field variables to be the spatial metric components $h_{\mu\nu}$, the shift vector $N_\mu = h_{\mu\nu}(\partial/\partial t)^\nu$, and the lapse function $N = (-g^{\mu\nu}\partial_\mu t\, \partial_\nu t)^{1/2}$. The usual Hilbert Lagrangian density, $\mathscr{L}_{\rm H} = \sqrt{-g}R$, for general relativity leads, via Eq. (2.19), to the formula for ω^μ given by Crnković and Witten,[3] which differs from the expression originally given by Friedman[2] by an identically conserved term of the form $\partial_\nu(\sqrt{-g}\,\delta_1 g^{\alpha[\mu}\delta_2 g^{\nu]\beta} g_{\alpha\beta})$. We obtain thereby relatively complicated expressions for the integrands defining θ_A and ω_{AB}. To simplify the situation, we choose instead the Lagrangian density

$$\mathscr{L} = N\sqrt{h}\,[\,{}^3R + K_{\mu\nu}K^{\mu\nu} - K^2\,], \tag{2.45}$$

where ${}^3R$ is the scalar curvature of $h_{\mu\nu}$ and $K_{\mu\nu}$ is the extrinsic curvature of Σ given in terms of $h_{\mu\nu}$, N, and N_μ by

$$K_{\mu\nu} = \frac{1}{2} N^{-1}\left[\frac{\partial h_{\mu\nu}}{\partial t} - 2D_{(\mu}N_{\nu)}\right], \tag{2.46}$$

where D_μ is the derivative operator on Σ associated with $h_{\mu\nu}$. This Lagrangian density differs from $\mathscr{L}_{\rm H}$ by addition of the divergence of a vector density v^μ which depends only on derivatives of the metric up to first order (see p. 464 of Ref. 9). Thus $\mathscr{L}_{\rm H}$ and $\mathscr{L}$ yield the same equations of motion. Note that by our general discussion above, the symplectic current ω^μ obtained from $\mathscr{L}$ differs from the current $\omega^\mu_{\rm H}$ obtained from $\mathscr{L}_{\rm H}$ by an identically conserved vector density locally constructed from the fields and their variations which can be computed using Eq. (2.15). Thus if Σ is compact, $\mathscr{L}$ and $\mathscr{L}_{\rm H}$ produce the same presymplectic form ω_{AB} on Γ; however, in the noncompact case, the two expressions for the presymplectic form may differ by a "surface term at spatial infinity." Further discussion of the relationship between ω^μ and $\omega^\mu_{\rm H}$ will be given elsewhere, and a generalization of $\omega^\mu_{\rm H}$ to the Einstein–Maxwell case also will be used to derive convenient expressions for conserved fluxes of gravitational and electromagnetic radiation.[10]

Since $\mathscr{L}$ has no dependence on second-order time–time or mixed time–space derivatives (or any higher-order derivatives) of $g_{\mu\nu}$, formulas (2.20) and (2.21) apply for the relevant components $\theta^\mu n_\mu$ and $\omega^\mu n_\mu$ (where $n_\mu = \nabla_\mu t$) of the current densities θ^μ and ω^μ. We obtain

$$\theta^\mu n_\mu = \pi^{\mu\nu}\,\delta h_{\mu\nu} \tag{2.47}$$

and

$$\omega^\mu n_\mu = (\delta_1 \pi^{\mu\nu})(\delta_2 h_{\mu\nu}) - (\delta_2 \pi^{\mu\nu})(\delta_1 h_{\mu\nu}), \tag{2.48}$$

where

$$\pi^{\mu\nu} = \sqrt{h}\,(K^{\mu\nu} - K h^{\mu\nu}). \tag{2.49}$$

Equations (2.47) and (2.48) define scalar densities on a hypersurface Σ that are independent of the choice of coordinates used in the construction. Note, however, that, as in Yang–Mills case, $\theta^{\mu}n_{\mu}$ and $\omega^{\mu}n_{\mu}$ fail to be gauge invariant in the sense that they do not vanish identically if $\delta g_{\mu\nu}$ is a gauge variation $\nabla_{(\mu}\xi_{\nu)}$. Nevertheless, as already mentioned above, it follows from the theorem proved at the end of the next section that such a gauge variation always is a degeneracy direction of the pullback $\overline{\omega}_{AB}$ of ω_{AB} to $\overline{\mathscr{F}}$.

From Eq. (2.48), we obtain

$$\omega_{AB}(\delta_1 g)^A(\delta_2 g)^B = \int_{\Sigma} [(\delta_1\pi^{\mu\nu})(\delta_2 h_{\mu\nu}) - (\delta_2\pi^{\mu\nu})(\delta_1 h_{\mu\nu})]. \quad (2.50)$$

Thus the degeneracy directions of ω_{AB} consist of those field variations $\delta g_{\mu\nu}$ for which $\delta h_{\mu\nu}$ and $\delta\pi^{\mu\nu}$ vanish on Σ. Consequently, we may identify phase space Γ with the pairs $(h_{\mu\nu},\pi^{\mu\nu})$ on Σ, with symplectic form Ω_{AB} determined by Eq. (2.50). This corresponds precisely to the usual choice of phase space for general relativity. Note that the constraint submanifold $\overline{\Gamma}$ consists of those pairs that satisfy

$$-{}^3R + h^{-1}(\pi^{\mu\nu}\pi_{\mu\nu} - \tfrac{1}{2}\pi^2) = 0, \quad (2.51)$$

$$D_\mu(h^{-1/2}\pi^{\mu\nu}) = 0, \quad (2.52)$$

where ${}^3R$ denotes the scalar curvature of $h_{\mu\nu}$.

As our last example we consider the parametrized massless, Klein–Gordon, scalar field theory, which is a generally covariant version of the ordinary scalar field theory in a background space-time. In this theory, the field variables are a real scalar field ψ on M, and a diffeomorphism y from M to another copy of the same manifold, which will be denoted $\widetilde{M}$. [Thus the field variable ϕ in the general discussion above corresponds to the pair (ψ,y) and we have $M' = \mathbb{R}\times\widetilde{M}$.] The manifold $\widetilde{M}$, in which the field y takes its value, is equipped with a fixed, nondynamical metric $\mathring{g}_{ab}$ and, introducing a coordinate volume element and derivative operator as before, the Lagrangian density of this theory, $\mathscr{L}$, is given by

$$\mathscr{L} = -\tfrac{1}{2}\sqrt{-g(y)}g^{\mu\nu}(y)\partial_\mu\psi\,\partial_\nu\psi, \quad (2.53)$$

where the metric $g_{\mu\nu}(y)$ denotes the pullback of the background metric $\mathring{g}_{ab}$ by the map y, i.e.,

$$g_{\mu\nu} = (y^*)^a{}_\mu (y^*)^b{}_\nu \mathring{g}_{ab}, \quad (2.54)$$

where $(y^*)^a{}_\mu$ denotes the induced map from the tangent space of $x\in M$ to the tangent space of $y(x)\in\widetilde{M}$ [or, equivalently, the pullback map from the cotangent space of $y(x)$ to the cotangent space of x].

From Eq. (2.11), the ψ component of the equations of motion is

$$0 = E_\psi = -\partial_\mu P^\mu, \quad (2.55)$$

where

$$P^\mu = -\sqrt{-g}g^{\mu\nu}\partial_\nu\psi. \quad (2.56)$$

This is the usual equation of motion for a massless, Klein–Gordon scalar field in the space-time $(M,g_{\mu\nu})$. The y^a component of the equations of motion is

$$0 = E_a = -(y_*)^\nu{}_a\sqrt{-g}\nabla_\mu T^\mu{}_\nu, \quad (2.57)$$

where $(y_*)^\nu{}_a \equiv (y^{-1*})^\nu{}_a$ is the inverse of $(y^*)^a{}_\mu$, ∇_μ is the derivative operator associated with $g_{\mu\nu}$, and $T_{\mu\nu}$ is the ordinary stress–energy tensor of ψ in the space-time $(M,g_{\mu\nu})$, i.e.,

$$T_{\mu\nu} = \partial_\mu\psi\,\partial_\nu\psi - \tfrac{1}{2}g_{\mu\nu}(y)g^{\alpha\beta}(y)\partial_\alpha\psi\,\partial_\beta\psi. \quad (2.58)$$

Since conservation of stress–energy, $\nabla_\mu T^\mu{}_\nu = 0$, is a consequence of the field equation for ψ, we see that Eq. (2.57) automatically is satisfied whenever Eq. (2.55) holds.

We proceed, now, to construct the phase space. Again, note that the Lagrangian density (2.53) does not depend upon second or higher derivatives of either ψ or y. Therefore, one can apply the formula (2.20) to obtain

$$\theta^\mu = P^\mu\,\delta\psi + (\sqrt{-g}T^\mu{}_\nu)(y_*)^\nu{}_a\,\delta y^a. \quad (2.59)$$

From this one obtains

$$\omega^\mu = \delta_1 P^\mu\,\delta_2\psi - \delta_2 P^\mu\,\delta_1\psi + \delta_1(\sqrt{-g}T^\mu{}_a)\delta_2 y^a - \delta_2(\sqrt{-g}T^\mu{}_a)\delta_1 y^a, \quad (2.60)$$

where

$$T^\mu{}_a = T^\mu{}_\nu(y_*)^\nu{}_a. \quad (2.61)$$

By integrating Eq. (2.60) over a hypersurface Σ we get

$$\omega_{AB}(\delta_1\psi,\delta_1 y)^A(\delta_2\psi,\delta_2 y)^B = \int_\Sigma [(\delta_1\pi\cdot\delta_2\psi - \delta_2\pi\cdot\delta_1\psi) + (\delta_1 H_a\,\delta_2 y^a - \delta_2 H_a\,\delta_1 y^a)], \quad (2.62)$$

where π and H_a are defined by

$$\pi \equiv P^\mu n_\mu = -(\sqrt{-g}g^{\mu\nu}\partial_\nu\psi)n_\mu, \quad (2.63)$$

$$H_a \equiv (\sqrt{-g}T^\nu{}_a)n_\nu. \quad (2.64)$$

Note that π is the usual canonical momentum of the Klein–Gordon scalar field in the space-time $(M,g_{\mu\nu})$. Note also that H_a on Σ is a function only of ψ, π, and y^a and the spatial derivatives of ψ and y^a.

From Eq. (2.62), and this property of H_a, we see immediately that any two field configurations having the same values of ψ, π, and y^a on Σ lie in the same degeneracy submanifold of ω_{AB}. Furthermore, by expanding $\delta_1 H_a$ in terms of $\delta_1\psi$, $\delta_1\pi$, and $\delta_1 y^a$, one can verify that the field variation $(\delta\psi,\delta\pi,\delta y^a)$ on Σ is a degeneracy direction of ω_{AB} if and only if

$$\delta\psi = h^\mu{}_\nu\,\delta y^\nu\,\partial_\mu\psi - u_\mu\,\delta y^\mu\pi/\sqrt{h}, \quad (2.65)$$

$$\delta\pi = \partial_\mu(h^\mu{}_\nu\,\delta y^\nu\pi) - \partial_\mu(u_\sigma\,\delta y^\sigma\sqrt{h}\,h^{\mu\nu}\,\partial_\nu\psi), \quad (2.66)$$

where $\delta y^\mu = (y_*)^\mu{}_a\,\delta y^a$, $h_{\mu\nu}$ is the spatial metric on Σ induced from $g_{\mu\nu}$, and u^μ is the unit normal to Σ. Thus we sweep out the degeneracy submanifold by allowing y to vary arbitrarily, with the corresponding variations of ψ and π given by Eqs. (2.65) and (2.66). Hence we can uniquely characterize each degeneracy submanifold of ω_{AB} by choosing a fixed diffeomorphism, $y: M\mapsto\widetilde{M}$, and specifying the values of ψ and π on Σ associated with this choice of y. In this manner, we may identify phase space (Γ,Ω_{AB}) with the manifold composed of the pairs (ψ,π) on Σ, with symplectic form

$$\Omega_{AB}(\delta_1\psi,\delta_1\pi)^A(\delta_2\psi,\delta_2\pi)^B$$
$$= \int_\Sigma [(\delta_1\pi)(\delta_2\psi) - (\delta_2\pi)(\delta_1\psi)]. \quad (2.67)$$

This is precisely the usual phase space of a scalar field in the fixed, nondynamical space-time $(M,g_{\mu\nu})$. Thus, when applied to this case, our general prescription for constructing phase space has the effect of "deparametrizing" the parametrized scalar field theory. Note that solutions to the equations of motion (2.55) and (2.57) exist for all (ψ,π) on Σ, so in this case there are no constraints, i.e., the constraint submanifold is all of phase space, $\overline{\Gamma} = \Gamma$.

III. LOCAL SYMMETRIES ON FIELD CONFIGURATION SPACE

In this section we shall define the notion of local symmetries on field configuration space $\mathscr{F}$ and will derive some of their properties. The principal result of this section is a theorem that directly implies that on the solution submanifold $\overline{\mathscr{F}}$ every local symmetry direction $(\hat{\delta}\phi)^A$ is a degeneracy direction of the pullback to $\overline{\mathscr{F}}$, $\overline{\omega}_{AB}$, of the presymplectic form ω_{AB}. In the next section, this result will be used to obtain important relationships between local symmetries and constraints on phase space.

Roughly speaking, a local symmetry is a field variation—such as the gauge transformations of Yang–Mills theory or the diffeomorphisms of general relativity—that keeps the action S invariant and that is "local" in a suitable sense to be made precise below. In this paper, we shall be concerned exclusively with infinitesimal local symmetries. Each infinitesimal local symmetry at each field configuration $\phi\in\mathscr{F}$ gives rise to a vector $(\delta\phi)^A$ in the tangent space to ϕ. In order to distinguish notationally such local symmetry vectors from arbitrary tangent vectors at $\phi\in\mathscr{F}$, we will place a caret over local symmetry vectors, i.e., $(\hat{\delta}\phi)^A$ will denote a field variation corresponding to an infinitesimal local symmetry, and similarly $\hat{\delta}\phi^a(x)$ will denote the tangent vector to M' for this infinitesimal local symmetry at the point $\phi(x)\in M'$. Variations of other quantities induced by local symmetry variations also will be denoted with a caret, e.g., $\hat{\delta}\mathscr{L}$ denotes the change in $\mathscr{L}$ resulting from $\hat{\delta}\phi^a$ [see Eq. (3.1)].

The most difficult part of formulating a notion of local symmetries for a general Lagrangian field theory—formulated within the framework of the previous section—is to capture the idea that one has "complete, local (in space-time) control" over the symmetry variations. The following definition provides such a notion in a form conveniently applicable to the proofs of the lemmas and theorem of this section. As we shall explain further below, this definition encompasses the standard notions of infinitesimal local symmetries in specific theories such as Yang–Mills theory and general relativity.

Definition: A set of pairs $(\hat{\delta}\phi^a,\alpha^\mu)$ consisting of a field variation $\hat{\delta}\phi^a$ on space-time (i.e., an M'-vector-valued scalar field on M) and a vector density α^μ on M will be said to comprise a collection of infinitesimal local symmetries at field configuration ϕ if the following three conditions are satisfied.

(i) The pairs form a vector space, i.e., if $(\hat{\delta}_1\phi^a,\alpha_1^\mu)$ and $(\hat{\delta}_2\phi^a,\alpha_2^\mu)$ are in the collection, then for all $c_1,c_2\in\mathbb{R}$ so is
$$(c_1\,\hat{\delta}_1\phi^a + c_2\,\hat{\delta}_2\phi^a, c_1\alpha_1^\mu + c_2\alpha_2^\mu).$$

(ii) For each pair $(\hat{\delta}\phi^a,\alpha^\mu)$ we have
$$\hat{\delta}\mathscr{L} = \nabla_\mu\alpha^\mu. \quad (3.1)$$
Furthermore, if Σ is noncompact, we require α^μ to be such that no "boundary terms at spatial infinity" arise from applying Gauss' law to integration of $\nabla_\mu\alpha^\mu$ over the region between two Cauchy surfaces Σ_1 and Σ_2. In the noncompact case, we also require $\hat{\delta}\phi^a$ to be such that no such spatial boundary terms arise from a similar application of Gauss' law to $\nabla_\mu\hat{\theta}^\mu$, where $\hat{\theta}^\mu$ is given by Eq. (2.12) with $\delta\phi^a$ replaced by $\hat{\delta}\phi^a$.

(iii) Given any pair $(\hat{\delta}\phi^a,\alpha^\mu)$ in the collection and given any two disjoint, closed subsets $\mathscr{C}_1,\mathscr{C}_2\subset M$ of space-time, then there exists a pair $(\hat{\delta}'\phi^a,\alpha'^\mu)$ in the collection such that
$$\left.\begin{aligned}\hat{\delta}'\phi^a(x) &= \hat{\delta}\phi^a(x)\\ \alpha'^\mu(x) &= \alpha^\mu(x)\end{aligned}\right\}\ \forall x\in\mathscr{C}_1; \quad (3.2a)$$
whereas
$$\left.\begin{aligned}\hat{\delta}'\phi^a(x) &= 0\\ \alpha'^\mu(x) &= 0\end{aligned}\right\}\ \forall x\in\mathscr{C}_2. \quad (3.2b)$$

Note that condition (i) of the definition is not restrictive in that if one has a set of pairs $(\hat{\delta}\phi^a,\alpha^\mu)$ satisfying conditions (ii) and (iii), one can obtain a set satisfying all the conditions by taking their linear span. Note also that Eq. (3.1) by itself places no restriction on $\hat{\delta}\phi^a$, since one could simply solve Eq. (3.1) for α^μ. Condition (ii) becomes restrictive only in conjunction with condition (iii), which states that one must be able to deform $\hat{\delta}\phi^a$ and α^μ to zero on $\mathscr{C}_2$ while preserving Eq. (3.1). It should be emphasized that condition (iii) applies to both $\hat{\delta}\phi^a$ and α^μ, i.e., it does not suffice to be able to deform only $\hat{\delta}\phi^a$ to zero on $\mathscr{C}_2$. Finally, it should be noted that the same field variation $\hat{\delta}\phi^a$ may occur many times in the collection of local symmetries, paired with different vector densities α^μ, i.e., a given local symmetry variation $\hat{\delta}\phi^a$ need not have a unique α^μ associated to it. Quantities defined below, such as the Noether current J^μ, will depend on the choice of α^μ.

Let W_ϕ denote the subspace of the tangent space V_ϕ to $\mathscr{F}$ at field configuration ϕ spanned by the local symmetry vectors $(\hat{\delta}\phi)^A$ at ϕ. An assignment of local symmetries to each $\phi\in\mathscr{F}$ will be said to comprise an *algebra of local symmetries on* $\mathscr{F}$ if the distribution of subspaces W_ϕ is integrable. By Frobenius' theorem, this is equivalent to requiring that the distribution W_ϕ comprise a sub-bundle of the tangent bundle to $\mathscr{F}$ and that the commutator of any two vector fields on $\mathscr{F}$ lying in W_ϕ also lies in W_ϕ. For the results obtained in this section, it is not necessary that the local symmetries comprise an algebra.

A collection of local symmetries at field configuration $\phi\in\mathscr{F}$ always is produced if all of the pairs $(\hat{\delta}\phi^a,\alpha^\mu)$ in the collection satisfy condition (ii) and are generated by formulas locally having the general form

$$\hat{\delta}\phi^a = T_0\Lambda + T_1\,\partial\Lambda + \cdots + T_l\,\partial^l\Lambda, \tag{3.3}$$

$$\alpha^\mu = U_0\Lambda + U_1\,\partial\Lambda + \cdots + U_m\,\partial^m\Lambda, \tag{3.4}$$

where, for a given $\phi\in\mathscr{F}$, the quantities $T_0,\ldots,T_l$ and $U_0,\ldots,U_m$ are fixed tensor fields on space-time, but the tensor field Λ can be chosen arbitrarily (within its specified index type) on space-time. [We have omitted all indices on the right sides of Eqs. (3.3) and (3.4) because the tensor fields T, U, and Λ may have arbitrary index structure (with respect to both M and M') and the index contractions may occur in an arbitrary fashion in each of the terms.] On account of the linearity of Eqs. (3.3) and (3.4) in Λ and its derivatives, it is clear that the pairs $(\hat{\delta}\phi^a,\alpha^\mu)$ comprise a vector space, so property (i) is satisfied. Since Λ is arbitrary, given disjoint closed sets $\mathscr{C}_1,\mathscr{C}_2\subset M$, we can choose Λ' so that $\Lambda'=\Lambda$ on $\mathscr{C}_1$ but $\Lambda'=0$ on $\mathscr{C}_2$. From this it follows that also property (iii) is satisfied.

In Yang–Mills theory, at field configuration $A^i{}_\mu$ the infinitesimal gauge transformations

$$\hat{\delta}A^i{}_\mu = \partial_\mu\Lambda^i + c^i{}_{jk}A^j{}_\mu\Lambda^k \tag{3.5}$$

satisfy Eq. (3.1) with $\alpha^\mu=0$, where Λ^i is an arbitrary Lie-algebra-valued scalar field on M. Thus the pairs $\{(\hat{\delta}A^i{}_\mu,0)\}$, with $\hat{\delta}A^i{}_\mu$ given by Eq. (3.5), comprise a collection of infinitesimal local symmetries at $A^i{}_\mu$. Furthermore, the tangent subspaces to $\mathscr{F}$ generated in this manner are integrable, so these gauge transformations define an algebra of local symmetries on $\mathscr{F}$.

For general relativity with Lagrangian density $\mathscr{L}_{\rm H}=\sqrt{-g}R$, at field configuration $g_{\mu\nu}$ the infinitesimal gauge transformations

$$\hat{\delta}g_{\mu\nu} = \pounds_\Lambda g_{\mu\nu} = 2\nabla_{(\mu}\Lambda_{\nu)} = 2\partial_{(\mu}\Lambda_{\nu)} - \Gamma^\sigma{}_{\mu\nu}\Lambda_\sigma \tag{3.6}$$

satisfy Eq. (3.1) with

$$\alpha^\mu = \Lambda^\mu\mathscr{L}_{\rm H}, \tag{3.7}$$

where Λ^μ is an arbitrary vector field on space-time. [For the modified Lagrangian (2.45), which differs from $\mathscr{L}_{\rm H}$ by addition of a term of the form $\nabla_\mu v^\mu$, the vector density α^μ would be changed by the addition of $\hat{\delta}v^\mu$.] Thus these gauge transformations comprise a collection of infinitesimal local symmetries at $g_{\mu\nu}$. Again, the tangent subspaces to $\mathscr{F}$ determined by Eq. (3.6) are integrable, so we obtain an algebra of local symmetries on $\mathscr{F}$.

In the parametrized massless scalar field theory, the infinitesimal transformations at field configuration (ψ,y^a),

$$\hat{\delta}\psi = \pounds_\Lambda\psi = \Lambda^\mu\,\partial_\mu\psi,\quad \hat{\delta}y^a = (y_*)^a{}_\mu\Lambda^\mu, \tag{3.8}$$

satisfy Eq. (3.1) with $\alpha^\mu=\Lambda^\mu\mathscr{L}$, where Λ^μ is an arbitrary vector field on space-time. Thus these transformations comprise a collection of infinitesimal local symmetries at (ψ,y^a). In this case, also, the tangent subspaces to $\mathscr{F}$ given by Eq. (3.8) are integrable, and we obtain an algebra of local symmetries. Thus the gauge transformations of Yang–Mills theory, general relativity, and parametrized scalar field theory are encompassed by our general definition of local symmetries.

In fact, the local symmetries of Yang–Mills theory, general relativity, and parametrized scalar field theory actually have more structure than indicated above. In all three cases, one can define an action of (infinite-dimensional) group $\mathscr{G}$ on $\mathscr{F}$ so that $\mathscr{F}$ is given the structure of a principal fiber bundle. The tangent subspaces W_ϕ of the infinitesimal local symmetries then correspond simply to the tangent subspaces to the fibers. As mentioned in Sec. I, in Yang–Mills theory, the group $\mathscr{G}$ consists of the set of maps from space-time M into the Yang–Mills group G, whereas in general relativity, the group $\mathscr{G}$ consists of the diffeomorphisms of M into itself. In parametrized theories, the role of $\mathscr{G}$ is also played by the group of diffeomorphisms on M.

The key additional structure provided by such a group action on $\mathscr{F}$ is that it allows one to define the notion of a "field-independent" infinitesimal local symmetry. Namely, we say that an infinitesimal local symmetry $(\hat{\delta}_1\phi)^A$ at $\phi_1\in\mathscr{F}$ is "the same symmetry" as $(\hat{\delta}_2\phi)^A$ at $\phi_2\in\mathscr{F}$ if $(\hat{\delta}_1\phi)^A$ and $(\hat{\delta}_2\phi)^A$ correspond to the action on $\mathscr{F}$ of the same element of the Lie algebra of $\mathscr{G}$. A vector field $(\hat{\delta}\phi)^A$ on $\mathscr{F}$ which is everywhere tangent to W_ϕ will be said to be a *field-independent infinitesimal local symmetry* if it represents the "same symmetry" in this sense at all points of $\mathscr{F}$. [Otherwise, $(\hat{\delta}\phi)^A$ will be referred to as "field dependent".] Clearly, the field-independent infinitesimal local symmetries comprise a subalgebra—isomorphic to the Lie algebra of $\mathscr{G}$—of the algebra of infinitesimal local symmetries. Since the field-independent symmetries span W_ϕ at each $\phi\in\mathscr{F}$, in most cases nothing is lost by restricting attention to them. Thus, in most discussions, only the field-independent symmetries are considered. For Yang–Mills theory the field-independent local symmetries are given by Eq. (3.5), with Λ^i chosen to be independent of $A^i{}_\mu$. For general relativity and parametrized scalar field theory the field-independent local symmetries are those for which Λ^μ (as opposed to, say, Λ_μ) is chosen to be independent of $g_{\mu\nu}$ in Eq. (3.6) and independent of ψ and y in Eq. (3.8).

One reason why we have not assumed the existence of the structure on $\mathscr{F}$ sufficient to enable us to define the notion of field-independent symmetries is that it is unnecessarily restrictive. A much more fundamental reason is that, as we shall explain in Sec. IV, even when this structure is present, it will be necessary to consider the "field-dependent" infinitesimal local symmetries in order to represent fully symmetries on phase space. Thus, for the purpose of this paper, the notion of field-independent local symmetries appears quite unnatural.

We define now the notion of the Noether charge Q of an infinitesimal local symmetry and obtain some of its properties. Let $(\hat{\delta}\phi^a,\alpha^\mu)$ be an infinitesimal local symmetry at field configuration ϕ. We define the Noether current J^μ by

$$J^\mu = \hat{\theta}^\mu - \alpha^\mu, \tag{3.9}$$

where $\hat{\theta}^\mu$ is given by Eq. (2.12), with $\hat{\delta}\phi^a$ substituted for $\delta\phi^a$. Thus J^μ is a vector density on space-time. Note that J^μ depends upon the choice of α^μ, i.e., if the same local symmetry field variation $\hat{\delta}\phi^a$ is paired with different α^μ's, different Noether currents will result.

Taking the divergence of Eq. (3.9), we obtain

$$\begin{aligned}\nabla_\mu J^\mu &= \nabla_\mu \hat{\theta}^\mu - \nabla_\mu \alpha^\mu \\ &= (\hat{\delta}\mathscr{L} - E_a\, \hat{\delta}\phi^a) - \hat{\delta}\mathscr{L} \\ &= -E_a \hat{\delta}\phi^a, \end{aligned} \tag{3.10}$$

where Eqs. (2.10) and (3.1) were used. Thus if ϕ satisfies the equations of motion, then the Noether current is conserved.

Given a Cauchy surface Σ, we define the Noether charge Q associated with $(\hat{\delta}\phi^a, \alpha^\mu)$ by

$$Q = \int_\Sigma J^\mu n_\mu. \tag{3.11}$$

In general, Q depends upon the choice of Σ, but by Eq. (3.10) [supplemented by condition (ii) of the above definition of local symmetries in the case where Σ is noncompact] if ϕ is a solution to the equations of motion, then Q is independent of the choice of Σ. Indeed we have the following stronger result.

Lemma 1: Let $(\hat{\delta}\phi^a, \alpha^\mu)$ be an arbitrary infinitesimal local symmetry at a solution ϕ. Then the Noether charge Q, defined by Eq. (3.11), vanishes:

$$Q = 0. \tag{3.12}$$

Proof: Choose disjoint closed sets $\mathscr{C}_1, \mathscr{C}_2 \subset M$ such that $\mathscr{C}_1$ contains an open neighborhood of the given Cauchy surface Σ appearing in Eq. (3.11), whereas $\mathscr{C}_2$ contains an open neighborhood of another Cauchy surface $\tilde{\Sigma}$. Using property (iii) of the definition of infinitesimal local symmetries, we choose $(\hat{\delta}'\phi^a, \alpha'^\mu)$ to be an infinitesimal local symmetry that agrees with $(\hat{\delta}\phi^a, \alpha^\mu)$ on $\mathscr{C}_1$ but vanishes on $\mathscr{C}_2$. Then, as proved above, the Noether charge Q' associated with $(\hat{\delta}'\phi^a, \alpha'^\mu)$ is conserved. However, clearly we have $Q' = Q$ on Σ whereas $Q' = 0$ on $\tilde{\Sigma}$. Thus we obtain $Q = 0$. □

Our next result provides a generalized Bianchi identity for all Lagrangian theories with local symmetries.

Lemma 2: Let ϕ be an arbitrary field configuration, and let $(\hat{\delta}\phi^a, \alpha^\mu)$ be an infinitesimal local symmetry such that both $\hat{\delta}\phi^a$ and α^μ have compact support on space-time. Then, we have

$$\int_M E_a\, \hat{\delta}\phi^a = 0. \tag{3.13}$$

Proof: Since $\hat{\delta}\phi^a$ and α^μ have compact support, it follows immediately that the Noether current J^μ also has compact support. The lemma then follows directly from integrating Eq. (3.10) over M. □

If the local symmetries are of the form (3.3) and (3.4), then Eq. (3.13) is equivalent to a local differential identity on E_a [obtained by integration of Eq. (3.13) by parts to remove derivatives of Λ and then setting the coefficient of Λ in the integrand equal to zero]. For Yang–Mills theory and general relativity, this yields the usual Bianchi identity. For the parametrized scalar field theory, it yields an identity relating the y^a equation of motion (i.e., conservation of stress–energy) to the ψ equation of motion (i.e., the Klein–Gordon equation).

The next result can be interpreted as saying that at any point ϕ of the solution submanifold $\overline{\mathscr{F}}$ of field configuration space $\mathscr{F}$, every local symmetry vector $(\hat{\delta}\phi)^A$ lies tangent to $\overline{\mathscr{F}}$. The basic argument in our proof of Lemma 3 was suggested to us by S. Anco.

Lemma 3: Let ϕ be a solution to the equations of motion, $E_a = 0$. Let $\hat{\delta}\phi^a$ be an infinitesimal local symmetry at ϕ. Then the induced variation of E_a vanishes:

$$\hat{\delta}E_a = 0, \tag{3.14}$$

i.e., $\hat{\delta}\phi^a$ is a solution of the linearized equations of motion.

Proof: We prove Eq. (3.14) first for the case of a local symmetry pair $(\hat{\delta}\phi^a, \alpha^\mu)$ such that both $\hat{\delta}\phi^a$ and α^μ are of compact support on space-time. Let $\phi(\lambda_1, \lambda_2): M \mapsto M'$ be a smooth two-parameter family of field configurations, with $\phi(0,0) = \phi$, such that, for all λ_1, the field variation

$$\hat{\delta}_2\phi^a \equiv \left.\frac{\partial\phi^a}{\partial\lambda_2}\right|_{\lambda_2=0} \tag{3.15}$$

is a local symmetry field variation of compact support, paired with an α^μ also of compact support. Then Eq. (3.13) holds at $\lambda_2 = 0$ for all λ_1. Taking the partial derivative of Eq. (3.13) with respect to λ_1 and evaluating at $\lambda_1 = 0$, we obtain

$$\begin{aligned} 0 &= \int_M [(\delta_1 E_a)(\hat{\delta}_2\phi^a) + E_a(\delta_1\hat{\delta}_2\phi^a)] \\ &= \int_M (\delta_1 E_a)(\hat{\delta}_2\phi^a), \end{aligned} \tag{3.16}$$

where we used $E_a = 0$ at ϕ in the second line. On the other hand, by Eq. (2.18) we have

$$0 = (\delta_1 E_a)(\hat{\delta}_2\phi^a) - (\hat{\delta}_2 E_a)(\delta_1\phi^a) + \nabla_\mu \omega^\mu. \tag{3.17}$$

Since $\hat{\delta}_2\phi^a$ has compact support, it follows that ω^μ also has compact support. Hence, integrating Eq. (3.17) over M, we obtain

$$0 = \int_M [(\delta_1 E_a)(\hat{\delta}_2\phi^a) - (\hat{\delta}_2 E_a)(\delta_1\phi^a)]. \tag{3.18}$$

Subtracting Eq. (3.18) from Eq. (3.16), we obtain

$$0 = \int_M (\hat{\delta}_2 E_a)(\delta_1\phi^a). \tag{3.19}$$

Since $\delta_1\phi^a$ is an arbitrary field variation, Eq. (3.19) implies that

$$\hat{\delta}_2 E_a = 0, \tag{3.20}$$

for an arbitrary local symmetry pair $(\hat{\delta}_2\phi^a, \alpha^\mu)$ of compact support.

To prove that Eq. (3.20) holds for an arbitrary local symmetry pair $(\hat{\delta}\phi^a, \alpha^\mu)$ not necessarily of compact support, we suppose Eq. (3.20) failed to hold for such a local symmetry at $x \in M$. Choose disjoint closed sets $\mathscr{C}_1, \mathscr{C}_2 \subset M$ such that $\mathscr{C}_1$ contains an open neighborhood of x and such that the complement of $\mathscr{C}_2$ has compact closure. The local symmetry pair $(\hat{\delta}'\phi^a, \alpha'^\mu)$ provided by property (iii) of the definition of local symmetries then would have compact support and also would fail to satisfy Eq. (3.20) at x, in contradiction with the above result. □

We now are ready to prove the main result of this section, which can be interpreted as stating that at each solution $\phi \in \overline{\mathscr{F}}$, and for each local symmetry $(\hat{\delta}\phi)^A$ at ϕ, we have

$$(dQ)_A = \omega_{AB}(\hat{\delta}\phi)^B, \tag{3.21}$$

where Q is the Noether charge associated with $(\hat{\delta}\phi)^A$, d denotes the exterior derivative on $\mathscr{F}$, and ω_{AB} is the "presymplectic form" on $\mathscr{F}$.

Theorem: Let ϕ be a solution to the equations of motion, $E_a = 0$, and let $\phi(\lambda_1,\lambda_2)$: $M \mapsto M'$ be a smooth two-parameter family of field configurations with $\phi(0,0) = \phi$ such that $\hat{\delta}_2\phi^a \equiv (\partial\phi^a/\partial\lambda_2)|_{\lambda_2=0}$ is the field variation of an infinitesimal local symmetry for all λ_1. Let $Q(\lambda_1)$ denote the Noether charge of this local symmetry at $\phi(\lambda_1,0)$. Then at ϕ we have

$$\delta_1 Q = \omega[\phi,\delta_1\phi,\hat{\delta}_2\phi], \tag{3.22}$$

where ω was defined by Eq. (2.23).

Proof: By Eq. (2.18) we have

$$(\delta_1 E_a)(\hat{\delta}_2\phi^a) - (\hat{\delta}_2 E_a)(\delta_1\phi^a) + \nabla_\mu\omega^\mu = 0. \tag{3.23}$$

The second term vanishes by Lemma 3. Since $E_a = 0$ at ϕ, we may write the first term as $\delta_1(E_a\,\hat{\delta}_2\phi^a)$. Using Eq. (3.10), we see that Eq. (3.23) takes the form

$$\nabla_\mu(-\delta_1 J^\mu + \omega^\mu) = 0. \tag{3.24}$$

Thus, by Gauss' law, the associated charge $\int_\Sigma(-\delta_1 J^\mu + \omega^\mu)n_\mu$ is conserved, i.e., independent of choice of Cauchy surface Σ. An exact repetition of the proof of Lemma 1 above then shows that, in fact, this charge vanishes. Thus we obtain

$$\begin{aligned} 0 &= \int_\Sigma(-\delta_1 J^\mu + \omega^\mu)n_\mu \\ &= -\delta_1 Q + \omega[\phi,\delta_1\phi,\hat{\delta}_2\phi], \end{aligned} \tag{3.25}$$

as we desired to show. □

Note that if we weaken the hypothesis of the above theorem by not requiring ϕ to be a solution, the calculation that previously led to Eq. (3.24) now yields

$$\nabla_\mu(-\delta_1 J^\mu + \omega^\mu) = \hat{\delta}_2(E_a\,\delta_1\phi^a). \tag{3.26}$$

Hence, if we happen to have a local symmetry vector field $(\hat{\delta}_2\phi)^A$ on $\mathscr{F}$ that satisfies $\hat{\delta}_2(E_a\,\delta_1\phi^a) = 0$ for all two-parameter families of the type specified in the theorem [but with $\phi = \phi(0,0)$ now arbitrary], then we again obtain Eq. (3.24), from which the result (3.25) again follows. Thus, for such very special, local symmetry vector fields on $\mathscr{F}$, Eq. (3.25) actually holds at all $\phi \in \mathscr{F}$, not just at solutions $\phi \in \overline{\mathscr{F}}$. As we shall discuss further in Sec. IV, the field-independent gauge transformations of Yang–Mills theory satisfy this property.

As already mentioned above, the Noether charge Q depends, in general, upon α^μ as well as $\hat{\delta}\phi^a$. An interesting corollary of the above theorem is that for any local symmetry $(\hat{\delta}\phi^a,\alpha^\mu)$ at a solution ϕ, the gradient of Q does not depend upon α^μ, nor does it depend upon how $(\hat{\delta}\phi)^A$ and α^μ are extended off of ϕ. [This result follows immediately from the fact that the right side of Eq. (3.22) is independent of these choices.] It also follows immediately from Eq. (3.22) that at a solution ϕ, the first variation of the Noether charge in a degeneracy direction of ω always vanishes.

A further, very important corollary of the above theorem may be stated as follows.

Corollary: Let ϕ^a be a solution to the equations of motion, $E_a = 0$; let $\delta_1\phi^a$ be a solution to the linearized equations of motion at ϕ^a (i.e., let $\delta_1\phi^a$ be such that $\delta_1 E_a = 0$); and let $\hat{\delta}_2\phi^a$ be an infinitesimal local symmetry field variation at ϕ^a. Then we have

$$\omega[\phi^a,\delta_1\phi^a,\hat{\delta}_2\phi^a] = 0. \tag{3.27}$$

Proof: This result is an immediate consequence of Eq. (3.22) together with Lemma 1. □

This corollary can be interpreted as saying that at any $\phi \in \overline{\mathscr{F}}$, every local symmetry vector $(\hat{\delta}\phi)^A$ is a degeneracy direction of the pullback $\overline{\omega}_{AB}$ of ω_{AB} to $\overline{\mathscr{F}}$, i.e., we have

$$\overline{\omega}_{AB}(\hat{\delta}\phi)^B = 0. \tag{3.28}$$

[Note that by Lemma 3, $(\hat{\delta}\phi)^A$ is always tangent to $\overline{\mathscr{F}}$, so the action of $\overline{\omega}_{AB}$ on $(\hat{\delta}\phi)^A$ is well defined.] The analysis of the next section will be based mainly on this result. Note also that this corollary can be viewed as stating that, as claimed in Sec. II, $\overline{\omega}_{AB}$ always is "gauge invariant," i.e., for any solution ϕ and for any pair of linearized solutions $(\delta_1\phi)^A$, $(\delta_2\phi)^A$ at ϕ, the quantity $\omega_{AB}(\delta_1\phi)^A(\delta_2\phi)^B$ depends only upon the gauge equivalence class of $(\delta_1\phi)^A$ and $(\delta_2\phi)^A$. Thus we have given a completely general proof, applicable to an arbitrary Lagrangian theory with local symmetries, of a result of Friedman[2] for the case of general relativity. This result for the cases of Yang–Mills theory and general relativity also was obtained previously (by means of detailed calculations) by Crnković and Witten.[3]

IV. LOCAL SYMMETRIES AND CONSTRAINTS ON PHASE SPACE

Recall that, in Sec. II, we introduced the manifold of field configurations, $\mathscr{F}$; we defined a "presymplectic form" ω_{AB} *on* $\mathscr{F}$; and we constructed phase space (Γ, Ω_{AB}) from $(\mathscr{F},\omega_{AB})$ by a reduction procedure. We thereby also obtained a projection map π: $\mathscr{F} \mapsto \Gamma$. The "constraint submanifold" $\overline{\Gamma}$ of Γ was defined to be the image under π of the solution submanifold $\overline{\mathscr{F}}$ of $\mathscr{F}$ and the restriction $\overline{\pi}$ of π to $\overline{\mathscr{F}}$ gave a similar projection map $\overline{\pi}$: $\overline{\mathscr{F}} \mapsto \overline{\Gamma}$. As in Sec. II, we shall continue to assume below that $\mathscr{F}$ has the structure of a fiber bundle over Γ with projection map π and that $\overline{\mathscr{F}}$ has the structure of a fiber bundle over $\overline{\Gamma}$ with projection map $\overline{\pi}$. We shall also assume that the nondegenerate symplectic form Ω_{AB} on Γ is invertible, so that the notion of a Poisson bracket is well defined on Γ. (Nondegeneracy does not automatically imply invertibility in infinite-dimensional spaces.) We denote the inverse of Ω_{AB} by Ω^{AB}, so that $\Omega_{AB}\Omega^{BC} = \delta_A{}^C$.

In the present section, we shall use the projection map π to "carry down" the notion and properties of local symmetries obtained in Sec. III from $\mathscr{F}$ to Γ and from $\overline{\mathscr{F}}$ to $\overline{\Gamma}$. We thereby shall obtain a relationship between local symmetries and constraints on phase space.

As already noted in Sec. II, the projection map π: $\mathscr{F} \mapsto \Gamma$ induces a corresponding map π^*: $V_\phi \mapsto V_\xi$ from the tangent space V_ϕ of any point $\phi \in \mathscr{F}$ to the tangent space V_ξ of the image point $\xi = \pi(\phi) \in \Gamma$. Thus, given a local symmetry vector $(\hat{\delta}\phi)^A$ at ϕ, we may project it to obtain the vector $\hat{X}^A \equiv \pi^*(\hat{\delta}\phi)^A$ at point ξ. However, since the map π is many-to-one, there is no reason why a local symmetry vector field $(\hat{\delta}\phi)^A$ on $\mathscr{F}$ should have a well defined projection onto Γ;

the vector $\pi^*(\hat{\delta}\phi)^A$ at $\xi\in\Gamma$ obtained by projection may depend upon the choice of point $\phi\in\mathscr{F}$ in fiber $\pi^{-1}[\xi]$ over ξ. Indeed, a necessary and sufficient condition for $(\hat{\delta}\phi)^A$ to have a projection from $\mathscr{F}$ to Γ is that for all degeneracy vector fields ψ^A on $\mathscr{F}$, we have that $\pounds_\psi(\hat{\delta}\phi)^A$ also is a degeneracy vector field, i.e., for all ψ^A satisfying

$$\omega_{AB}\psi^B = 0, \tag{4.1}$$

we have

$$0 = \omega_{AB}\pounds_\psi(\hat{\delta}\phi)^B = \pounds_\psi(\omega_{AB}(\hat{\delta}\phi)^B). \tag{4.2}$$

We shall see below that in the case of parametrized theories, there do not exist any local symmetry vector fields on $\mathscr{F}$ satisfying this condition that project to a nonvanishing vector field on Γ. However, in the case where there are local symmetry vector fields on $\mathscr{F}$ that have a well defined, nontrivial projection to Γ, a notion of local symmetries on Γ can be defined. In fact, for most purposes the requirement that a local symmetry have a well defined projection from all of $\mathscr{F}$ to all of Γ is too strong; it would exclude the nonspatial diffeomorphisms of general relativity (see below). Rather, all that is needed for most of our results is that the local symmetry project from $\bar{\mathscr{F}}$ to $\bar{\Gamma}$. Therefore, we introduce the following two definitions.

Definitions: A vector field $\hat{X}^A$ on Γ will be called an *infinitesimal local symmetry on* Γ if it is of the form $\hat{X}^A = \pi^*(\hat{\delta}\phi)^A$, where $(\hat{\delta}\phi)^A$ is the field variation of an infinitesimal local symmetry defined on all of $\mathscr{F}$. Similarly, a vector field $\hat{X}^A$ on $\bar{\Gamma}$ will be called an *infinitesimal local symmetry on* $\bar{\Gamma}$ if it is of the form $\hat{X}^A = \pi^*(\hat{\delta}\phi)^A$, where $(\hat{\delta}\phi)^A$ is the field variation of an infinitesimal local symmetry defined on $\bar{\mathscr{F}}$.

We explore, now, some of the properties of infinitesimal local symmetries on $\bar{\Gamma}$. Since, by Lemma 3 of Sec. III $(\hat{\delta}\phi)^A$ is always tangent to $\bar{\mathscr{F}}$, for any infinitesimal symmetry $\hat{X}^A$ on $\bar{\Gamma}$, we have

$$\hat{X}^A = \pi^*(\hat{\delta}\phi)^A = \bar{\pi}^*(\hat{\delta}\phi)^A, \tag{4.3}$$

where $\bar{\pi}^*$ is the tangent space map associated to the projection map $\bar{\pi}$: $\bar{\mathscr{F}}\mapsto\bar{\Gamma}$. (We also use the same notation $\bar{\pi}^*$ to denote the associated pullback map on forms.) By Eq. (2.34) of Sec. II, we have

$$\bar{\omega}_{AB} = \bar{\pi}^*\bar{\Omega}_{AB}, \tag{4.4}$$

where $\bar{\omega}_{AB}$ denotes the pullback of the presymplectic form ω_{AB} to $\bar{\mathscr{F}}$, and $\bar{\Omega}_{AB}$ denotes the pullback of Ω_{AB} to $\bar{\Gamma}$. By the elementary properties of the map $\bar{\pi}^*$, we thus obtain

$$\bar{\omega}_{AB}(\hat{\delta}\phi)^B = \bar{\pi}^*(\bar{\Omega}_{AB}\hat{X}^B). \tag{4.5}$$

The key result of this section now follows by invoking the corollary to the theorem proved at the end of Sec. III. According to that corollary, the left side of Eq. (4.5) always vanishes. Thus we obtain

$$\bar{\pi}^*(\bar{\Omega}_{AB}\hat{X}^B) = 0, \tag{4.6}$$

and, consequently, we find that every infinitesimal local symmetry $\hat{X}^A$ on $\bar{\Gamma}$ satisfies

$$\bar{\Omega}_{AB}\hat{X}^B = 0. \tag{4.7}$$

Equation (4.7) directly implies that *to each infinitesimal local symmetry on* $\bar{\Gamma}$, *there corresponds a constraint on phase space*. To see this, we note that since Ω_{AB} is nondegenerate, if the infinitesimal local symmetry $\hat{X}^A$ is nonvanishing at $\xi\in\bar{\Gamma}$, then the one-form

$$\mu_A = \Omega_{AB}\hat{X}^B \tag{4.8}$$

also is nonvanishing at ξ. But Eq. (4.7) states that $\mu_A T^A = 0$, for all T^A tangent to $\bar{\Gamma}$ at ξ. Thus the codimension of $\bar{\Gamma}$ in Γ must be at least 1. Each additional independent infinitesimal local symmetry $\hat{X}'^A$ at ξ results in an additional independent μ'_A and thus increases the codimension of $\bar{\Gamma}$ in Γ by 1. Since, as mentioned in Sec. II, Γ has the interpretation as representing the "kinematically possible" states whereas $\bar{\Gamma}$ represents the "dynamically possible" states, each such increase in the codimension of $\bar{\Gamma}$ corresponds to an additional constraint.

Furthermore, in the case where the infinitesimal local symmetries on $\mathscr{F}$ form an algebra, then *the constraints associated with infinitesimal local symmetries on* $\bar{\Gamma}$ *always are first class*, i.e., these constraints close under the Poisson bracket in the sense explained below. To show this, given an infinitesimal local symmetry $\hat{X}^A$ on $\bar{\Gamma}$, we let C be any function on Γ such that on $\bar{\Gamma}$ we have

$$C|_{\bar{\Gamma}} = 0 \tag{4.9}$$

and

$$(dC)_{A|\bar{\Gamma}} = \Omega_{AB}\hat{X}^B, \tag{4.10}$$

where d denotes the exterior derivative on Γ. [On account of Eq. (4.7), Eq. (4.10) is merely a condition on the normal derivatives of C at $\bar{\Gamma}$, so there are no futher integrability conditions that need be satisfied in order to obtain such a C.] We refer to any C satisfying (4.9) and (4.10) as a "constraint function" associated with $\hat{X}^A$. Note that, by Eq. (3.21), the pullback $\pi^*(dC)_A$ of $(dC)_A$ to $\mathscr{F}$ is equal to $(dQ)_A$ on $\bar{\mathscr{F}}$, where Q is the Noether charge associated with the local symmetry that projects to $\hat{X}^A$ on $\bar{\Gamma}$. For a given choice of constraint function C, it is convenient to extend $\hat{X}^A$ to a vector field X^A on Γ by setting

$$X^A = \Omega^{AB}(dC)_B. \tag{4.11}$$

Then X^A automatically satisfies

$$\pounds_X\Omega_{AB} = 0. \tag{4.12}$$

Now, consider two infinitesimal local symmetries, $\hat{X}_1{}^A$ and $\hat{X}_2{}^A$, on $\bar{\Gamma}$. Under the hypothesis that the infinitesimal local symmetries on $\mathscr{F}$ comprise an algebra, their commutator $\hat{X}^A = \pounds_{\hat{X}_1}\hat{X}_2{}^A$ also will be an infinitesimal local symmetry on $\bar{\Gamma}$. On $\bar{\Gamma}$, we have

$$\begin{aligned}\Omega_{AB}\hat{X}^B &= \Omega_{AB}\pounds_{\hat{X}_1}\hat{X}_2{}^B = \Omega_{AB}\pounds_{X_1}\hat{X}_2{}^B \\ &= \pounds_{X_1}(\Omega_{AB}\hat{X}_2{}^B) = \pounds_{X_1}(dC_2)_A \\ &= (d(X_1^D(dC_2)_D))_A \\ &= -(d(\Omega^{BD}(dC_1)_B(dC_2)_D))_A \\ &= (d\{C_1,C_2\})_A,\end{aligned} \tag{4.13}$$

where the identity (2.27) was used in the third line and where the Poisson bracket $\{C_1,C_2\}$ is defined by

$$\{C_1,C_2\} = -\Omega^{AB}(dC_1)_A(dC_2)_B. \tag{4.14}$$

Note that Eq. (4.13) holds on $\overline{\Gamma}$ for any choice of constraint functions C_1,C_2 satisfying (4.9) and (4.10). Furthermore, we have $\{C_1,C_2\} \doteq 0$ on $\overline{\Gamma}$. Hence $C=\{C_1,C_2\}$ is a constraint function for $\hat{X}^A$. Thus the above calculations establish the following results: To each infinitesimal local symmetry $\hat{X}^A$ on $\overline{\Gamma}$ we may associate an equivalence class $[C]$ of "constraint functions," where $[C]$ consists of all functions on phase space satisfying Eqs. (4.9) and (4.10). The Poisson bracket algebra of these constraint functions is isomorphic to the Lie bracket algebra of the infinitesimal local symmetries on $\overline{\Gamma}$ in the sense that if C_1 is any constraint function for $\hat{X}_1{}^A$ and C_2 is any constraint function for $\hat{X}_2{}^A$, then $\{C_1,C_2\}$ is a constraint function for $\hat{X}^A = £_{\hat{X}_1}\hat{X}_2{}^A$. In particular, since the Lie bracket algebra of the $\hat{X}^A$'s on $\overline{\Gamma}$ close when the infinitesimal local symmetries on $\mathscr{F}$ comprise an algebra, it follows that the Poisson bracket algebra of the constraint functions close. This corresponds to the usual notion of first class constraints, so, in this sense, the constraints associated with local symmetries are first class. Of course, our analysis does not preclude the possibility that the theory may possess other constraints in addition to those implied by the presence of local symmetries, so the entire set of constraints need not be first class.

The above discussion and results apply to all Lagrangian field theories with local symmetries. Hence any difference in the structure of the resulting constraints on phase space for different theories with local symmetries must be attributable to differences in the structure of the local symmetries on $\mathscr{F}$ and/or differences in the manner in which these local symmetries project to $\overline{\Gamma}$. We now shall elucidate the special features of the local symmetries on phase space arising in Yang–Mills theory, general relativity, and the parametrized scalar field theory. We shall see that the differences in the structure of the constraints on phase space in these theories arise primarily from the manner in which the local symmetries project to phase space.

In Yang–Mills theory, the "field-independent" infinitesimal local symmetries on $\mathscr{F}$ give rise to local symmetries on phase space that satisfy much stronger properties than in the general case considered above. As already mentioned in Sec. III, the "field-independent" symmetries on $\mathscr{F}$ are given by

$$\hat{\delta}A^i{}_\mu = \partial_\mu\Lambda^i + c^i{}_{jk}A^j{}_\mu\Lambda^k, \tag{4.15}$$

where Λ^i is a fixed (i.e., independent of field configuration $A^i{}_\mu$) Lie-algebra-valued function, Λ: $M\mapsto L(G)$, on spacetime. In particular, on Σ we have

$$\hat{\delta}A^i{}_\mu = \partial_\mu\Lambda^i + c^i{}_{jk}A^j{}_\mu\Lambda^k, \tag{4.16}$$

$$\hat{\delta}E_i{}^\mu = c_i{}^j{}_k E_j{}^\mu\Lambda^k. \tag{4.17}$$

Furthermore, as discussed in Sec. II, the degeneracy submanifolds of ω_{AB} consist of those field configurations with the same pullback of $A^i{}_\mu$ to Σ and the same $E_i{}^\mu$ on Σ. Thus if we consider the field-independent infinitesimal local symmetry (4.15) at two field configurations on the same degeneracy submanifold, it is clear from Eqs. (4.16) and (4.17) that the fields $\hat{\delta}E^i{}_\mu$ on Σ and the pullback of $\hat{\delta}A_i{}^\mu$ to Σ will be the same. This implies that in Yang–Mills theory, *all the field-independent infinitesimal local symmetries on $\mathscr{F}$ project to Γ and thus define infinitesimal local symmetries on all of Γ.* We shall refer to these local symmetries on Γ as "field-independent." The Lie algebra of the field-independent local symmetries on Γ is isomorphic to the Lie algebra of the group $\mathscr{G}/\mathscr{H}$, where $\mathscr{G}$ is the gauge group acting on $\mathscr{F}$ and $\mathscr{H}$ is the normal subgroup of $\mathscr{G}$ consisting of the gauge transformations that keep $E_i{}^\mu$ and the pullback of $A^i{}_\mu$ fixed on Σ. Furthermore, Γ thereby naturally acquires the structure of a principal fiber bundle, with group $\mathscr{G}/\mathscr{H}$. The constraint submanifold $\overline{\Gamma}$ similarly acquires a natural principal fiber bundle structure, also with group $\mathscr{G}/\mathscr{H}$.

In fact, the field-independent local symmetries of Yang–Mills theory possess further remarkable properties. It is not difficult to verify that the field-independent local symmetries on $\mathscr{F}$ satisfy the condition $\hat{\delta}_2(E_a\,\delta_1\phi^a)=0$ discussed after the theorem at the end of Sec. III. Consequently, the equation

$$(dQ)_A = \omega_{AB}(\hat{\delta}\phi)^B \tag{4.18}$$

holds for field-independent local symmetries $(\hat{\delta}\phi)^A$ at *all* $\phi\in\mathscr{F}$ (i.e., not merely for $\phi\in\overline{\mathscr{F}}$, as guaranteed by the theorem). An immediate consequence of Eq. (4.18) is that the Noether charge Q is constant on the degeneracy submanifolds of ω_{AB}. Consequently, we may project the function Q from $\mathscr{F}$ to Γ so as to obtain a function C on Γ that satisfies

$$(dC)_A = \Omega_{AB}\hat{X}^B, \tag{4.19}$$

everywhere on Γ, where $\hat{X}^A$ is the projection of $(\hat{\delta}\phi)^A$. Note that, in particular, Eq. (4.19) implies that $\hat{X}^A$ is symplectic, i.e., $£_{\hat{X}}\Omega_{AB}=0$. Consequently, in Yang–Mills theory, for any field-independent local symmetry $\hat{X}^A$ on Γ, we can uniquely assign a constraint function C such that $C\doteq 0$ on $\overline{\Gamma}$ and Eq. (4.10) holds on all of Γ, not merely on $\overline{\Gamma}$. Consequently, the Poisson bracket algebra of these constraint functions on Γ is isomorphic to the Lie bracket algebra of the field-independent local symmetries on Γ, which, in turn, is isomorphic to the Lie algebra of $\mathscr{G}/\mathscr{H}$. Thus in Yang–Mills theory the field-independent local symmetries on Γ satisfy much stronger properties than hold in the general case.

In Yang–Mills theory, the most general infinitesimal symmetry vector field $(\hat{\delta}\phi)^A$ on $\mathscr{F}$ that projects to Γ is given by Eq. (4.15), where Λ^i is constant on the degeneracy submanifolds of ω_{AB}, i.e., the map Λ: $M\mapsto L(G)$ is an arbitrary functional of $E_i{}^\mu$ on Σ and the pullback of $A^i{}_\mu$ to Σ. The local symmetries that project from $\overline{\mathscr{F}}$ to $\overline{\Gamma}$ are characterized similarly. Such field-dependent local symmetries on $\mathscr{F}$ fail, in general, to satisfy Eq. (4.18). Their projection to Γ fails, in general, to be symplectic, and thus one cannot associate to them a constraint function on all of Γ. (Of course, as with all local symmetries, their projection to $\overline{\Gamma}$ is associated with an equivalence class of constraint functions in the manner described above.) However, since the field-independent local symmetries on Γ span the tangent subspace of infinitesimal local symmetries at each point of Γ, there is little reason to consider field-dependent local symmetries on Γ.

In general relativity, a certain subclass of the field-independent local symmetries on $\mathscr{F}$—namely, the spatial diffeomorphisms—satisfy special properties similar to the field-independent local symmetries of Yang–Mills theory. Recall from the previous section that the field-independent local symmetries are given by

$$\hat{\delta} g_{\mu\nu} = \pounds_\Lambda g_{\mu\nu}, \tag{4.20}$$

where the vector field Λ^μ is independent of the field configuration $g_{\mu\nu}$ on space-time. (Note that, therefore, $\Lambda_\mu = g_{\mu\nu}\Lambda^\nu$ does depend upon field configuration.) Consider, now, the purely spatial infinitesimal diffeomorphisms, i.e., consider the case where Λ^μ on Σ is everywhere tangential to Σ. Then on Σ we have

$$\hat{\delta} h_{\mu\nu} = \pounds_\Lambda h_{\mu\nu} = 2D_{(\mu}\Lambda_{\nu)}, \tag{4.21}$$

$$\hat{\delta}\pi^{\mu\nu} = \pounds_\Lambda \pi^{\mu\nu} = \partial_\sigma(\Lambda^\sigma \pi^{\mu\nu}) - 2\pi^{\sigma(\mu}\partial_\sigma \Lambda^{\nu)}, \tag{4.22}$$

where D_μ denotes the derivative operator on Σ associated with $h_{\mu\nu}$. Since the degeneracy submanifolds of ω_{AB} consist of field configurations with the same $h_{\mu\nu}$ and $\pi^{\mu\nu}$ on Σ, it follows immediately from Eqs. (4.21) and (4.22) that the field-independent infinitesimal spatial diffeomorphisms project from all of $\mathscr{F}$ to Γ. The Lie algebra of the resulting local symmetries on Γ (which we again refer to as "field independent") is isomorphic to the Lie algebra of the diffeomorphism group of Σ. Furthermore, both Γ and $\bar{\Gamma}$ thereby naturally acquire the structure of a principal fiber bundle, with this structure group. However, unlike the Yang–Mills case, the field-independent infinitesimal spatial diffeomorphisms on $\mathscr{F}$ do not satisfy the condition $\hat{\delta}_2(E_a\,\delta_1\phi^a) = 0$ discussed following the theorem at the end of Sec. III. Instead, we obtain

$$\begin{aligned}\hat{\delta}_2(E_a\,\delta_1\phi^a) &= -\pounds_{\Lambda_2}(\sqrt{-g}\,G^{\alpha\beta}\,\delta_1 g_{\alpha\beta})\\ &= -\partial_\mu(\Lambda_2^\mu\sqrt{-g}\,G^{\alpha\beta}\,\delta_1 g_{\alpha\beta}).\end{aligned} \tag{4.23}$$

Nevertheless, remarkably, when this term is inserted into Eq. (3.25), it makes no contribution, since $\Lambda_2^\mu n_\mu = 0$. Consequently, for the field-independent infinitesimal spatial diffeomorphisms $(\hat{\delta}\phi)^A$ on $\mathscr{F}$, we again obtain

$$(dQ)_A = \omega_{AB}(\hat{\delta}\phi)^B, \tag{4.24}$$

everywhere on $\mathscr{F}$. Hence, by exactly the same arguments as in the Yang–Mills case, to each of the resulting field-independent local symmetries on Γ we can uniquely associate a constraint function on all of Γ in such a way that the Poisson bracket algebra of these functions is isomorphic to the Lie bracket algebra of the local symmetries, which, in turn, is isomorphic to the Lie algebra of diffeomorphisms on Σ. Thus the structure for the field-independent spatial diffeomorphisms in general relativity is completely analogous to the structure of the field-independent gauge transformations of Yang–Mills theory.

However, the structure for the nonspatial diffeomorphisms of general relativity is very different. For general Λ^μ, the formulas corresponding to Eqs. (4.21) and (4.22) are [see, e.g., Eqs. (E.2.35) and (E.2.36) of Ref. 9 for the case $G_{\mu\nu} = 0$]

$$\hat{\delta} h_{\mu\nu} = 2\alpha h^{-1/2}(\pi_{\mu\nu} - \tfrac{1}{2}h_{\mu\nu}\pi) + 2D_{(\mu}\beta_{\nu)}, \tag{4.25}$$

$$\begin{aligned}\hat{\delta}\pi^{\mu\nu} = {}& \alpha h^{1/2}h^{\mu\sigma}h^{\nu\rho}G_{\sigma\rho} - \alpha h^{1/2}({}^3R^{\mu\nu} - \tfrac{1}{2}{}^3Rh^{\mu\nu})\\ &+ \tfrac{1}{2}\alpha h^{-1/2}h^{\mu\nu}(\pi^{\rho\sigma}\pi_{\rho\sigma} - \pi^2)\\ &- 2\alpha h^{-1/2}(\pi^{\mu\sigma}\pi_\sigma{}^\nu - \tfrac{1}{2}\pi\pi^{\mu\nu})\\ &+ h^{1/2}(D^\mu D^\nu\alpha - h^{\mu\nu}D^\sigma D_\sigma\alpha)\\ &+ \partial_\sigma(\beta^\sigma\pi^{\mu\nu}) - 2\pi^{\sigma(\mu}D_\sigma\beta^{\nu)},\end{aligned} \tag{4.26}$$

where $\alpha = -u_\mu\Lambda^\mu$ (with u_μ the unit normal to Σ) and $\beta^\mu = h^\mu{}_\nu\Lambda^\nu$ are the lapse function and shift vector associated with Λ^μ, and $G_{\mu\nu}$ is the Einstein tensor of $g_{\mu\nu}$. Note that the term involving the Einstein tensor is the only term on the right sides of Eqs. (4.25) and (4.26) not determined by the specification of $h_{\mu\nu}$, $\pi^{\mu\nu}$, and Λ^μ on Σ. Indeed, on a given degeneracy submanifold (i.e., for fixed $h_{\mu\nu}$ and $\pi^{\mu\nu}$ on Σ), the space-time metric $g_{\mu\nu}$ can be chosen so that the spatial projection of $G_{\mu\nu}$ can be specified arbitrarily. Taking this fact into account, we see from Eqs. (4.25) and (4.26) that the only way α and β^μ can be chosen on a given degeneracy submanifold so that $\hat{\delta}h_{\mu\nu}$ and $\hat{\delta}\pi^{\mu\nu}$ are constant (i.e., so that a consistent projection to Γ is obtained) is to choose $\alpha = 0$ and β^μ independent of field configuration. Thus the most general local symmetries that project from $\mathscr{F}$ to Γ are the infinitesimal spatial diffemorphisms, with $\Lambda^\mu = \beta^\mu$ chosen to be an arbitrary functional of $(h_{\mu\nu},\pi^{\mu\nu})$ on Σ. In particular, *there is no notion of local symmetries on all of* Γ *corresponding to nonspatial diffeomorphisms.*

However, this situation improves considerably if one seeks only local symmetries that project from $\bar{\mathscr{F}}$ to $\bar{\Gamma}$. In that case, we have $G_{\mu\nu} = 0$ and it is clear from Eqs. (4.25) and (4.26) that $\hat{\delta}h_{\mu\nu}$ and $\hat{\delta}\pi^{\mu\nu}$ will be constant on each degeneracy submanifold if and only if α and β^μ are constant there. Thus *the most general local symmetries that project from* $\bar{\mathscr{F}}$ *to* $\bar{\Gamma}$ *are those for which* α *and* β^μ *are arbitrary functionals of* $h_{\mu\nu}$ *and* $\pi^{\mu\nu}$ on Σ. In particular, *none of the "field-independent" local symmetries with nonspatial* Λ^μ project from $\bar{\mathscr{F}}$ *to* $\bar{\Gamma}$, because if Λ^μ is constant on a degeneracy submanifold, then $\alpha = -u_\mu\Lambda^\mu$ and $\beta^\mu = h^\mu{}_\nu\Lambda^\nu$ are nonconstant, since u_μ *and* $h^\mu{}_\nu = \delta^\mu{}_\nu + u^\mu u_\nu$ depend upon properties of the space-time metric $g_{\mu\nu}$ that are not determined by $h_{\mu\nu}$ and $\pi^{\mu\nu}$ on Σ.

The infinitesimal local symmetries on $\bar{\Gamma}$ comprise an algebra that is "as large" an algebra as one would have expected to obtain if the field-independent local symmetries had projected down. Thus the field variations corresponding to the nonspatial infinitesimal diffeomorphisms are fully represented on $\bar{\Gamma}$. Furthermore, by our general analysis above, for each infinitesimal local symmetry on $\bar{\Gamma}$ we obtain an equivalence class of constraint functions whose Poisson bracket algebra is isomorphic to the Lie bracket algebra of the local symmetries on $\bar{\Gamma}$. However, since there is no subalgebra of local symmetries on $\bar{\Gamma}$ corresponding to the field-independent, nonspatial, infinitesimal diffeomorphisms on $\mathscr{F}$, there is no natural action on $\bar{\Gamma}$ of the group of field-independent local symmetries on $\mathscr{F}$ ($\cong$ the group of diffemorphism of M) or any of its factor groups, apart from the spatial diffeomorphisms.

In summary, we see that the difference in the manner in which the local symmetries project from field configuration space to phase space fully accounts for the following differences between Yang–Mills theory and general relativity.

(1) In Yang–Mills theory the notion of local symmetries is naturally defined on all of phase space Γ. In general relativity, although the local symmetries corresponding to spatial diffeomorphisms also are naturally defined on Γ, the local symmetries corresponding to nonspatial diffeomorphisms are defined only on the constraint submanifold $\bar{\Gamma}$.

(2) In Yang–Mills theory the "field-independent" local symmetries on $\mathscr{F}$ project to a natural subalgebra of local symmetries on Γ, also referred as "field independent." To each such field-independent local symmetry on Γ one can uniquely associate a constraint function C defined on all of Γ and satisfying the properties discussed above. In general relativity, the field-independent local symmetries on $\mathscr{F}$ corresponding to spatial diffeomorphisms similarly project to Γ, and, for the resulting field-independent local symmetries on Γ, constraint functions also can be defined on Γ that satisfy similar properties. However, for nonspatial diffeomorphisms, the field-independent local symmetries do not even project from $\bar{\mathscr{F}}$ to $\bar{\Gamma}$, so one does not obtain a similar subalgebra. Furthermore, for the nonspatial diffeomorphisms, constraint functions on Γ are naturally defined only up to equivalence class as described above.

(3) In Yang–Mills theory field-independent local symmetries on phase space naturally endow Γ and $\bar{\Gamma}$ with the structure of principal fiber bundles, with the structure group being a factor group of the group $\mathscr{G}$ of field-independent local symmetries on $\mathscr{F}$. For general relativity, similar results hold for the spatial diffeomorphisms. However, since there do not exist any local symmetries on Γ or $\bar{\Gamma}$ corresponding to the field-independent nonspatial diffeomorphisms of $\mathscr{F}$, neither Γ nor $\bar{\Gamma}$ naturally acquire a similar principal fiber bundle structure with regard to an appropriately large factor group of the diffeomorphism group of M (i.e., a factor group that includes any nonspatial diffeomorphisms).

With regard to this last point, since in general relativity the local symmetries corresponding to the nonspatial diffeomorphisms are not even defined off of $\bar{\Gamma}$, there is, of course, no sense in which Γ can be given a principal bundle structure associated with them. However, as described above, the local symmetries corresponding to all diffeomorphisms are fully represented on $\bar{\Gamma}$. Suppose that one could find a subalgebra $\mathscr{A}$ of these local symmetries on $\bar{\Gamma}$ such that each $\hat{X}^A \in \mathscr{A}$ is nonvanishing at each point of $\bar{\Gamma}$ (unless $\hat{X}^A$ vanishes identically) and such that at each point of $\bar{\Gamma}$, the local symmetry vectors in $\mathscr{A}$ span the tangent subspace of local symmetries. Then $\bar{\Gamma}$ would acquire principal bundle structure with respect to the subgroup $\mathscr{D}_{\mathscr{A}}$ of diffeomorphisms of $\bar{\Gamma}$ generated by $\mathscr{A}$. If so, then $\mathscr{D}_{\mathscr{A}}$ could be viewed as the local symmetry group of the phase space formulation of geneal relativity. Note that, presumably, $\mathscr{D}_{\mathscr{A}}$ would not be a factor group or subgroup of the diffeomorphism group of M. The issue of whether there exists such a subalgebra $\mathscr{A}$ of the local symmetries on $\bar{\Gamma}$ appears worthy of further investigation. Note that if such a subalgebra $\mathscr{A}$ exists, the constraint functions associated with elements of $\mathscr{A}$ would suffice to enforce all the constraints of general relativity. The Poisson bracket algebra of these constraint functions then would close on $\bar{\Gamma}$ with "structure constants" rather than "structure functions." [Indeed, one way of searching for such an $\mathscr{A}$ would be to try to choose the field dependence of α and β^μ on Γ such that the Poisson bracket algebra corresponding to the algebra (1.4), (1.6), and (1.7) obtained for "field-independent" α nd β^μ now closes with structure constants on $\bar{\Gamma}$.] Thus the existence of such an $\mathscr{A}$ could provide a possible remedy to the difficulties, mentioned in the Introduction, that arise when one attempts to impose the constraints in quantum gravity.

Finally, we note that the situation for local symmetries on phase space for the parametrized scalar field is completely different from Yang–Mills theory and general relativity. For an arbitrary field configuration (ψ,y) on space-time the local symmetries (3.8) yield the following field variations on Σ:

$$\hat{\delta} y^a = (y_*)^a{}_\mu \Lambda^\mu, \tag{4.27}$$

$$\hat{\delta}\psi = \Lambda^\mu \partial_\mu \psi = \beta^\mu \partial_\mu \psi + \alpha h^{-1/2}\pi, \tag{4.28}$$

$$\hat{\delta}\pi = \partial_\mu(\beta^\mu \pi) + \partial_\mu(\alpha\sqrt{h}\, h^{\mu\nu}\partial_\nu \psi) - \partial_\mu(\sqrt{-g}\, g^{\mu\nu}\partial_\nu\psi). \tag{4.29}$$

Comparing this with Eqs. (2.65) and (2.66), we see that on $\bar{\mathscr{F}}$ [where the last term in Eq. (4.29) vanishes], the local symmetry vector fields are degeneracy directions of ω_{AB}. Hence the map $\bar{\pi}: \bar{\mathscr{F}} \to \bar{\Gamma}$ projects all local symmetries on $\bar{\mathscr{F}}$ to zero on $\bar{\Gamma}$. Indeed, since in this case we have $\Gamma = \bar{\Gamma}$, it follows that if $(\hat{\delta}\phi)^A$ is a local symmetry vector field on any degeneracy submanifold of $\mathscr{F}$ that projects to a vector $\hat{X}$ at corresponding point of Γ, then we have $\hat{X} = \pi^*(\hat{\delta}\phi)^A = 0$. Thus, in parameterized scalar field theory, there are no local symmetries on phase space, as might be expected in view of the fact that—as noted in Sec. II—the construction of phase space, in effect, deparametrized the theory. The absence of nontrivial local symmetries on phase space also is implied by our general arguments above from the fact that no constraints are present.

ACKNOWLEDGMENTS

We want to thank Stephen Anco, Gregory Burnett, and Robert Geroch for several helpful remarks and Karel Kuchař for suggesting that we include an analysis of the parametrized scalar field theory in our general framework.

This research was supported, in part, by National Science Foundation Grant No. PHY 84-16691 to the University of Chicago.

APPENDIX: HAMILTONIAN FORMULATIONS

In this appendix, we analyze the circumstances under which the general type of Lagrangian field theory considered in Sec. II can be given a Hamiltonian formulation on phase space. For simplicity, we shall restrict attention to the "time-independent" case (see below).

Let t^μ be a complete vector field on space-time M having the property that the diffeomorphisms $\Lambda(t)$ generated by t^μ map the Cauchy surface Σ used in the definition of ω_{AB} [see Eq. (2.23)] into Cauchy surfaces. We will refer to t^μ as a "time translation," although it is not necessary that t^μ be timelike. The diffeomorphism $\Lambda(t)$ can be made to act on any field configuration, $\phi: M \to M'$, to produce a one-parameter family, $\phi(t) = \phi \circ \Lambda(t)$, of "time translated" field configurations. We denote the infinitesimal field variation associated with this one-parameter family by $\delta_t \phi^a$.

For simplicity, we now restrict attention to the case where, if nondynamical background fields γ are present in $\mathscr{L}$ [see Eq. (2.2) above], then the time translations generated by t^μ leave γ invariant. Thus the following discussion is applicable, for example, to general relativity (where no background fields are present) and to a field theory in a curved, stationary space-time (with t^μ chosen as the stationary Killing field) but is not applicable to, say, a field theory in a background space-time without symmetries. This restriction corresponds to the consideration of "time-independent" systems in ordinary particle mechanics and we will use that terminology here. (In order to treat the "time-dependent" case, we would need to enlarge $\mathscr{F}$ to include the background fields γ in order to define time evolution properly.) Note that in the "time-independent" case considered here, if ϕ is a solution to the equations of motion, then $\phi(t)$ also will be a solution for all t.

Let τ^A be a vector field on the solution submanifold $\bar{\mathscr{F}}$ such that at each $\phi \in \bar{\mathscr{F}}$, τ^A is of the form $(\delta_t \phi)^A$, for some t^μ that leaves the background fields (if any) invariant. [In order to encompass theories like general relativity, we do not require that t^μ be the same vector field on space-time for different points $\phi \in \bar{\mathscr{F}}$, i.e., t^μ may be "field dependent." In order to encompass Yang–Mills theory, we would have to permit τ^A to differ from $(\delta_t \phi)^A$ by a (field-dependent) local symmetry vector field (i.e., an infinitesimal gauge transformation) $(\hat{\delta}\phi)^A$.] We view τ^A as representing time evolution on $\bar{\mathscr{F}}$. Note that τ^A is tangent to $\bar{\mathscr{F}}$. The result of Sec. II that $\omega[\phi^a, \delta_1 \phi^a, \delta_2 \phi^a]$ is independent of the choice of Cauchy surface Σ used to define ω when ϕ is a solution and $\delta_1 \phi^a$ and $\delta_2 \phi^a$ are linearized solutions implies that

$$\pounds_\tau \bar{\omega}_{AB} = 0, \tag{A1}$$

where $\bar{\omega}_{AB}$ denotes the pullback to $\bar{\mathscr{F}}$ of $\bar{\omega}_{AB}$. Note that, here, we have used the time translation invariance of γ, since varying Σ corresponds to time translating both ϕ and γ, whereas τ^A represents the variation of only the dynamical fields ϕ.

Suppose, now, that we can find a nontrivial vector field τ^A on $\bar{\mathscr{F}}$ as in the previous paragraph such that τ^A has a well defined projection to a vector field T^A on $\bar{\Gamma}$. Note that the existence of such a τ^A is intimately related to the existence of a well posed initial value formulation for the equations of motion for ϕ; if the projection map $\bar{\pi}$: $\bar{\mathscr{F}} \to \bar{\Gamma}$ is many-to-one (i.e., if many solutions exist having the same "initial data"), there need not exist any such nonvanishing τ^A. If such a τ^A does exist, we may view T^A as representing the time evolution on $\bar{\Gamma}$. For such a T^A, Eq. (A1) then directly implies

$$\mathscr{L}_T \bar{\Omega}_{AB} = 0, \tag{A2}$$

where $\bar{\Omega}_{AB}$ is the pullback of Ω_{AB} to $\bar{\Gamma}$.

Equation (A2) is the necessary and sufficient condition that, locally, there exists a function H on phase space Γ such that on $\bar{\Gamma}$ we have

$$(dH)_A = \Omega_{AB} T^B. \tag{A3}$$

To see this, we note that Eq. (A2) is just the integrability condition for the pullback of Eq. (A3) to $\bar{\Gamma}$ since, setting $\gamma_A = -\Omega_{AB} T^B$, we have, using the identity (2.27),

$$\begin{aligned}(\overline{d\lambda})_{AC} &= \pounds_T \bar{\Omega}_{AC} - T^B (\overline{d\Omega})_{BAC} \\ &= \pounds_T \bar{\Omega}_{AC}.\end{aligned} \tag{A4}$$

Hence, if Eq. (A2) is satisfied, we can obtain a function H on $\bar{\Gamma}$ satisfying the pullback of Eq. (A3) to $\bar{\Gamma}$. The remaining components of Eq. (A3) can be satisfied by appropriate choice of the normal derivatives of H on $\bar{\Gamma}$. Of course, as in our discussion of constraints earlier in this section, the extension of H to Γ is highly nonunique, since only the normal derivatives of H and $\bar{\Gamma}$ are determined by Eq. (A3).

Equation (A3) can be rewritten as

$$T^A = \Omega^{AB} (dH)_B, \tag{A5}$$

where evaluation on $\bar{\Gamma}$ is understood. This is the standard form of Hamilton's equations of motion on a symplectic manifold. Thus we have shown that in the "time-independent" case, *the Lagrangian field theories considered in this paper can be given a Hamiltonian formulation provided only that an appropriate time evolution vector field τ^A on $\bar{\mathscr{F}}$ projects to $\bar{\Gamma}$.*

Finally, we remark that, in particular, all Lagrangian field theories that are "generally covariant" (i.e., where the field variations induced by diffeomorphisms are local symmetries) can be given a Hamiltonian formulation whenever there exists a local symmetry vector field on $\bar{\mathscr{F}}$ representing nontrivial time translations that projects from $\bar{\mathscr{F}}$ to $\bar{\Gamma}$. In such a case, the Hamiltonian H can be chosen to be a constraint function for this local symmetry and hence can be chosen to vanish on $\bar{\Gamma}$.

[1] C. J. Isham and K. V. Kuchař, Ann. Phys. (NY) **164**, 288, 316 (1985).
[2] J. L. Friedman, Commun. Math. Phys. **62**, 247 (1978).
[3] C. Crnković and E. Witten, in *Three Hundred Years of Gravitation*, edited by S. W. Hawking and W. Israel (Cambridge U.P., Cambridge, 1987).
[4] P. G. Bergmann, in *Encyclopedia of Physics*, edited by S. Flügge (Springer, Berlin, 1962), Vol. 4, p. 203.
[5] P. G. Bergmann and A. Komar, Int. J. Theor. Phys. **5**, 15 (1972).
[6] D. C. Salisbury and K. Sundermeyer, Phys. Rev. D **27**, 740, 757 (1983).
[7] S. Lang, *Differential Manifolds* (Springer, New York, 1985).
[8] N. Woodhouse, *Geometric Quantization* (Oxford U.P., Oxford, 1980).
[9] R. M. Wald, *General Relativity* (Univ. Chicago, Chicago, 1984).
[10] G. Burnett and R. M. Wald, to be published.

Superconformal algebras and Clifford algebras

Z. Hasiewicz,[a] K. Thielemans, and W. Troost[b]
Instituut voor Theoretische Fysica, Universiteit Leuven, B-3030 Leuven, Belgium

(Received 8 August 1989; accepted for publication 20 September 1989)

It is shown that superconformal Lie superalgebras contain a Clifford algebra structure. This is used to present a classification of these algebras under the following assumptions: only fields with (half-)integer conformal dimension between 2 and $\frac{1}{2}$, and a nonzero central element. It turns out that the number of supersymmetry generators must be less than or equal to 4. More generally, the so(N) series, as well as some nonassociative algebras connected to S^7, also have a natural place in this Clifford algebra framework.

I. INTRODUCTION

In conformal field theories, the Virasoro algebra plays a central role. It can be extended in a natural way to include a number (N) of supersymmetries, i.e., the superconformal algebras. It is well known that, for $N=2$, it follows from Jacobi identities that the algebra necessarily also contains a field of dimension 1, a "U(1) Kac–Moody" algebra. For $N=3$, the algebra necessarily contains three dimension-1 fields forming an SU(2) Kac–Moody algebra, and besides a fermionic field of dimension $\frac{1}{2}$, a free fermion. Also for $N=4$, these ingredients lead to superconformal Lie algebras.

In Ref. 1, a classification was initiated of conventional superconformal Lie superalgebras (CSCLS's) having precisely these characteristics. It did not include the $N=1$ algebras of Kac and Todorov.[2] Also, the algebra resulting from the $N=4$ case in Ref. 3, by replacing the derivative of the dimension-0 field by an arbitrary field of dimension 1, was not included.

In addition, another $N=4$ algebra was recently written down,[4,5] which reduces to the previous cases only for special values of a parameter. All these cases have a finite sub-Lie algebra of the Kac–Moody zero modes that is not simple, but only reductive (semisimple + Abelian). A list of finite sub-superalgebras (formed by the 0 and $\pm\frac{1}{2}$ modes and the ± 1 Virasoro modes) was given in Ref. 6, and could be used as a basis for a complete classification. The possible nonsimplicity of the Lie algebra, however, makes this a tedious and nonsimple proposition.

In this paper we uncover, for each superconformal Lie algebra, an underlying Clifford algebra structure. This Clifford algebra is represented by transformations mixing the spaces corresponding to operators of different dimensions: odd elements map fermionic to bosonic spaces, and the even elements, though respecting statistics, still mix dimensions $\frac{3}{2}$ and $\frac{1}{2}$. The Clifford algebra representation may be reducible.

The information contained in the Clifford algebra representation is not enough to reconstruct the superconformal algebra completely, nor does it guarantee it to be a Lie algebra. The additional structure required is precisely Lie algebra covariance. Imposing this on an arbitrary representation, singles out $N=4$ as a special case, and allows one to prove that $N>4$ is impossible for this conventional type of superconformal Lie algebra. Also it leads to a completion of the classification of the $N\leqslant 4$ cases. Since we work throughout the paper with algebras defined over the field $\mathbf{R}$, there are some distinctions here: for example, it is shown that, for $N=4$, with signatures (3,1) or (1,3), the "small"algebra without dimension-$\frac{1}{2}$ fields does not exist, and the "large" algebra with a SO(3,1) $\oplus$ U(1) Kac–Moody algebra exists but has no parameter.

The same methods are used to derive possible extensions of these algebras, by including more fields of dimension 1 or $\frac{1}{2}$. It turns out that only $N=1,2$ and the "small" $N=4$ algebras allow this type of extension. The extensions are, in fact, representation spaces for the basic algebras.

We then enlarge the scope of applications of the Clifford algebra representations. We show that actually all the superconformal Lie algebras in the so(N) list of Ref. 3 have a natural Clifford algebra structure, although they contain fields with dimensions smaller than $\frac{1}{2}$ for $N\geqslant 4$, and are therefore not included in our general investigation. Finally we show that also the non-Lie superalgebra for $N=8$ given in Ref. 7 follows our general pattern: it is based on an irreducible Clifford algebra $\mathscr{C}(8)$ representation, without imposing the additional covariance structure.

In Sec. II we introduce the structure that is our starting point, the conventional superconformal Lie superalgebra (CSCLS), devoting some space to making clear our assumptions. We introduce a basis-free notation, and write out the Jacobi identities. In Sec. III we deduce the existence of some exact sequences, which lead to the exposure of a Clifford algebra representation. The next section introduces a distinction between "basic" algebras (essentially the smallest that still contains all N supersymmetry generators) and enlarged ones. The consequences of the additional Lie algebra structure is explored, and lead to a no-go theorem for $N>4$. In Sec. V we describe the rich $N=4$ structure in detail, for the different signatures. In Sec. VI we treat the enlarged cases. The next two sections show that Clifford algebra methods are more widely applicable, by going beyond the CSCLS's, showing how the so(N) algebras of Ref. 3 and the nonassociative $N=8$ algebra of Ref. 7 fit in.

In Ref. 5, the classification of a class of superconformal

[a] On leave of absence from the Institute of Theoretical Physics, University of Wrocław, PL-50-205 Wrocław, Poland.
[b] Bevoegdverklaard navorser, N.F.W.O. Belgium, BITNET FGBDA19 at BLEKUL11.

 0022-2488/90/030744-13$03.00

Lie superalgebras is also studied. That class is less restrictive than the one studied in Sec. II, in particular concerning the structure of the central term. Our result in Sec. IV is compatible with the classification conjecture in Ref. 5.

One word about technique. For representations, we found it very convenient to use the method that takes (a subspace of) the Clifford algebra itself as the representation space. Readers who are unfamiliar with it can get some idea of the method from Appendix A, and more details, for example, from Ref. 8.

II. CONVENTIONAL SUPERCONFORMAL LIE SUPERALGEBRAS

The structure we call conventional superconformal Lie superalgebra corresponds roughly to that described in Ref. 1. To make this notion precise we have to fix the notation and formulate a few definitions. Here and in further considerations the object of our interest, CSCLS, will be denoted by $\mathscr{S}$. By definition $\mathscr{S}$ is a Lie superalgebra, and consequently it splits in a natural way into two subspaces:

$$\mathscr{S} = \mathscr{S}_{(0)} \oplus \mathscr{S}_{(1)}, \tag{1}$$

the even one, $\mathscr{S}_{(0)}$, being an ordinary Lie algebra, while the odd one, $\mathscr{S}_{(1)}$, carries a representation of the former. We will assume $\mathscr{S}_{(0)}$ to be the sum of some Kac–Moody Lie algebra KM($\mathscr{L}$), with $\mathscr{L}$ as underlying Lie algebra, and Virasoro algebra Vir:

$$\mathscr{S}_{(0)} = \mathrm{KM}(\mathscr{L}) + \mathrm{Vir}. \tag{2}$$

From the above it follows that we have two subalgebras contained in $\mathscr{S}_{(0)}$.

The first is the Kac–Moody subalgebra

$$[T_m(\Sigma),T_n(\Sigma')] = T_{m+n}([\Sigma,\Sigma']) - mK(\Sigma,\Sigma')\delta(m+n)c, \tag{3}$$

where $\Sigma,\Sigma' \in \mathscr{L}$ are elements of $\mathscr{L}$, the Lie algebra, which can be identified with the subalgebra of "zero modes" contained in Eq. (3). This K is an arbitrary, bilinear, symmetric form on $\mathscr{L}$ satisfying

$$K(\mathrm{ad}_\Sigma \cdot,\cdot) + K(\cdot,\mathrm{ad}_\Sigma \cdot) = 0, \tag{4}$$

i.e., it is ad_Σ invariant. This relation is a consequence of the Jacobi identity involving Eq. (3).

The second is the Virasoro subalgbra

$$[L_m,L_n] = (m-n)L_{m+n} + (m^3-m)\delta(m+n)c/4. \tag{5}$$

Notice that the central element c is common to KM($\mathscr{L}$) and Vir. Moreover KM($\mathscr{L}$) forms a linear representation of Vir,

$$[L_m,T_n(\Sigma)] = -nT_{m+n}(\Sigma), \tag{6}$$

corresponding to a conformal dimension equal to 1. To conclude our discussion of the structure of $\mathscr{S}_{(0)}$, we want to stress that no assumption of semisimplicity or reductivity of $\mathscr{L}$ has been made. Consequently, the form K is assumed neither to be positive (or negative) nor even to be nondegenerate. The only requirement is Eq. (4). The odd subspace $\mathscr{S}_{(1)}$ of $\mathscr{S}$ will be assumed to have a sum structure

$$\mathscr{S}_{(1)} = Q \oplus G, \tag{7}$$

with the subspace Q containing elements of conformal dimension $\frac{1}{2}$, while G contains elements of conformal dimension $\frac{3}{2}$. This is the same as to say that there is a fixed Virasoro algebra action on Q and G. Let V and W be the spaces of corresponding orbits of Vir. Then we can say that V and W are the underlying vector spaces generating Q and G in the same sense as $\mathscr{L}$ underlies KM($\mathscr{L}$). Hence in addition to Eq. (7) we can perform the more detailed decomposition

$$Q = \oplus_m Q_m(V), \qquad G = \oplus_m G_m(W), \quad m \in \mathbb{Z} + \alpha, \tag{8}$$

and then we have

$$[L_m,Q_n(v)] = -(m/2+n)Q_{m+n}(v), \quad v \in V, \tag{9}$$

$$[L_m,G_n(w)] = (m/2-n)G_{m+n}(w), \quad w \in W. \tag{10}$$

For the space $Q(V)$ we postulate standard fermionic anticommutation relations

$$\{Q_m(v),Q_n(v')\} = -b(v,v')\delta(m+n)c, \tag{11}$$

where b is a symmetric bilinear form on V. Moreover, we assume that $Q(V)$ is a representation for KM($\mathscr{L}$) in the following sense:

$$[T_m(\Sigma),Q_n(v)] = Q_{m+n}(R(\Sigma)v), \quad \Sigma \in \mathscr{L}, \quad v \in V, \tag{12}$$

where $R(\Sigma)$ is an endomorphism of V corresponding to $\Sigma \in \mathscr{L}$. The Jacobi identity involving the triple (T,T,Q) implies immediately that

$$\mathscr{L} \ni \Sigma \to R(\Sigma) \in \mathrm{End}\, V \tag{13}$$

is a representation of the Lie algebra $\mathscr{L}$. Similarly, the Jacobi identity for (T,Q,Q) forces b to be R invariant:

$$b(R(\Sigma)\cdot,\cdot) + b(\cdot,R(\Sigma)\cdot) = 0. \tag{14}$$

The above is the only requirement for the form b. Notice that the subspace

$$\mathscr{S}_{(0)} \oplus Q \tag{15}$$

forms a Lie subsuperalgebra of $\mathscr{S}$. We admit the possibility of V being one point (zero) only, i.e., zero dimensional. We will assume the following as the most general form of the anticommutator of two elements of G:

$$\{G_m(w),G_n(w')\} = 2B(w,w')L_{m+n} + B(w,w')(m^2-\tfrac{1}{4})\delta(m+n)c - (m-n)T_{m+n}(\varphi(w,w')), \quad w,w' \in W. \tag{16}$$

In the above formula B is a bilinear symmetric form on W. We will assume that all dimension-$\frac{3}{2}$ elements correspond to supersymmetry generators, i.e., that they "square" to the Virasoro algebra. Correspondingly, we assume that this time the form B is nondegenerate. The dimension of W is then called the number of supersymmetries. The mapping

$$W \times W \ni (w,w') \to \varphi(w,w') \in \mathscr{L} \tag{17}$$

in the last term of Eq. (16) is forced to be antisymmetric. We will call the image of this map I_φ. In order to fix the structural relations of $\mathscr{S}$ completely we need two more bilinear and one linear mapping. The first one

$$W \times V \ni (w,v) \to \psi(w,v) \in \mathscr{L} \tag{18}$$

defines the anticommutator:

$$\{G_m(w),Q_n(v)\} = T_{m+n}(\psi(w,v)), \quad w\in W, \quad v\in V. \tag{19}$$

Notice that an additional central term could be added to the right-hand side, as in Refs. 5 and 9. This would imply a nonzero expectation value for the product of two fields of dimensions $\frac{1}{2}$ and $\frac{3}{2}$, respectively. In this paper only the most conventional case is considered.

The next two mappings,

$$\mathscr{L} \ni \Sigma \to \Lambda(\Sigma) \in \mathrm{End}\, W \tag{20}$$

and

$$\mathscr{L} \times W \ni (\Sigma,w) \to d(\Sigma,w) \in V, \tag{21}$$

determine the affine covariance relation of G with respect to $\mathrm{KM}(\mathscr{L})$. Namely,

$$[T_m(\Sigma),G_n(w)] = G_{m+n}(\Lambda(\Sigma)w) + mQ_{m+n}(d(\Sigma,w)). \tag{22}$$

Obviously, the mappings introduced above are not completely arbitrary, and there is a set of (very strong) relations they must satisfy in order to define the structure of a Lie superalgebra; these follow from the Jacobi identities.

First of all, from the (T,T,G) Jacobi identity it follows that Eq. (20) is a representation of $\mathscr{L}$ on W. This is the third one (in addition to the adjoint and R) built into $\mathscr{S}$. The next observation we can make by inspection of Jacobi identities is that the form B in Eq. (16) is invariant with respect to Λ, i.e.,

$$B(\Lambda(\Sigma)\cdot,\cdot) + B(\cdot,\Lambda(\Sigma)\cdot) = 0, \quad \Sigma\in\mathscr{L} \quad (TGG). \tag{23}$$

Then there is a series of covariance relations for the bilinear mappings we have defined. We will list them with the indication of the triple, for which the Jacobi identity enforces it:

$$\mathrm{ad}_\Sigma \varphi(\cdot,\cdot) = \varphi(\Lambda(\Sigma)\cdot,\cdot) + \varphi(\cdot,\Lambda(\Sigma)\cdot) \quad (TGG), \tag{24}$$

$$\mathrm{ad}_\Sigma \psi(\cdot,\cdot) = \psi(\Lambda(\Sigma)\cdot,\cdot) + \psi(\cdot,R(\Sigma)\cdot) \quad (TGQ), \tag{25}$$

$$\mathrm{R}(\Sigma)d(\cdot,\cdot) = d(\mathrm{ad}_\Sigma\cdot,\cdot) + d(\cdot,\Lambda(\Sigma)\cdot) \quad (TTG), \tag{26}$$

where $\Sigma\in\mathscr{L}$ is arbitrary.

Those are the relations one could expect from the very beginning. The identities of the above type are typical for Lie superalgebras, and in fact they are closely related to the Lie algebra structure of the even subalgebra.

We assume in the following a nonzero central extension ($c\neq 0$). By comparing central elements, we get two relations containing the forms

$$K(\Sigma,\psi(w,v)) = b(d(\Sigma,w),v) \quad (TGQ), \tag{27}$$

$$K(\Sigma,\varphi(w,w')) = B(\Lambda(\Sigma)w,w') \quad (TGG); \tag{28}$$

$\Sigma\in\mathscr{L}$, $v\in V$, $w,w'\in W$ are arbitrary elements. Their origin indicates that they are specific for SCLS's. If one assumes[1] that, in addition to B, the forms K and b are also nondegenerate, then using the above relations one can express two of the structural mappings of $\mathscr{S}$ in terms of the others. However, this assumption is not needed in further considerations.

There are other relations imposed by Jacobi identities on the structural mappings of $\mathscr{S}$. This series, together with Eqs. (27) and (28), is crucial for the Clifford algebra structure we are introducing in the next sections:

$$\begin{aligned} 2B(w,w')\Sigma &= \psi(w,d(\Sigma,w')) + \psi(w',d(\Sigma,w)) \\ &\quad + \varphi(w,\Lambda(\Sigma)w') - \varphi(\Lambda(\Sigma)w,w') \quad (GGT), \end{aligned} \tag{29}$$

$$\begin{aligned} 2B(w,w')w'' &= B(w',w'')w + B(w,w'')w' \\ &\quad + \Lambda(\varphi(w',w''))w + \Lambda(\varphi(w,w''))w' \quad (GGG), \end{aligned} \tag{30}$$

$$d(\varphi(w,w'),w'') + d(\varphi(w,w''),w') = 0 \quad (GGG), \tag{31}$$

$$2B(w,w')v = d(\psi(w,v),w') + d(\psi(w',v),w) \quad (GGQ), \tag{32}$$

where $w,w',w''\in W$, $\Sigma\in\mathscr{L}$, $v\in V$ are arbitrary elements. Finally,

$$R(\psi(w,v))v' - R(\psi(w,v'))v = 0 \quad (GQQ), \tag{33}$$

$$\begin{aligned} 2R(\varphi(w,w'))v &= d(\psi(w,v),w') - d(\psi(w',v),w) \quad (GGQ), \end{aligned} \tag{34}$$

where $v,v'\in V$, $w,w'\in W$.

As one sees, the structure of CSCLS depends on many data. In the next section, a Clifford algebra representation will bring more order into them.

Finishing this section we will rewrite the basic definitions (for the readers who prefer notation with indices) in terms of bases of the underlying spaces $\mathscr{L}$, V, and W. Let $\{\Sigma_A\}$ be the basis in $\mathscr{L}$. Then $\{T_m(\sigma_A) := T_{mA}\}$ form a basis for $\mathrm{KM}(\mathscr{L})$. Relations (3) can be rewritten as follows:

$$[T_{mA},T_{nB}] = C^D_{AB}T_{m+nD} - mK_{AB}\delta(m+n)c, \tag{35}$$

where $K_{AB} := K(\Sigma_A,\Sigma_B)$, C^D_{AB} - structure constants of $\mathscr{L}$. Choosing a basis $\{v_\alpha\}$ in V we can form the basis $\{Q_{n\alpha} := Q_n(v_\alpha)\}$ of Q. Then Eqs. (11) and (12) can be rewritten as follows:

$$\begin{aligned} &\{Q_{n\alpha},Q_{m\beta}\} = -b_{\alpha\beta}\delta(m+n)c, \quad b_{\alpha\beta} := b(v_\alpha,v_\beta), \\ &[T_{mA},Q_{n\beta}] = R^{\alpha}_{A\beta}Q_{m+n\alpha}. \end{aligned} \tag{36}$$

Finally, choosing a basis $\{w_j\}$ in W we can form the basis $\{G_{mj} := G_m(w_j)\}$ of G. Then for Eq. (16) we can write

$$\begin{aligned} \{G_{mj},G_{ni}\} &= 2B_{ij}L_{m+n} + B_{ij}(m^2 - \tfrac{1}{4})\delta(m+n)c \\ &\quad - (m-n)\varphi^A_{ji}T_{m+nA}, \end{aligned} \tag{37}$$

$$B_{ij} = B(w_i,w_j),$$

while Eqs. (19) and (22) look as follows:

$$\begin{aligned} &\{G_{mi},Q_{n\alpha}\} = \psi^A_{i\alpha}T_{m+nA}, \\ &[T_{mA},G_{ni}] = \Lambda^j_{iA}G_{m+nj} + mD^{\alpha}_{Ai}Q_{m+n\alpha}. \end{aligned} \tag{38}$$

We are not going to rewrite the relations between the structural mappings in this notation as, first of all, we will never use them in this form, and second, they may actually hinder easy understanding when we use the different vector spaces in Sec. III.

III. EXACT SEQUENCES AND THE CLIFFORD ALGEBRA

Let us fix an arbitrary nonisotropic element $w\in W$ $[B(w,w)\neq 0]$. We can consider the bilinear map $\varphi(w,w')$ from $W\times W$ to $\mathscr{L}$ as a linear map from W to $\mathscr{L}$.

$$\varphi_w\colon W \ni w' \to \varphi(w,w') \in \mathscr{L},$$

keeping w fixed. Similary, let us define

$$d_w:\quad \mathscr{L}\ni\Sigma\to d(\Sigma,w)\in V,$$

$$i_w:\quad \mathbb{R}\ni a\to aw\in W.$$

Then we have the following exact sequence (i.e., the image of each of the maps in the sequence is the kernel of the next):

$$\mathbb{R}\xrightarrow{i_w} W\xrightarrow{\varphi_w}\mathscr{L}\xrightarrow{d_w} V\to 0.$$

Proof: At W, (i) $\varphi(w,aw)=a\varphi(w,w)=0$ by antisymmetry of φ; and (ii) $\varphi(w,w_1)=0$ implies, by Eq. (30), that we can write

$$B(w,w)w_1=B(w,w_1)w.$$

At $\mathscr{L}$, (i) $d(\varphi(w,w_1),w)=0$ by Eq. (31); and (ii) $d(\Sigma,w)=0$ implies, by Eq. (29), that we can write

$$B(w,w)\Sigma=\varphi(w,w_1),\quad \text{with } w_1=\Lambda(\Sigma)w.$$

At V, $d(\mathscr{L},w)$ generates the whole space V by Eq. (32). ■

Similary, defining, always keeping w fixed,

$$\psi_w:\quad V\ni v\to\psi(w,v)\in\mathscr{L},$$

$$\Lambda_w:\quad \mathscr{L}\ni\Sigma\to\Lambda(\Sigma)w\in W,$$

$$B_w:\quad W\ni w'\to B(w,w')\in\mathbb{R},$$

we also have the following sequence:

$$0\to V\xrightarrow{\psi_w}\mathscr{L}\xrightarrow{\Lambda_w} W\xrightarrow{B_w}\mathbb{R}.$$

Proof: At V, $\psi(w,v)=0$ implies $v=0$ by Eq. (32).

At $\mathscr{L}$, (i) $\Lambda(\psi(w,v))w=0$ by Eqs. (27), (28), and (31); and (ii) $\Lambda(\Sigma)w=0$ implies, by Eq. (29), that we can write

$$B(w,w)\Sigma=\psi(w,v)v=d(\Sigma,w).$$

At W, (i) $B(w,\Lambda(\Sigma)w)=0$ by Eq. (23); and (ii) $B(w,w_1)=0$ implies, by Eq. (30), that we can write

$$B(w,w)w_1=\Lambda(\Sigma)w,\quad \text{with } \Sigma=\varphi(w,w_1).\qquad ■$$

Note that each of these exact sequences separately implies the following relation between the dimensions of the vector spaces:

$$|W|+|V|=|\mathscr{L}|+1. \tag{39}$$

This relation was also derived in Ref. 1, under the additional assumption that in each of the spaces W, V, $\mathscr{L}$ the metric should be nondegenerate. No such condition was imposed here.

Taken together, the exact sequences suggest a combination of the maps φ_w, ψ_w, and B_w from $W\oplus V$ to $\mathscr{L}\oplus\mathbb{R}$, and combining d_w, Λ_w, and i_w to a map from $\mathscr{L}\oplus\mathbb{R}$ to $W\oplus V$.

Clifford Algebras: As a result we are led to define, for all $w_1\in W$, a map in the space $S=W+V+\mathscr{L}+\mathbb{R}$ as follows. Let w, $w_1\in W$, $v\in V$, $\Sigma\in\mathscr{L}$, $a\in\mathbb{R}$. Define the endomorphisms

$$\Gamma_w:\quad S\ni x\to\Gamma_w(x)\in S,$$

$$\begin{aligned}\Gamma_w(w_1+v+\Sigma+a)&\\ &=(aw+\Lambda(\Sigma)w)+d(\Sigma,w)\\ &\quad+(\varphi(w,w_1)+\psi(w,v))+B(w,w_1).\end{aligned} \tag{40}$$

The set of maps Γ_w, $w\in W$, satisfies the following important property.

Property: The Γ_w represent the Clifford algebra $\mathscr{C}(B)$, i.e.,

$$\Gamma_w\Gamma_{w'}+\Gamma_{w'}\Gamma_w=2B(w,w').$$

Its verification follows by direct computation, using Eqs. (23) and (29)–(32).

On the space S, we can define a metric by making a linear combination of the metrics on each of the terms, introduced previously, and taking a natural metric on $\mathbb{R}$. Let us define the metric

$$\begin{aligned}\theta(w+v+\Sigma+a,w'+v'+\Sigma'+a')&\\ &=B(w,w')+b(v,v')-K(\Sigma,\Sigma')-aa'.\end{aligned}$$

This metric is symmetric, $\theta(x,y)=\theta(y,x)$, and also has the following invariance property. Define the operator β_- in $\mathscr{C}$ by

$$\beta_-(\Gamma_w)=-\Gamma_w,$$

$$\beta_-(AB)=\beta_-(B)\beta_-(A)\quad (A,B\in\mathscr{C}).$$

Then θ is β_- invariant, i.e., $\forall x,y\in S$,

$$\theta(Ax,y)=\theta(x,\beta_-(A)y)\quad (A\in\mathscr{C}). \tag{41}$$

To prove this property, it is, in fact, sufficient to consider the special case $A=\Gamma_w$:

$$\theta(\Gamma_w x,y)+\theta(x,\Gamma_w y)=0.$$

This is easily verified using Eqs. (27) and (28). Note that the existence of this metric is not guaranteed if the central extension is zero.

The representation of $\mathscr{C}$ in S has some additional properties, which follow directly from its definition.

(i) The representation space S splits into two parts, which we call odd and even:

$$S=S_-\oplus S_+,\quad S_+=\mathscr{L}\oplus\mathbb{R},\quad S_-=W\oplus V.$$

(ii) $\mathscr{C}$ itself, as a vector space, splits into an even ($\mathscr{C}_+$) [resp. odd ($\mathscr{C}_-$)] subspace, being linear combinations of products of an even (resp. odd) number of Γ_{w_i}. We have that

$$\mathscr{C}_+S_+=S_+,\quad \mathscr{C}_-S_+=S_-,$$

$$\mathscr{C}_+S_-=S_-,\quad \mathscr{C}_-S_-=S_+.$$

Therefore, as a representation of $\mathscr{C}_+$, S is necessarily reducible in two subspaces with the same dimension. If the irreducible $\mathscr{C}$ representation splits in this way, then S may be an arbitrary sum of $\mathscr{C}$-irreducible parts. However, it can happen that an irreducible representation of $\mathscr{C}$ remains irreducible as a representation space of $\mathscr{C}_+$. In that case the smallest building block for S will consist of the sum of two irreducible $\mathscr{C}_+$ representations. This will be illustrated for $N=3$ shortly.

(iii) Γ_w maps $W\oplus V$ to $\mathscr{L}\oplus\mathbb{R}$ and $\mathscr{L}\oplus\mathbb{R}$ to $W\oplus V$. By changing the sign of one of these maps, one can obtain the Clifford algebra $\mathscr{C}(-B)$.

The presence of a Clifford algebra will be a powerful tool to analyze the superconformal algebra. It may be remembered that also in Ref. 1 a Clifford algebra was used to analyze the case when the space V is zero. That Clifford algebra has no direct relation with the one introduced here: it corresponds to the form $K(\Sigma,\Sigma')$ on $\mathscr{L}$, which they assumed to be nondegenerate. This assumption is not needed in the present framework.

To conclude this section, let us illustrate the above construction for the case $N=3$, with a compact form B. We choose a basis $\{w_i,\ i=1,2,3\}$ in W such that $B(w_i,w_j) = -\delta_{ij}$. The smallest representation of $\mathscr{C}(0,3)$ is four-dimensional. The even subalgebra $\mathscr{C}_+(0,3)$ is isomorphic to $\mathscr{C}(0,2)$, which also has a four-dimensional irreducible representation. Therefore S will consist of a direct sum of eight-dimensional representations. Let us examine a single copy. Because of Eq. (39) we have that V is one-dimensional and $\mathscr{L}$ is three-dimensional. To reconstruct the superconformal algebra, let us choose a vector r in the four-dimensional subspace S_+ to correspond to the unity in the $\mathbb{R}$ component of $\mathscr{L}+\mathbb{R}$. More explicitly, we can choose as a representation

$$\Gamma_1 = i\sigma_2\otimes 1\otimes 1,\quad i\sigma_2=\begin{pmatrix} & +1\\ -1 & \end{pmatrix},$$
$$\Gamma_2 = -\sigma_1\otimes\sigma_3\otimes i\sigma_2,\quad \sigma_1=\begin{pmatrix} & +1\\ +1 & \end{pmatrix},$$
$$\Gamma_3 = -\sigma_1\otimes i\sigma_2\otimes 1,\quad \sigma_3=\begin{pmatrix} +1 & \\ & -1\end{pmatrix},$$
$$r=\begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\0\end{pmatrix}.$$

Then it follows from the definition [Eq. (40)] that

$$w_1=\Gamma_1 r=\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\0\end{pmatrix},$$
$$w_2=\Gamma_2 r=\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix},$$
$$w_3=\Gamma_3 r=\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}1\\0\end{pmatrix}.$$

Further, by applying definition (40) again, with $x=w_i$,

$$M_3=\varphi(w_2,w_1)=\Gamma_2\Gamma_1 r=\begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix},$$
$$M_1=\varphi(w_3,w_2)=\Gamma_3\Gamma_2 r=\begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix},$$
$$M_2=\varphi(w_1,w_3)=\Gamma_1\Gamma_3 r=-\begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}1\\0\end{pmatrix},$$

exhausting the space $\mathscr{L}$. It is straightforward to see that the β_- operation corresponds to taking the transpose of the representation matrix, and that the only allowed metric θ is given by

$$\theta(Ar,Br) = -\tfrac{1}{8}\mathrm{tr}(A^T B).$$

The one-dimensional vector space V is spanned by

$$v=\Gamma_3\Gamma_2\Gamma_1 r=-\begin{pmatrix}1\\0\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix}\otimes\begin{pmatrix}0\\1\end{pmatrix}.$$

The maps $\Lambda(\Sigma)$, ψ, and d can also be read off from Eq. (40); for example,

$$\psi(M_i,v)=w_i$$

and

$$d(M_i,w_j)=\delta_{ij}v.$$

The Clifford algebra structure implies that the relations (29)–(32) are satisfied, and the invariance of the metric guarantees relations (27) and (28). The relations (24)–(26) imposed by the requirements that $\mathscr{L}$ form a Lie algebra, that Λ and R be representations of $\mathscr{L}$ in W and V, respectively, and that the maps φ, d, and ψ be $\mathscr{L}$ covariant, are at this point not guaranteed automatically. In the example at hand, it is easy to check them, resulting in

$$R(\Sigma)=0$$

and

$$[M_i,M_j]=\epsilon_{ijk}M_k,\quad i,j,k\in\{1,2,3\}.$$

IV. RECONSTRUCTION AND LIE ALGEBRA STRUCTURE

As is clear from the example above, the embedding of the different spaces W, V, and $\mathscr{L}$ can be identified if we choose a specific vector r representing the unit element from $\mathbb{R}$, by successive application of the basis elements Γ_i of $\mathscr{C}$ to r. The first application identifies the embedding of the space W,

$$w=\Gamma_w r,$$

where we make a slight abuse of notation by identifying W with its embedding in S. A second application results in the identification

$$\varphi(w,w')=\tfrac{1}{2}(\Gamma_w\Gamma_{w'}-\Gamma_{w'}\Gamma_w)r.$$

Another iteration, multiplying with $\Gamma_{w''}$, results in another vector of the space $W\oplus V$. To separate out the different parts, one needs the metric θ to make a projection on the orthogonal spaces W and V. The different parts then fix $\Lambda(\varphi(w,w'))w''$ and $d(\varphi(w,w',w))$:

$$\begin{aligned}\Lambda(\varphi)w'' &= \theta(w^i,\Gamma_{w''}\varphi)w_i,\\ d(\varphi,w'') &= \Gamma_{w''}\varphi-\Lambda(\varphi)w'',\end{aligned}\tag{42}$$

where $\varphi\in\mathscr{L}$, and $\{w_i\}$, $\{w^i\}$ are B-dual bases of W. A continuation of this procedure may or may not generate the whole space S. In the former case we will call the algebra "basic," in the latter case we will call it "enlarged." The reconstruction of the superconformal algebra can then be continued by choosing a new vector from the remaining orthogonal complement in S_+. It will then generate further vectors in the spaces V and $\mathscr{L}$ only. We will treat the enlarged case further in Sec. VI.

Lie algebra structure: To continue the reconstruction of a superconformal Lie algebra from the corresponding Clifford algebra representation, more information is needed. In particular, up to now, the Lie algebra structure reflected in Eqs. (24)–(26) has not been used fully.

Since the form B on W is invariant with respect to the action Λ of $\mathscr{L}$ on W, the matrices Λ form a subalgebra of the Lie algebra so(B). As a consequence, to each $\Sigma\in\mathscr{L}$ there exist one or more elements in $\mathscr{C}(B)$, and more specifically a unique element $\hat{\Sigma}$ in the so(B) Lie algebra $\mathscr{C}^2(B)$ [$\mathscr{C}^n(B)$ denotes the space of antisymmetrized n-tuple products of generators of $\mathscr{C}(B)$, i.e., Γ matrices], such that

$$[\hat{\Sigma},\Gamma_w]_{\mathscr{C}}=\Gamma_{\Lambda(\Sigma)w},$$

where $[\ ,\ ]_{\mathscr{C}}$ denotes the commutator in $\mathscr{C}$. From Eq. (42)

we can explicitly calculate this element corresponding to $\varphi(w,w')$:

$$\hat{\varphi}(w,w') = \tfrac{1}{8}\theta(\Gamma^i\Gamma^j r,[\Gamma_w,\Gamma_{w'}]_{\mathscr{C}}\, r)\Gamma_i\Gamma_j, \tag{43}$$

where we have abbreviated $\Gamma_{w_i} = \Gamma_i$ and $\Gamma_{w^i} = \Gamma^i$.

The covariance equations (24)–(26) now also imply the following relations.

(i) If $\Sigma_1 = [\Sigma_2,\Sigma_3]$ $(\Sigma_i \in \mathscr{L})$,

$$\text{then also } \hat{\Sigma}_1 = [\hat{\Sigma}_2,\hat{\Sigma}_3]_{\mathscr{C}}. \tag{44}$$

(Note that [,] denotes the Lie bracket in $\mathscr{L}$ whereas $[\ ,\ ]_{\mathscr{C}}$ denotes the commutator in $\mathscr{C}$.) Indeed,

$$\begin{aligned}
[\hat{\Sigma}_1,\Gamma_w]_{\mathscr{C}} &= \Gamma_{\Lambda(\Sigma_1)w}\\
&= \Gamma_{\Lambda(\Sigma_2)\Lambda(\Sigma_3)w} - \Gamma_{\Lambda(\Sigma_3)\Lambda(\Sigma_2)w} \qquad (45)\\
&= [\hat{\Sigma}_2,\Gamma_{\Lambda(\Sigma_3)w}]_{\mathscr{C}} - [\hat{\Sigma}_3,\Gamma_{\Lambda(\Sigma_2)w}]_{\mathscr{C}}\\
&= [\hat{\Sigma}_2,[\hat{\Sigma}_3,\Gamma_w]_{\mathscr{C}}]_{\mathscr{C}} - [\hat{\Sigma}_3,[\hat{\Sigma}_2,\Gamma_w]_{\mathscr{C}}]_{\mathscr{C}} \qquad (46)\\
&= [[\hat{\Sigma}_2,\hat{\Sigma}_3]_{\mathscr{C}},\Gamma_w]_{\mathscr{C}}, \qquad (47)
\end{aligned}$$

and Eq. (44) follows from the uniqueness of $\hat{\Sigma}$. In view of this property, we will in the sequel denote the commutator in $\mathscr{C}$ by [,], also.

(ii) One can identify the vector space W with the vector space Γ_w, $w \in W$. Now $\varphi(w,w')$ is an antisymmetric bilinear function on $W \times W$, so we can also view it as a linear function on the space of commutators $\Gamma_w \wedge \Gamma_{w'} = \tfrac{1}{2}[\Gamma_w,\Gamma_{w'}]$ so that, taking w, $w' \in W$ with $B(w,w') = 0$, we can write

$$\varphi(w,w') = \varphi(\Gamma_w \wedge \Gamma_{w'}).$$

The covariance equation [Eq. (24)],

$$[\Sigma,\varphi(w,w')] = \varphi(\Lambda(\Sigma)w,w') + \varphi(w,\Lambda(\Sigma)w')$$

can now be written with the help of the Lie algebra element Σ as follows:

$$\begin{aligned}
&[\Sigma,\varphi(\Gamma_w \wedge \Gamma_{w'})]\\
&\quad= \varphi(\Gamma_{\Lambda(\Sigma)w} \wedge \Gamma_{w'}) + \varphi(\Gamma_w \wedge \Gamma_{\Lambda(\Sigma)w'})\\
&\quad= \varphi([\hat{\Sigma},\Gamma_w] \wedge \Gamma_{w'}) + \varphi(\Gamma_w \wedge [\hat{\Sigma},\Gamma_{w'}])\\
&\quad= \varphi([\hat{\Sigma},\Gamma_w\Gamma_{w'}]). \qquad (48)
\end{aligned}$$

This equation allows the computation of the Lie algebra I_φ from the metric θ and commutator algebra in $\mathscr{C}$. The requirement that the φ algebra calculated in this way is indeed a Lie algebra, amounts to two conditons:

(a) antisymmetry of the bracket, and (b) Jacobi identity. The second of these conditions follows from Eq. (44), and the fact that commutators in $\mathscr{C}$ satisfy the Jacobi identities automatically. Indeed, let $[\varphi(b_1,\varphi(b_2)] = c$; then we have

$$\begin{aligned}
&[[\varphi(b_1),\varphi(b_2)],\varphi(b_3)]\\
&\quad= \varphi([\hat{c},b_3]) = \varphi([[\hat{\varphi}(b_1),\hat{\varphi}(b_2)],b_3])\\
&\quad= \varphi([\hat{\varphi}(b_1),[\hat{\varphi}(b_2),b_3]]) + \varphi([[\hat{\varphi}(b_1),b_3],\hat{\varphi}(b_2)])\\
&\quad= [\varphi(b_1),[\varphi(b_2),\varphi(b_3)]]\\
&\qquad - [\varphi(b_2),[\varphi(b_1),\varphi(b_3)]],
\end{aligned}$$

where the b_i are antisymmetric bilinear products of two Γ's.

Surprisingly, the antisymmetry of the bracket is nontrivial and leads to a condition on the form θ. Let us choose an orthonormal basis $\{w_i\}$. The antisymmetry of $[\varphi(w_i,w_j), \varphi(w_k,w_l)]$ (where $i \neq j$, $k \neq l$) is equivalent to the following equation:

$$\varphi([\hat{\varphi}(w_i,w_j),\Gamma_k\Gamma_l]) = \varphi([\Gamma_i\Gamma_j,\hat{\varphi}(w_k,w_l)]).$$

Using Eq. (43) and working out the commutators one obtains

$$\begin{aligned}
&\varphi(\Gamma^a,\Gamma_l)\theta(\Gamma_a\Gamma_k r,\Gamma_i\Gamma_j r) - \varphi(\Gamma^a,\Gamma_k)\theta(\Gamma_a\Gamma_l r,\Gamma_i\Gamma_j r)\\
&\quad= -\varphi(\Gamma^a,\Gamma_j)\theta(\Gamma_a\Gamma_i r,\Gamma_k\Gamma_l r)\\
&\qquad + \varphi(\Gamma^a,\Gamma_i)\theta(\Gamma_a\Gamma_j r,\Gamma_k\Gamma_l r), \qquad (49)
\end{aligned}$$

where a sum over a is implied. If we specify that the two pairs (i,j) and (k,l) have an index in common, Eq. (49) becomes trivial, by using

$$\begin{aligned}
\theta(\Gamma_a\Gamma_b r,\Gamma_a\Gamma_c r) &= -\theta(\Gamma_b r,(\Gamma_a)^2\Gamma_c r)\\
&= +B_{aa}\theta(r,\Gamma_b\Gamma_c r)\\
&= B_{aa}B_{bc}\theta(r,r),
\end{aligned}$$

which vanishes if $b \neq c$, so that for $N \leqslant 3$ we have nothing to prove. Let us take i, j, k, l all different. The terms where a takes on a value from the set $\{i,j,k,l\}$ cancel one another, so that, also for $N = 4$, Eq. (49) is always satisfied. For $N > 4$, one can take the scalar product of the vector above with the vector $\Gamma_i\Gamma_j\Gamma_k\Gamma_l r$, to obtain

$$\begin{aligned}
0 = \sum_{a \notin \{i,j,k,l\}} &(\theta(\Gamma_a\Gamma_k r,\Gamma_i\Gamma_j r)^2 B_{ll}\\
&+ \theta(\Gamma_a\Gamma_l r,\Gamma_i\Gamma_j r)^2 B_{kk}\\
&+ \theta(\Gamma_a\Gamma_i r,\Gamma_k\Gamma_l r)^2 B_{jj}\\
&+ \theta(\Gamma_a\Gamma_j r,\Gamma_k\Gamma_l r)^2 B_{ii}). \qquad (50)
\end{aligned}$$

This forces

$$\theta(\Gamma_a\Gamma_k r,\Gamma_i\Gamma_j r) = 0 \quad (a,k,i,j \text{ all different}). \tag{51}$$

If the form B has a definite signature, this is obvious. For nondefinite signature, the argument is more involved, and can be found in Appendix A. As a result, for $N > 4$, expression (43) becomes

$$\hat{\varphi}(w,w') = \tfrac{1}{2}[\Gamma_w,\Gamma_{w'}] \quad (N > 4) \tag{52}$$

and the Lie algebra I_φ is isomorphic to so(B). Let us continue the discussion for $N > 4$, postponing $N = 4$ to the next section. Then Eq. (52) means that the map "hat" is the identity, i.e., it maps each bivector (viewed as an element of the antisymmetric tensor product of W) to itself [viewed as an element of $\mathscr{C}^2(B)$]. Consequently, the map Λ is the standard vector representation of so(B) in W.

From the determination of θ in Appendix A and Eq. (51) it follows that the space $\mathscr{C}^3(B)r \subset S_-$ is orthogonal to W, and that $\mathscr{C}^4(B)r$ is orthogonal to $I_\varphi \sim$ so(B). This leads to a contradiction, as we now show.

As a first step, let us examine the R representation of so(B) on $\mathscr{C}^3(B)r$. Let P_3 denote the orthogonal projection of S_- onto this subspace. For any $\varphi \in$ so(B) and $w \in W$, we have

$$d(\varphi,w) = P_3(\hat{\varphi}\Gamma_w r) = \tfrac{1}{2}[\hat{\varphi},\Gamma_w]_+ r, \tag{53}$$

where $\tfrac{1}{2}[\hat{\varphi},\ \Gamma_w]_+ = \tfrac{1}{2}(\hat{\varphi}\Gamma_w + \Gamma_w\hat{\varphi})$ is exactly the three-vector component of the product $\hat{\varphi}\Gamma_w$. Obviously, any element of $\mathscr{C}^3(B)r$ can be written in the form (53). Now we rewrite the covariance relation, Eq. (26), for arbitrary $\varphi' \in \mathrm{so}(B)$,

$$R(\varphi')d(\varphi,w) = d([\varphi',\varphi],w) + d(\varphi,\Lambda(\varphi')w),$$

in terms of [Eq. (53)]

$$R(\varphi')\tfrac{1}{2}[\hat{\varphi},\Gamma_w]_+ r = \tfrac{1}{2}[[\hat{\varphi}',\hat{\varphi}],\Gamma_w]_+ r + \tfrac{1}{2}[\hat{\varphi},[\hat{\varphi}',\Gamma_w]]_+ r. \tag{54}$$

This assures us that R is just the natural representation of $\mathrm{so}(B)$ on three-vectors induced by the one on W. A similar reasoning allows us to write any element of $\mathscr{C}^4(B)r$ in terms of the ψ mapping by using the spaces $\mathscr{C}^3(B)$ and W. Take an arbitrary element $v = ar$ of $\mathscr{C}^3(B)r$. Then we can write

$$\psi(w,v) = \Gamma_w ar. \tag{55}$$

It is obvious that any element of $\mathscr{C}^4(B)$ can be written in this way. The covariance formula (25) applied to Eq. (55) implies, using Eq. (54), that

$$[\hat{\varphi},\psi(w,v)] = \mathrm{ad}_\varphi \psi(w,v) = [\hat{\varphi},\Gamma_w a]r, \tag{56}$$

which in turn means that the representation of $\mathrm{so}(B)$ on four-vectors is again a natural one (i.e., induced on four-vectors by Λ).

This representation is never trivial, unless dim $W \leqslant 4$. In the case of strict inequality the representation space is zero, while for dim $W=4$ it is spanned by a single element—the unit pseudoscalar.

Let us now show, as a second step, that the action of the elements of $\mathscr{C}^4(B)r$ must be trivial on W, i.e., $\Lambda(\mathscr{C}^4(B)r) \equiv 0$. For any element $l \in \mathscr{C}^4(B)r$, we have

$$\Lambda(l)w = P_W(\Gamma_w l), \quad w \in W, \tag{57}$$

where P_w denotes the projection onto the W subspace of S_-. However, since $\mathscr{C}^4 r \perp I_\varphi$, we see that $P_W(\Gamma_w l) = 0$ and consequently $\Lambda(l) \equiv 0$. This in turn implies

$$[\varphi,l] = 0, \tag{58}$$

for any $\varphi \in \mathrm{so}(B)$.

When the right-hand side of Eq. (56) is not identically zero, which is always the case for dim $W > 4$, we are led to a contradiction with Eq. (58). Hence we have the following proposition.

Proposition 4.1: For dim $W > 4$, CSCLS cannot exist. The theorem we formulated above means that the structure, which contains more than four two-dimensional conformal conventional supersymmetries, does not exist in the category of Lie superalgebras. However, this statement does not mean that such description is completely impossible. In Sec. VIII, we indicate the direction of possible generalization to a category that could make such a description possible.

V. $N=4$

In this section, an explicit description is given of the basic algebras for $N=4$. Since we are discussing algebras over $\mathbb{R}$, we have to distinguish the different signatures of the form B.

Let us first take a negative definite form B. The $\mathscr{C}(0,4)$ contains one element $J = \Gamma_1\Gamma_2\Gamma_3\Gamma_4$ with square equal to 1. There are then two inequivalent constructions, depending on the choice of the 8- or 16-dimensional representation. Taking as our representation space the Clifford algebra itself (see Appendix A), these correspond to the choices $r = (1+J)$ or $r = 1$. We continue for the moment with the larger one. The element determining the metric has the general form $\Theta = -1 - \alpha J$, where α is an arbitrary real parameter. With a choice of basis in $\mathscr{C}$, $\Gamma_i\Gamma_j + \Gamma_j\Gamma_i = -2\delta_{ij}$, we can make the following identification:

$$r = 1,$$

$$w_i = \Gamma_i \quad \text{(choice of orthogonal basis in } W\text{)},$$

$$\varphi_{ij} = \varphi(w_i,w_j) = \Gamma_i\Gamma_j \quad (i \neq j).$$

The V-space is identified as the odd subspace orthogonal to W. We can take as a basis

$$v_i = \Gamma_i(J - \alpha),$$

since $\mathrm{tr}(\Theta\Gamma_j\Gamma_i(J-\alpha)) = 0$. The basis is completed by the element

$$\mathscr{L} \ni \sigma = J - \alpha.$$

To exhibit the maps d, ψ, Λ, and R explicitly, we have only to make an orthogonal decomposition

$$\Gamma_k\varphi_{ij} = \Gamma_k\Gamma_i\Gamma_j = \delta_{jk}\Gamma_i - \delta_{ik}\Gamma_j + \epsilon_{ijkl}\Gamma_l J = \delta_{jk}\Gamma_i - \delta_{ik}\Gamma_j + \alpha\epsilon_{ijkl}\Gamma_l + \epsilon_{ijkl}\Gamma_l(J-\alpha),$$

to read off that

$$\Lambda(\varphi_{ij})w_k = \delta_{jk}w_i - \delta_{ik}w_j + \alpha\epsilon_{ijkl}w_l, \tag{59}$$

$$d(\varphi_{ij},w_k) = \epsilon_{ijkl}v_l. \tag{60}$$

Similarly,

$$\psi(w_i,v_j) = \Gamma_i\Gamma_j(J-\alpha) = -\alpha\varphi_{ij} - \tfrac{1}{2}\epsilon_{ijkl}\varphi_{kl} - \delta_{ij}\sigma,$$

and

$$d(\sigma,w_k) = v_k. \tag{61}$$

Furthermore, Eq. (34) gives

$$R(\varphi_{ij})v_k = \delta_{jk}v_i - \delta_{ik}v_j + \alpha\epsilon_{ijkl}v_l.$$

Since $\Lambda(\sigma)w_k = 0$, and also $R(\sigma)w_k = 0$ from Eqs. (61) and (26), we can conclude that the representation R is isomorphic to Λ.

Finally, one should determine the Lie algebra structure of $\mathscr{L}$. First of all, $\Lambda(\sigma) = 0$ implies $[\sigma,\mathscr{L}] = 0$. Second, the structure of I_φ, which, according to Eq. (24), is an ideal in L, can be most easily recognized if one considers the six linear combinations

$$\varphi_{ij} \pm \tfrac{1}{2}\epsilon_{ijkl}\varphi_{kl}.$$

Namely, we have

$$\Lambda(\varphi_{ij} \pm \tfrac{1}{2}\epsilon_{ijkl}\varphi_{kl})w_n = (1 \mp \alpha)(\delta_{jn}w_i - \delta_{in}w_j \mp \epsilon_{ijnm}w_m), \tag{62}$$

and for $\alpha^2 \neq 1$, $\mathscr{L}$ is the algebra $\mathrm{so}(4) \simeq \mathrm{so}(3) \oplus \mathrm{so}(3)$. The metric K is proportional to the Cartan–Killing metric, multiplied by $1 \mp \alpha$. On the space V, on the other hand, we have

$$b(v_i,v_j) = +\mathrm{tr}((1+\alpha J)(J-\alpha)\Gamma_i\Gamma_j(J-\alpha)) = -\delta_{ij}(1-\alpha^2).$$

Summarizing, the structure of this "large" $N=4$ algebra is the same as that of Ref. 4, with $\alpha = 1 - 2\gamma$.

Now we examine different signatures of B. For the signatures (4,0) and (2,2), the above construction can be taken over without changes other than some signs, so we will not repeat the formulas. The resulting Lie algebra in the case (2,2) contains so $(2,2) \simeq \mathrm{so}(2,1) \oplus \mathrm{so}(2,1)$. For the signatures (1,3) and (3,1), however, the element J has square minus one. As discussed in Appendix A, this implies that the parameter α disappears from the metric, the only possible choice being $\Theta = -1$. Note that the Lie algebra in these cases contains so(3,1), which is simple over $\mathbb{R}$.

Returning to the compact case, we can examine what happens in the special case $\alpha = 1$. In that case, half of I_φ, viz., $\varphi_{ij}(J-1)$, is represented trivially on W [see Eq. (62)]. In fact, the algebra reduces to the $N=4$ superconformal subalgebra without dimension-$\frac{1}{2}$ fields, and a representation of this subalgebra by four fields of dimension $\frac{1}{2}$ and four fields of dimension 1 forming an Abelian algebra. The metric on this part of S vanishes. A slightly more general representation of this $N=4$ subalgebra can also be obtained by taking a contraction of the $\alpha \neq 1$ algebra in the following way. Consider the three self-dual combinations

$$L_{ij} = (1-\alpha)^{-1/2}(\varphi_{ij} + \tfrac{1}{2}\epsilon_{ijkl}\varphi_{kl}) \quad (\alpha < 1),$$

and write down the commutation relations with this basis. Then take the limit $\alpha \to 1$, keeping L_{ij} fixed. The element σ and the basis of V also have to be rescaled in this limit. The result is again that one obtains the $N=4$ subalgebra, plus a representation with four fields each of dimensions 1 and $\frac{1}{2}$, with the difference that now the metric does not vanish.

This contraction is of interest for two purposes. First, it relieves us of the task of constructing the "small" $N=4$ algebra, given by the eight-dimensional representation obtained by taking $r = 1 + J$, $\Theta = -1 - J$, since this construction obviously leads to the $N=4$ subalgebra for $\alpha = 1$ in the case above. Second, it, in fact, corresponds to an extension of the basic small $N=4$ algebra, in the sense of Sec. VI.

To conclude this section, let us mention what happens for the other signatures. Again for signatures (0,4), (2,2), and (4,0) the situation is completely analogous. For (3,1) and (1,3), however, the parameter α could only take the value zero. Therefore, as real algebras, there is no "small" $N=4$ algebra for these signatures.

VI. EXTENSIONS

We now return to the case of enlarged algebras. We recall that the basic algebras are generated by $\mathscr{C} r$, acting with the Clifford algebra on a single vector of the representation space S. The enlarged algebras are obtained when S is larger than $\mathscr{C} r$. One can then take another vector $r^{(1)}$, orthogonal to $\mathscr{C} r$, and consider the space $\mathscr{C} r^{(1)}$. It is important to realize that then

$$\theta(\mathscr{C} r^{(1)}, \mathscr{C} r) = 0$$

and consequently $\mathscr{C} r^{(1)}$ contains elements in $\mathscr{L}$ and V only, not in W. Note that this conclusion cannot be drawn (and is, in fact, not true) if the central extension is zero, since then the metric θ is not defined. In the general case, this step may have to be reiterated with vectors $r^{(2)}, r^{(3)}, \ldots$, leading to components $\mathscr{C} r^{(1)}, \mathscr{C} r^{(2)}, \ldots$, which are all mutually orthogonal. We now examine the structure of one of these components. Let $l, l', \ldots \in \mathscr{L} \cap \mathscr{C} r^{(1)} = L^{(1)}$ be new Lie algebra elements, and $u, u', \ldots \in V \cap \mathscr{C} r^{(1)} = U^{(1)}$ be new elements corresponding to the operators of dimension $\frac{1}{2}$. The following properties are then easily verified.

(i) $\Lambda(l) = 0$ since $\Gamma_w l$ is orthogonal to W for all w.

(ii) $[l, \varphi(w,w')] = 0$ from covariance, Eq. (24). More generally, covariance implies that l commutes with the whole basic algebra: $R(l) d(\varphi(w,w'),w'') = 0$, etc.

(iii) If $u = d(w,l)$, $R(\varphi)u = d(\Lambda(\varphi)w,l)$ from covariance, Eq. (26), for all $\varphi \in I_\varphi = \{\varphi(w,w') \mid w,w' \in W\}$. This implies that, as an I_φ module, U splits into a sum of representations that are isomorphic to the representation Λ on W.

This third property excludes the possibility of enlarging the $N=3$ basic algebra, since the dimension of U is not a multiple of 3. It also excludes enlarging the big $N=4$ algebras, but for a more subtle reason. To see this, choose some fixed l, and take

$$u_i = d(l,w_i) = \Gamma_i l,$$
$$\psi(w_j,u_i) = \Gamma_j u_i = \Gamma_j \Gamma_i l.$$

We can now compute the R representation in two ways:

$$\begin{aligned} R(\varphi_{ij})u_k &= d(l, \Lambda(\varphi_{ij})w_k) \\ &= \delta_{jk}u_i - \delta_{ik}u_j + \alpha\epsilon_{ijkl}u_l, \end{aligned}$$

by Eq. (42); and, on the other hand, by Eq. (34),

$$\begin{aligned} R(\varphi_{ij})u_k &= d(\psi(w_i,u_k),w_j) \\ &= \Gamma_j \Gamma_i \Gamma_k l \\ &= \delta_{jk}u_i - \delta_{ik}u_j - \epsilon_{ijka}\Gamma_a J l. \end{aligned}$$

This shows that there is an inconsistency unless

$$Jl = -\alpha l.$$

As a result, the (3,1) and (1,3) signatures cannot be enlarged, since in these cases $\alpha = 0$ and $J^2 = -1$. The other signatures have $J^2 = +1$, so they can only be enlarged if $\alpha^2 = 1$. But this corresponds precisely to the cases where the "large" $N=4$ algebra reduces to the "small" one.

The remaining possibilities ($N=1,2$ and $N=4$, small algebras) all allow for extensions, for $N=4$ this follows from Sec. V. Let $\mathscr{L}$ and U be the subspaces of $\mathscr{L}$ and V corresponding to the extension, i.e., orthogonal to the basic algebra, and let $l, l' \in L$. Then

$$R(l)d(l',w) = d([l,l'],w),$$

so that U carries the adjoint representation of $\mathscr{L}$. The only thing that remains to be specified is the Lie algebra structure of $\mathscr{L}$ and the allowed metric on $\mathscr{L}$. The metric on U follows from that on $\mathscr{L}$.

For $N=1$, we can be short. To enlarge the algebra, one takes a number of copies of the two-dimensional representation, which splits in even and odd subspaces of dimension 1. There are no restrictions on the Lie algebra of $\mathscr{L}$. The metric should be $\mathscr{L}$ invariant and symmetric, but otherwise arbitrary. These algebras can be found in Ref. 2.

For $N=2$, we distinguish the compact and the noncompact cases. For the noncompact case, the smallest representation is two dimensional and can be constructed in $\mathscr{C}$ with the help of $\Gamma_1\Gamma_2$, since $(\Gamma_1\Gamma_2)^2 = +1$ in this case (see Appendix A). We obtain that the extension can be made with $l^- = 1 - \Gamma_1\Gamma_2$, leading to a pair $l^-\in L$, $u^-\in U$ and the relations

$$d(l^-,w_1) = -d(l^-,w_2) = u^-,$$
$$\psi(w_1,u^-) = \psi(w_2,u^-) = l^-,$$
$$R(\varphi_{12})u^- = -u^-.$$

The other two-dimensional representation, constructed on $l^+ = 1 + \Gamma_1\Gamma_2$, gives similar results:

$$d(l^+,w_1) = d(l^+,w_2) = u^+,$$
$$\psi(w_2,u^+) = -\psi(w_2,u^+) = l^+,$$
$$R(\varphi_{12})u^+ = u^+,$$

and it is now clear that the space U splits according to the eigenvalue of $R(\varphi_{12})$. Each of these subspaces (U^+ and U^-) can be multidimensional. The space $\mathscr{L}$ follows this splitting, $L^\pm = \psi(U^\pm, w_1)$. Choosing bases, one can write for the representation of $\mathscr{L}$ on Q that

$$R(l_i^+)u_j^+ = f_{ij}^{+\,k}u_k^+,$$

where on the right-hand side there are no terms with u_k^- due to the fact that φ_{12} and $\mathscr{L}$ commute. Since R should be the adjoint of $\mathscr{L}$, we conclude that $\mathscr{L}$ is the direct sum of two ideals $L^+ \oplus L^-$, i.e., $[L^+,L^-] = 0$.

To complete the description, we have to specify the possible forms of the metric. Invariance for φ_{12} gives $\theta(U^+,U^+) = \theta(U^-,U^-) = 0$ and consequently also $\theta(L^+,L^+) = \theta(L^-,L^-) = 0$. In contrast, $\theta(L^+,L^-)$ may be different from zero. It is further restricted by noticing that $\theta([l_1^+,l_2^+],l^-) = -\theta(l_2^+,[l_1^+,l^-]) = 0$, i.e., the metric vanishes on the derived algebra $[L,L]$. This property is similar to a result obtained in Ref. 10 for $N=2$ Kac–Moody superalgebras. This completes the description of the $N=2$ case with signature $(1,1)$. The structure of the resulting enlarged superconformal Lie algebra is exhibited in Appendix B, in an isotropic basis.

For the compact $N=2$ cases, there is no essential distinction between $(0,2)$ and $(2,0)$. The smallest representation we can use is four dimensional, and is constructed in $\mathscr{C}$ on a vector $l_1 = 1$. The structure that results is then [for signature $(0,2)$]

$$d(l_1,w_1) = u_1, \quad \psi(w_1,u_1) = -l_1,$$
$$d(l_1,w_2) = u_2, \quad \psi(w_2,u_1) = -l_2,$$
$$d(l_2,w_1) = -u_2, \quad \psi(w_1,u_2) = l_2,$$
$$d(l_2,w_2) = u_1, \quad \psi(w_2,u_2) = l_1,$$
$$R(\varphi_{12})u_1 = -u_2,$$
$$R(\varphi_{12})u_2 = u_1.$$

The last two equations show that the action of $R(\varphi_{12})$ on U defines a complex structure. The map from U to $\mathscr{L}$ provided by the operation $\psi(w,\cdot)$ ($w\in W$ is fixed) then transfers this complex structure to the Lie algebra $\mathscr{L}$:

$$I_{12}l = \psi(w,R(\varphi_{12})d(w,l))/B(w,w), \quad (I_{12})^2 = -1. \tag{63}$$

From $[\varphi_{12},l] = 0$ and $\Lambda(l)W = 0$ it follows that this complex structure is a Lie complex structure, i.e., it commutes with the Lie bracket:

$$I_{12}[l,l'] = [l,I_{12}l'].$$

The invariance of the form θ implies the statement that on $\mathscr{L}$ the metric is Hermitian with respect to the complex structure:

$$\theta(I_{12}l,I_{12}l') = \theta(l,l').$$

As a result, the algebra $\mathscr{L}$ becomes a complex Lie algebra. If we extend the field $\mathbb{R}$ over which it is defined to $\mathbb{C}$, we can diagonalize the complex structure, and the resulting description follows the same pattern as the noncompact case: the eigenspaces L^{+i} and L^{-i} of I_{12} each form a separate Lie algebra, the two parts are complex conjugates to each other, and commute. The metric satisfies $\theta(L^{+i},L^{+i}) = 0 = \theta(L^{-i},L^{-i})$ and, in addition, also vanishes on the derived algebra.

Note the difference between the indefinite and definite signatures: in the former case, the extension proceeds in units of one boson and one fermion field, the representations of φ_{12} being one dimensional. In the latter case, it proceeds in units of two boson and two fermion fields, with two-dimensional representations being used.

The construction of enlarged $N=4$ algebras proceeds in the same way as for $N=2$. Recall that the basic $N=4$ algebras that can be extended are the small ones. The Lie subalgebra contained in this small subalgebra is three dimensional.

We illustrate the construction by treating in detail the extension of the $(2,2)$ superalgebra. The other cases are similar.

Let us take $1 = \Gamma_1^2 = \Gamma_2^2 = -\Gamma_3^2 = -\Gamma_4^2$, and let the relations in the basic algebra be $\varphi_{12} + \varphi_{34} = \varphi_{13} + \varphi_{24} = \varphi_{14} - \varphi_{23} = 0$. The minimal extensions can be constructed as representations of $\mathscr{C}(2,2)$ by constructing left ideals on $l^- = (1 - \Gamma_1\Gamma_3)(1 - \Gamma_2\Gamma_4)$ or $l^+ = (1 + \Gamma_1\Gamma_3)\times(1 + \Gamma_2\Gamma_4)$. (The mixed cases are not compatible with the relations in the basic algebra.) We illustrate the case l^-. The representation is four dimensional. The even subspace $\mathscr{L}$ is spanned by $l_1 = l^-$ and $l_2 = \Gamma_1\Gamma_2 l^-$. The odd subspace U is spanned by $u_1 = \Gamma_1 l^- = -\Gamma_3 l^-$ and $u_2 = \Gamma_2 l^- = -\Gamma_4 l^-$. The following tables summarize the structural maps:

d	w_1	w_2	w_3	w_4
l_1	u_1	u_2	$-u_1$	$-u_2$
l_2	u_2	$-u_1$	u_2	$-u_1$
ψ				
u_1	l_1	$-l_2$	l_1	l_2
u_2	l_2	l_1	$-l_2$	l_1

R	u_1	u_2
φ_{12}	u_2	$-u_1$
φ_{13}	$-u_1$	u_2
φ_{14}	$-u_2$	$-u_1$

Clearly, U carries the real spin-$\frac{1}{2}$ representation of the $O(2,1)$ algebra. In fact, $R(\varphi)$ acts as a representation of the Clifford algebra $\mathscr{C}_+(2,2) \simeq \mathscr{C}(2,1)$: this is the generalization of the complex structure [$\mathscr{C}_+(2,0) \simeq \mathscr{C}_+(0,2) \simeq \mathscr{C}(0,1)$] or $\mathscr{C}_+(1,1) \simeq \mathscr{C}(1,0)$ algebras represented in U for $N=2$. Again, this can be extended to anticommuting operators I_{12}, I_{13}, I_{14} on $\mathscr{L}$, as in Eq. (63), which commute with the adjoint action in $\mathscr{L}$. As a consequence, $\mathscr{L}$ must be Abelian:

$$I_{12}I_{13}[l,l'] = I_{12}[I_{13}l,l'] = [I_{13},I_{12}l'],$$
$$I_{13}I_{12}[l,l'] = I_{13}[l,I_{12}l'] = [I_{13}l,I_{12}l'].$$

Anticommutation of I_{12} and I_{13} proves that $[l,l'] = 0$. This remains valid for representations that are sums of spaces isomorphic to $\mathscr{C}l^-$ (given above) or to $\mathscr{C}l^+$ (which simply amounts to a change of basis $l_1 \leftrightarrow l_2$).

The construction of a possible metric proceeds as in the case of the basic algebra, but now there is no privileged element r. Its form is given by

$$\theta(Al^-,Bl^-) = \mathrm{tr}(Ml^+\beta_-(A)Bl^-), \tag{64}$$

since $\beta_-(l^-) = l^+$ and the metric is β_- invariant. The arguments in Appendix A lead to the result that M is of the form $a + b\Gamma_1\Gamma_2\Gamma_3\Gamma_4$, and then θ vanishes on this minimal extension. The general extension of the basic $N=4$ algebra is a direct sum of copies of the minimal extension described above. The metric vanishes on the subspaces

$$U_1 = \{u_1 | R(\varphi_{13})u_1 = -u_1\}$$

and

$$U_2 = \{u_2 | R(\varphi_{13})u_2 = +u_2\},$$

but can couple vectors between these spaces. The corresponding metric on the $\mathscr{L}$ space then follows by β_- invariance of θ [see Eq. (41)].

The structure of the resulting enlarged superconformal Lie algebra is exhibited in components in Appendix B.

The extension of the compact $N=4$ algebras follows the same pattern, but the extension proceeds with eight-dimensional units, containing four fermionic and four bosonic fields, each being a representation space of $\mathscr{C}_+(0,4) \simeq \mathscr{C}_+(4,0) \simeq \mathscr{C}(0,3)$ with three complex structures.

VII. so(N) ALGEBRAS

In this section, we consider a type of superconformal Lie algebras more general than in Sec. II. We shall still limit ourselves to conformal fields with integer or half-integer dimensions, but also allow dimension $h \leqslant 0$. The aim is to show that the well known so(N) series[3] also has a natural formulation in terms of the Clifford algebra $\mathscr{C}(N)$.

In the algebra with N supersymmetry generators,[3] one has fields with conformal dimension 2 down to $2 - N/2$. The dimension-1 fields form a so(N) Kac–Moody algebra. We consider the real form of these algebras [which can be easily derived from the original one by multiplying the operators J^R (which have conformal dimension $2 - R/2$) with $i^{(-1/2)R(R-1)}$]. The number of operators in these algebras is 2^N, and this corresponds to the dimension of a Clifford algebra with N generators. This suggests replacing the sets of antisymmetric indices in the notation of Ref. 3 with products of γ matrices. This has the advantage that it is now easy to introduce a different metric on the Clifford algebra (which corresponds to B above).

If we denote by γ^R an antisymmetric product of R γ matrices, the algebra becomes, in this new form,

$$[J^R_m(\gamma^R),J^S_n(\gamma^S)]_{RS+1}$$
$$= \begin{cases} a_{R,S}(m(2-S) - n(2-R))J^{R+S}_{m+n}(\gamma^R\gamma^S), & k=0,\\ b_{R,S}J^{R+S-2}_{m+n}(\gamma^R\gamma^S), & k=1,\\ 0, & k \geqslant 2, \end{cases} \tag{65}$$

with k the number of γ's common in γ^R and γ^S.

Note that if one writes this in operator produce expansion form, one sees that terms with $k \geqslant 2$ are nonsingular, so they do not appear in the commutation rules.

Jacobi identities now put restrictions on $a_{R,S}$ and $b_{R,S}$. One can prove that by rescaling the operators, one can put them in the simple form $a_{R,S} = 1$ and $b_{R,S} = \pm 1$. However, changing the sign of b is equivalent to changing the metric in the Clifford algebra, so this does not give a new freedom.

It is also possible to write these algebras in the same framework as before. This gives rise to a general formulation of SCA's which includes the case of Sec. II. To this end, we introduce for each R a vector space V_R with elements v_R, and the corresponding conformal fields $J^R(v_R)$. We put $V_0 = \mathbf{R}$ to have only one dimension-2 field. Defining the two sets of maps

$$\varphi^{R,S}\colon\ V_R \otimes V_S \to V_{R+S},$$
$$\psi^{R,S}\colon\ V_R \otimes V_S \to V_{R+S-2},$$

we can now write down the general form:

$$[J^R_m(v_R),J^S_n(v_S)]_{RS+1}$$
$$= (m(2-S) - n(2-R))J^{R+S}_{m+n}(\varphi(v_R,v_S))$$
$$+ J^{R+S-2}_{m+n}(\psi(v_R,v_S)).$$

Putting $\varphi^{R,0}(v_R,1) = v_R$ and $\psi^{R,0}(v_R,1) = 0$, it follows that $J^0 = 2L$, and J^R are primary fields with conformal dimension $2 - R/2$. Jacobi identities imply the following relations for φ and ψ:

$$\varphi(v_R,\varphi(v_S,v_T)) = (-1)^{RS}\varphi(v_S,\varphi(v_R,v_T)),$$
$$(R+S-4)\varphi(\psi(v_R,v_S),v_T)$$
$$= \{(S-2)\psi(v_R,\varphi(v_S,v_T))$$
$$+ (R-2)\varphi(v_R,\psi(v_S,v_T))\}$$
$$+ (-1)^{RS+1}\{R \leftrightarrow S\},$$
$$\psi(\psi(v_R,v_S),v_T)) = \psi(v_R,\psi(v_S,v_T))$$
$$+ (-1)^{RS+1}\psi(v_S,\psi(v_R,v_T)).$$

We can now, as before, define the maps Γ_w on the sum $\mathscr{S}$ of all the vector spaces:

$$\Gamma_w\colon\ S \ni x \to \Gamma_w(x) \in S,\quad w \in V_1,$$
$$\Gamma_w(v_R) = \varphi^{1,R}(w,v_R) + \psi^{1,R}(w,v_R).$$

Again, the set of maps Γ_w $w \in V_1$, satisfies the following important property.

Property: The Γ_w represent the Clifford algebra $\mathscr{C}(\psi^{1,1})$.

This follows from all Jacobi identities with $R = S = 1$.

To conclude this section, we note that one could identify the construction in Eq. (65) with the singular part of an operator product expansion. It may be interesting to extend our method to the full OPE, where nonsingular terms could correspond to multiple contractions, $k \geqslant 2$ in Eq. (65).

VIII. $N>4$ SUPERCONFORMAL (NOT LIE) SUPERALGEBRAS

The formalism we developed indicates a possible generalization of the notion of two-dimensional superconformal algebra. To go beyond the limit $N=4$, we have to violate the covariance relations, especially that of Eq. (24), which led us directly to the no-go theorem.

The above means, in turn, that we are forced to leave the category of Lie superalgebras as too constraining. It is not completely clear to us how to identify the category in which we should work, but some concrete examples point to a generalization related to symmetric spaces.

Let us notice that, if we take $r\in S_+$ as a generating element of the superalgebra, then the fundamental mapping we used,

$$\mathscr{C}^2(B)\ni(\Gamma_w\wedge\Gamma_{w'})\to\varphi(w,w'):=\tfrac{1}{2}[\Gamma_w,\Gamma_{w'}]r\in S_+, \quad (66)$$

can have (this is the general case) a nontrivial kernel depending on the rank of this element in $\mathscr{C}(B)$. It is not difficult to see that this kernel is a Lie subalgebra of $\mathrm{so}(B)$ (identified as a commutator algebra of bivectors). The image of Eq. (66) can thus be identified with the coset space

$$\mathrm{so}(B)/\mathrm{Ker}\ \varphi. \quad (67)$$

We cannot expect that Eq. (67) admits a natural Lie algebra structure unless $B=(4,0)$, $(0,4)$, $(2,2)$ [those are the only signatures for which $\mathrm{so}(B)$ is not simple, and as we noticed, the existence of the "small" $N=4$ superalgebra is directly related to the nonsimplicity of $\mathrm{so}(4)$ and $\mathrm{so}(2,2)$].

Surprisingly the kernels of Eq. (66) (we assume B to be Euclidean here), do form a uniform series, parametrized by four non-negative integers:

$$\mathrm{Ker}\ \varphi\sim\mathrm{su}(2)^m\oplus\mathrm{su}(3)^k\oplus g(2)^l\oplus\mathrm{so}(7)^r, \quad (68)$$

where g^s stands for the direct sum of s copies of the Lie algebra g. The corresponding coset spaces of the Lie algebras are the following:

$$K(N|m,k,l,r):=\mathrm{so}(N)/\mathrm{su}(2)^m\otimes\mathrm{su}(3)^k\otimes g(2)^l\oplus\mathrm{so}(7)^r, \quad (69)$$

where $N\geqslant 4m+6k+7l+8r$.

Because S is a spin representation space, the corresponding group cosets are not coset spaces of $\mathrm{SO}(N)$, but rather of its twofold (hence universal in the Euclidean case) covering $\mathrm{Spin}(N)$.

In order to illustrate what we have in mind, let us work out two examples corresponding to $N=7$ and $N=8$.

(i) $N=8$: Spin (8)/Spin (7). The element $r\in\mathscr{C}_+(8,0)$ generating the irreducible representation of $\mathscr{C}(8,0)$ can be chosen as follows:

$$r=\tfrac{1}{16}(1-\Gamma_{1234})(1-\Gamma_{1256})(1-\Gamma_{1357})(1-\Gamma_{3478}). \quad (70)$$

The kernel of the mapping [Eq. (66)] is an $\mathrm{so}(7)$ subalgebra of $\mathrm{so}(8)$ [in fact, for any element $\eta\in S_+$ its stabilizer subalgebra of $\mathrm{so}(8)$ is isomorphic to $\mathrm{so}(7)$; this is a manifestation of the fact that this algebra does not depend on the particular choice of r, but rather on its rank].

A convenient (and orthonormal) basis in S_+ can be constructed as follows. Let us take an arbitrary vector $w_0\in W$ (eight-dimensional Euclidean space); such that $w_0^2=1$. Let $\{w_i\}_1^7$ be an orthonormal basis in the orthogonal complement $w_0^\perp$ of w_0 in W, and let, finally, $\{\Gamma_a\}_0^7$ be the corresponding set of generators of $\mathscr{C}(8,0)$. Then the elements

$$M_i:=\Gamma_0\Gamma_i r,\quad i=1,\ldots,7, \quad (71)$$

form an orthonormal basis in $\varphi(\mathscr{C}^2(8,0))$ and together with r a basis in S_+.

The corresponding (also orthonormal) basis in S_- looks as follows:

$$u_a:=\Gamma_a r,\quad a=0,\ldots,7. \quad (72)$$

Notice that the image of W fills all of S_-.

This construction allows for a very easy identification of the mapping Λ (which is not a representation of any Lie algebra):

$$\begin{aligned}\Lambda(M_i)u_0&=\Gamma_0\Gamma_0\Gamma_i r=\Gamma_i r=u_i,\\ \Lambda(M_i)u_j&=\Gamma_j\Gamma_0\Gamma_i r=\epsilon_{ijk}u_k.\end{aligned} \quad (73)$$

It can be shown[11] that the above formulas define the octonionic multiplication in S_- (and in S_+ and W as well), i.e., ϵ_{ijk} are the octonionic algebra structure constants. This in turn implies that Λ is just the right multiplcation by imaginary octonions. A more detailed study of this approach can be found in Ref. 11. The "current" superalgebra, for which the above is the underlying structure, is identical to one of those described in Ref. 7 (viz., the one in Sec. V [(38)]). It should be pointed out that this algebra stands out as the simplest of the "S^7 constructions," in that there is no violation of (G,G,G) Jacobi identity. In our approach, this is essential for the Clifford algebra construction.

(ii) $N=7$: $\mathrm{Spin}(7)/G_2$. Following exactly the same way one can construct the superalgebra corresponding to $N=7$. This time the elements of W-space (dimension-$\frac{3}{2}$ generators) form a seven-dimensional "adjoint representation" for octonionic imaginary units, and there is one dimension-$\frac{1}{2}$ element, corresponding to the octonionic unity.

The above means that both cases correspond to the same (set theoretical) coset space but with different geometric structures living on it. In the case of $N=8$, the dimension-$\frac{3}{2}$ generators form some kind of spin space over S^7, whereas for $N=7$, they are just tangent vectors. Surprisingly the quotient spaces corresponding to $N=5,6$ are also (set theoretically) seven-spheres with again two different geometric structures put on them.

Work in this direction could provide us a reasonable generalization of the supersymmetry algebra. Further development will be presented elsewhere.

IX. CONCLUSIONS

We have shown that the presence of a Clifford algebra representation in CSCLS provides one with a powerful tool for their analysis. It leads to a complete classification of this type of algebra.

Moreover, it appears that their presence is more universal than the class of algebras analyzed in detail in the first part of this paper. It suggest also a possible direction in which to look for higher N, possible nonassociative algebras, similar to the $N=8$ case. Another possibility would be to include composite operators in the treatment. We leave this for further investigations.

ACKNOWLEDGMENTS

We thank A. Van Proeyen for discussions, and F. Defever for a critical reading of the manuscript.

One of us (Z.H.) also wishes to thank the Institute for Theoretical Physics at Leuven for its hospitality.

APPENDIX A: CONSTRUCTION OF REPRESENTATIONS

In this appendix we remind the reader of some facts about Clifford algebras, and then apply them to deduce some properties of the metric θ, including the solution to the invariance equations of Sec. IV.

Starting from a fixed vector r of a representation space S of the Clifford algebra $\mathscr{C}$, the vector space $\mathscr{C}r$ will be a subspace of S that also carries a (generally reducible) representation of $\mathscr{C}$, with as maximal dimension the dimension of $\mathscr{C}$. A convenient way[8] to construct these representations is to choose, as the representation space, the Clifford algebra itself. This is done by taking an arbitrary element $r\in\mathscr{C}$, and considering the left ideal $\mathscr{C}r$. Depending on r, this representation may be reducible or irreducible. We now explain how irreducible representations arise.

To this end, recall that it is possible to find a maximum of k commuting elements $E_A\in\mathscr{C}$, $E_A^2=1$, that generate a group of order 2^k, where k depends on the dimension and the signature of $\mathscr{C}$. Then $\frac{1}{2}(1+E_A)$ are commuting projection operators, and an irreducible representation of $\mathscr{C}$ is obtained by the left ideal $\mathscr{C}r$ with

$$r=r(\{\epsilon_A\})=\prod_{A=1}^{k}\frac{1}{2}(1+\epsilon_A E_A)$$

and $\epsilon_A\in\{-1,1\}$. A general representation is obtained by taking an arbitrary linear combination r of such irreducible $r(\{\epsilon_A\})$. Some of these combinations may be equivalent.

The metric θ introduced in the main text can be rewritten in this representation as

$$\theta(Ar,Br)=\mathrm{tr}(M\beta_-(Ar)Br)=\mathrm{tr}(rM\beta_-(r)\beta_-(A)B),$$

where $M\in\mathscr{C}$ is an arbitrary β_--invariant element, $\beta_-(M)=M$, and $\mathrm{tr}(F)$ means the component of $F\in\mathscr{C}$ proportional to unity. In addition, by making linear combinations of the irreducible representations that are present, one can arrange that the metric does not mix irreducible components. This implies that there exists a choice of elements E_A such that M is a function into $\mathscr{C}$ of this set of elements, and we can write that

$$\Theta=rM\beta_-(r)=\sum_D a_D\prod_{A\in D}E_A, \qquad (\mathrm{A}1)$$

where the sum is over subsets D of the index set of E_A. The following requirements follow from the properties of θ.

(i) $\theta(r,r)=-1$, and therefore the coefficient in Eq. (A1) corresponding to the term with $D=\varnothing$ has the value

$$a_\varnothing=-1.$$

(ii) Orthogonality of the even and odd subspaces $S_+=\mathscr{L}+\mathbb{R}$ and $S_+=W+V$ implies that Eq. (A1) contains no terms that are odd elements of $\mathscr{C}$.

(iii) $\beta_-(\Theta)=\Theta$ implies that Eq. (A1) contains only terms that are products of a number of Γ matrices that is equal to zero modulo 4. So, for $N<4$, $\Theta=-1$.

(iv) If Eq. (A1) contains a term with the product of four Γ matrices, the product of their squares must necessarily be $+1$, since all E_A commute, and $E_A^2=1$.

The last property proves that, if $\theta(r,\Gamma_i\Gamma_j\Gamma_k\Gamma_l r)\neq 0$, then $(\Gamma_i\Gamma_j\Gamma_k\Gamma_l)^2=+1$ for any signature of the Clifford algebra $\mathscr{C}$.

Now we can prove Eq. (51) for all signatures. If we put $(\Gamma_i\Gamma_j\Gamma_k\Gamma_l)^2=\epsilon=\pm 1$, the first term in Eq. (50) does not vanish iff $(\Gamma_a\Gamma_k\Gamma_i\Gamma_j)^2=(\Gamma_a\Gamma_l)^2\epsilon=1$, which implies that we can replace the B_{ll} coefficient by ϵ. This also holds for the other terms, so that we again obtain Eq. (51).

For constructing the representations of the basic algebras, only the projections formed with four Γ matrices (or more) are useful.

(i) A projector $P=\frac{1}{2}(1+\Gamma_1)$, $\Gamma_1^2=1$, interferes with the splitting in even/odd subspaces $S_\pm$, since $\Gamma_1 P=1\cdot P$.

(ii) A projector $P=\frac{1}{2}(1+\Gamma_1\Gamma_2)$, $\Gamma_1^2=1=-\Gamma_2^2$, cannot be used for the basic algebra, since it would make w_1 and w_2 proportional.

(iii) A projection $P=\frac{1}{2}(1+\Gamma_1\Gamma_2\Gamma_3)$ would also interfere with the splitting $S_\pm$. This argument holds for the enlarged algebras also. Thus, for $N<4$, $r=1$.

APPENDIX B: EXPLICIT FORMULATION OF SOME ENLARGED SCA's NONCOMPACT CASE

We list only the nonvanishing (anti)commutators that are not related to the Virasoro algebra.

1. $N=2$

We have

$$\begin{aligned}
&\{G_m^+,G_n^-\}\\
&\quad=2L_{m+n}+(m-n)T_{m+n}+(m^2-\tfrac{1}{4})\delta_{m+n}c,\\
&[T_m,G_n^\pm]=\pm G_{m+n}^\pm,\\
&[T_m^{a\pm},G_n^\pm]=(m/\sqrt{2})Q_{m+n}^{a\pm},\\
&\{Q_m^{a\mp},G_n^\pm\}=(1/\sqrt{2})T_{m+n}^{a\mp},\\
&[T_m^{a\pm},Q_n^{b\pm}]=f_c^{ab\pm}Q_{m+n}^{c\pm},\\
&[T_m^{a\pm},T_n^{b\pm}]=f_c^{ab\pm}T_{m+n}^{c\pm},\\
&[T_m^{a+},T_n^{b-}]=-mK^{ab}\delta_{m+n}c,\\
&[T_m,T_n]=m\delta_{m+n}c,\\
&\{Q_m^{a+},Q_n^{b-}\}=-K^{ab}\delta_{m+n}c.
\end{aligned}$$

We remind the reader of the fact (Sec. VI) that the metric K^{ab} vanishes on the derived algebra, i.e., if T^a can be written as a commutator, then $K^{ab}=0$, for all b.

2. $N=4$

The nonvanishing commutators for the $N=2+2$ case are

$$
\begin{aligned}
&\{G^{i+}_m, G^{i-}_n\} \\
&\quad = 2L_{m+n} + (m-n)T_{m+n} + (m^2 - \tfrac{1}{4})\delta_{m+n}c \\
&\quad\ (i = 1,2), \\
&[T_m, G^{i\pm}_n] = \pm G^{i\pm}_{m+n}, \\
&[T_m, T_n] = m\delta_{m+n}c, \\
&[T^+_m, G^{1-}_n] = G^{2+}_{m+n}, \\
&[T^+_m, G^{2-}_n] = -G^{1+}_{m+n}, \\
&[T^-_m, G^{1+}_n] = G^{2-}_{m+n}, \\
&[T^-_m, G^{2+}_n] = -G^{1-}_{m+n}, \\
&\{G^{1+}_m, G^{2+}_n\} = -2(m-n)T^+_{m+n}, \\
&\{G^{1-}_m, G^{2-}_n\} = -2(m-n)T^-_{m+n}, \\
&[T^+_m, T^-_n] = -T_{m+n} - (m/2)\delta_{m+n}c, \\
&[T_m, Q^{a\pm}_n] = \pm Q^{a\pm}_{m+n}, \\
&[T^+_m, Q^{a-}_n] = Q^{a+}_{m+n}, \\
&[T^-_m, Q^{a+}_n] = -Q^{a-}_{m+n}, \\
&\{G^{1+}_m, Q^{a-}_n\} = U^{a1}_{m+n}, \\
&\{G^{2-}_m, Q^{a+}_n\} = U^{a1}_{m+n}, \\
&\{G^{2+}_m, Q^{a-}_n\} = U^{a2}_{m+n}, \\
&\{G^{1-}_m, Q^{a+}_n\} = -U^{a2}_{m+n}, \\
&[G^{1+}_m, U^{a2}_n] = 2nQ^{a+}_{m+n}, \\
&[G^{2+}_m, U^{a1}_n] = -2nQ^{a+}_{m+n}, \\
&[G^{1-}_m, U^{a1}_n] = -2nQ^{a-}_{m+n}, \\
&[G^{2-}_m, U^{a2}_n] = -2nQ^{a-}_{m+n}, \\
&[U^{a1}_m, U^{b2}_n] = -mK^{ab}\delta_{m+n}c, \\
&\{Q^{a+}_m, Q^{b-}_n\} = b^{ab}\delta_{m+n}c,
\end{aligned}
$$

with the condition on the metric

$$
K^{ab} = -K^{ba} = 2b^{ba}. \tag{B1}
$$

[1] P. Ramond and J. H. Schwarz, Phys. Lett. B **64**, 75 (1976).
[2] V. G. Kac and I. T. Todorov, Commun. Math. Phys. **102**, 337 (1985).
[3] M. Ademollo, L. Brink, A. D'Adda, R. D'Auria, E. Napolitano, S. Sciuto, E. Del Giudice, P. di Vecchia, S. Ferrara, F. Gliossi, R. Musto, and R. Peltorino, Phys. Lett. B **62**, 105 (1976); Nucl. Phys. B **114**, 297 (1976); M. Ademollo, L. Brink, A. D'Adda, R. D'Auria, E. Napolitano, S. Sciuto, E. Del Giudice, P. di Vecchai, S. Ferrara, F. Gliossi, R. Musto, R. Peltorino, and J. Schwarz, Nucl. Phys. B **111**, 77 (1976).
[4] A. Sevrin, W. Troost, and A. Van Proeyen, Phys. Lett. B **208**, 447 (1988); K. Schoutens, Nucl. Phys. B **295** [FS21], 634 (1988); E. A. Ivanov, S. O. Krivonos, and V. M. Leviant, *ibid.* **304**, 601 (1988).
[5] V. G. Kac and J. W. van de Leur, in *Strings 88*, edited by S. J. Gates, Jr., C. R. Preitschopf and W. Siegel (World Scientific, Singapore, 1989), p. 77.
[6] W. Nahm, Nucl. Phys. B **135**, 149 (1978).
[7] F. Englert, A. Sevrin, W. Troost, A. Van Proeyen, and Ph. Spindel, J. Math. Phys. **29**, 281 (1988).
[8] P. Lounesto, Found. Phys. **11**, 721 (1981).
[9] M. Chaichian and J. Lukierski, Helsinki University preprint HU-TFT-89-16, 1989.
[10] M. Chaichian and J. Lukierski, Phys. Lett. B **212**, 461 (1988).
[11] Z. Hasiewicz and A. K. Kwasniewski, J. Math. Phys. **26**, 6 (1985).

Scattering and complete integrability in conformally invariant nonlinear theories

John C. Baez
Department of Mathematics and Computer Science, University of California, Riverside, California 92521

(Received 28 March 1989; accepted for publication 4 October 1989)

Conformally invariant nonlinear wave equations in four dimensions, corresponding to multicomponent massless scalar fields with a quartic interaction, are studied. It is proved that the scattering operator S on the space $\mathbf{H}$ of finite-Einstein-energy Cauchy data has infinitely many fixed points, as well as periodic points of all orders. There are also $\Phi \in \mathbf{H}$ such that $S^n\Phi$ is almost periodic, but not periodic and $\Phi \in \mathbf{H}$ such that $S^n\Phi$ is not almost periodic. It is also proved that $\mathbf{H}$ admits no conformally invariant Kähler metrics, but infinitely many distinct Kähler metrics invariant under the Poincaré group and scale transformations. Moreover, it is proved that time evolution for these nonlinear wave equations is completely integrable on the space $\mathbf{H}$.

I. INTRODUCTION

For most relativistic nonlinear wave equations the construction of a scattering operator defined on a physically natural space of solutions presently requires quite careful arguments. For conformally invariant nonlinear wave equations, however, the scattering operator arises rather simply in terms of the conformal embedding of Minkowski space in the Einstein universe. This facilitates the study of certain questions that are still intractable for nonlinear wave equations that are Poincaré invariant, but not conformally invariant. In this paper we consider massless scalar fields with a quartic interaction, i.e., equations of the form

$$\Box f + q'(f) = 0, \tag{1}$$

where q is a non-negative homogeneous polynomial of degree four in the multicomponent scalar field f. We work with the space $\mathbf{H}$ of "finite-Einstein-energy solutions." A solution f has finite Einstein energy if the sum of its energy and that of its transform under conformal inversion is finite, or equivalently, if f extends to a finite-energy solution of the corresponding equation on the Einstein universe, as described below. More concretely, a solution f has finite Einstein energy if

$$\int_{x_0=0} \left\{ \left(1+\frac{r^2}{4}\right)((\nabla f)^2+\dot{f}^2) - \frac{1}{2} f^2 \right\} d^3x$$

is finite. The advantage of the space $\mathbf{H}$ is that it is preserved by the action of the conformal group, which is not the case for the traditional space of finite-energy solutions.

Some time ago Segal pointed out the significance of the physical vacuum in classical field theories as an elliptic fixed point for Minkowski time evolution and hence for the scattering operator.[1] For $\Box f + q'(f) = 0$ with $q \neq 0$ we show that there are infinitely many other fixed points for the scattering operator, as well as periodic points of all orders. The question of the stability of these fixed and periodic points is closely related to Floquet theory; using classical results on intervals of instability the presence of "pseudohyperbolic" fixed points can be established, which in turn implies the existence of solutions Φ such that $S^n\Phi$ fails to be almost periodic. This contrasts interestingly with investigations of conformally invariant quantum field theories,[2] which suggest (in part on a nonrigorous basis) that the generator of Einstein time evolution should have a pure point spectrum, implying almost periodicity of $S^n\Phi$ for all states Φ of the quantized system. The situation is reminiscent of the prevalence of bounded, but not almost periodic trajectories in finite-dimensional classical systems whose corresponding quantized Hamiltonians have a pure point spectrum (the "absence of quantum chaos").

A related question concerns the existence of invariant Kähler metrics on the solution manifold of a classical field theory. The solution manifold typically has a natural symplectic structure; a choice of "Kählerization" of the solution manifold is closely related to a choice of vacuum state for the quantized theory.[3] The solution manifold of the linear equation $\Box f = 0$, for example, has a flat Kähler metric that is invariant under the conformal group and states of the corresponding quantized theory may be represented as antiholomorphic functions on the solution manifold via the "complex wave" or "Fock–Bargmann–Segal" representation.[4] We show that for $q \neq 0$ the symplectic structure of the solution manifold $\mathbf{H}$ of $\Box f + q'(f) = 0$ has no conformally invariant Kählerizations, but has infinitely many that are Poincaré invariant and flat. The proof does not rule out the existence of a conformally invariant Kähler metric on a dense open subset of $\mathbf{H}$, nor does it rule out the existence of conformally invariant polarizations which are not Kähler. Thus a manifestly conformally invariant "geometric quantization" may still be possible.[5]

We also show that Minkowski time evolution for $\Box f + q'(f) = 0$ is completely integrable on the space $\mathbf{H}$. This contrasts with the results of Nikolaevskii and Shchur[6] (see, also, Savvidy[7] and Steeb *et al.*[8]), who embed the manifold of space-independent solutions of $\Box f + q'(f) = 0$, where $f = (f_1, f_2)$ and $q(f) = f_1^2 f_2^2$, in the space of smooth solutions of the Yang–Mills equations. By proving that time evolution on this manifold is nonintegrable, Nikolaevskii and Shchur conclude that the Yang–Mills equations are not integrable. However, these space-independent solutions are not of finite Einstein energy, nor of finite energy in the usual sense, and their relevance to the physics of localized systems

 0022-2488/90/030757-06$03.00

is presumably only indirect. There is thus no contradiction: While the complete integrability of some well-known equations in two-dimensional space-time is rather robust with respect to the space of solutions considered,[9] there is no *a priori* reason to expect this in general. Indeed, it is possible that the Yang–Mills equations are completely integrable on a suitable space of solutions modulo gauge transformations. In a separate paper[10] we show by using scattering theory that there are infinitely many gauge-invariant conserved quantities for sufficiently regular solutions of these equations.

II. REVIEW

We begin with a review of the Goursat problem approach to scattering theory for conformally invariant nonlinear wave equations.[11–15] Let $\mathbf{M}_0$ denote Minkowski space and let $\widetilde{\mathbf{M}}$ denote the universal cover of the conformal compactification of $\mathbf{M}_0$, which we identify with the "Einstein universe" $\mathbf{R}\times S^3$ and to which we give the coordinates (τ,u). Let V denote a finite-dimensional real Hilbert space and let q be a homogeneous polynomial of degree four on V, bounded from below. Let $\mathbf{H} = L^{2,1}(S^3,V)\oplus L^2(S^3,V)$ as a real Hilbert space with the inner product given as in Ref. 15, where $L^{2,1}(S^3,V)$ consists of functions which, together with their first derivatives, lie in $L^2(S^3,V)$. Elements of $\mathbf{H}$, the space of *finite-Einstein-energy Cauchy data*, will be denoted by capital Greek letters, e.g., $\Phi = (\Phi_1,\Phi_2)$. Given $\Phi\in\mathbf{H}$ there is a function $\varphi:\widetilde{\mathbf{M}}\to V$, unique modulo functions vanishing a.e., such that

$$(\Box + 1)\varphi + q'(\varphi) = 0 \tag{2}$$

and $(\varphi,\dot\varphi)|_{\tau=0} = \Phi$, where these equations are taken in the sense of distributions. We call such φ a *finite-Einstein-energy solution* of (2) and denote the space of such by $\mathbf{E}_q$, which may be given a Hilbert manifold structure such that the map $\Phi\mapsto\varphi$ is a diffeomorphism.

The group $\widetilde{\mathbf{G}} = \widetilde{\mathrm{SO}}(2,4)$ acts as conformal diffeomorphisms of $\widetilde{\mathbf{M}}$ and the subgroup preserving $\mathbf{M}_0$ is the universal cover of $\mathbf{P}$, the extension of the Poincaré group $\mathbf{P}_0$ by scale transformations. The group $\widetilde{\mathbf{G}}$ acts on $\mathbf{E}_q$ via

$$(g\varphi)(x) = \|dg_x^{-1}\|\varphi(g^{-1}(x)),\quad g\in\widetilde{\mathbf{G}},\quad x\in\widetilde{\mathbf{M}}; \tag{3}$$

this action gives rise to an action U_q of $\widetilde{\mathbf{G}}$ as diffeomorphisms of $\mathbf{H}$.

Let ι: $\mathbf{M}_0\to\widetilde{\mathbf{M}}$ denote the conformal embedding. Given a finite-Einstein-energy solution φ of (2) and defining $f:\mathbf{M}_0\to V$ by $f(x) = p(x)\varphi(\iota(x))$, where p is the conformal factor associated with ι, then (1) holds in the distributional sense and f is said to be a finite-Einstein-energy solution of (1). This correspondence allows one to study scattering of finite-Einstein-energy solutions of (1) in terms of Eq. (2). The *Minkowski time* evolution subgroup of $\widetilde{\mathbf{G}}$, with the generator $\mathbf{T}_0$, acts on $\widetilde{\mathbf{M}}$ in such a way that

$$\exp(t\mathbf{T}_0)\iota(x_0,\vec{x}) = \iota(x_0+t,\vec{x})$$

for all $(x_0,\vec{x})\in\mathbf{M}_0$. Given $\Phi\in\mathbf{H}$ there exist "in and out fields" $\Phi_\pm\in\mathbf{H}$ such that

$$\lim_{t\to\pm\infty}\|U_q(\exp(-t\mathbf{T}_0))\Phi - U(\exp(-t\mathbf{T}_0))\Phi_\pm\| = 0,$$

where $U = U_0$. [Note here that $U_q(\exp(-t\mathbf{T}_0))$ evolves a Cauchy datum *forward* by time t due to the presence of g^{-1} in Eq. (3). Also, while the convergence above was shown in Ref. 14 only for a weaker topology, a similar argument proves convergence in the norm topology of $\mathbf{H}$.] The maps $\Phi\mapsto\Phi_\pm$ are diffeomorphisms. Thus for any given q there is a unique diffeomorphism $S:\mathbf{H}\to\mathbf{H}$, the *scattering operator*, such that $S\Phi_- = \Phi_+$.

The boundary of $\mathbf{M}_0$ as embedded in $\widetilde{\mathbf{M}}$ is the union of the "cones at future and past infinity":

$$C_\pm = \{(\tau,u)\in\mathbf{R}\times S^3 : \rho = \pm(\pi-\tau)\},$$

where ρ is the angle from u to the north pole of S^3. We identify C_- with S^3 by means of the map $(\tau,u)\mapsto u$, but identify C_+ with S^3 by $(\tau,u)\mapsto -u$. Let $\mathbf{H}(C)$, the space of *finite-Einstein-energy Goursat data*, denote $L^{2,1}(S^3,V)$ as a real Hilbert space with the inner product as given in Ref. 15. There are diffeomorphisms $W_{q,\pm}:\mathbf{H}\to\mathbf{H}(C)$, the *wave transforms*, determined by the property that

$$W_{\pm,q}\Phi = \varphi\,|C_\pm$$

for all $\Phi\in\mathbf{H}$ such that the corresponding $\varphi\in\mathbf{E}_q$ is continuous. For all $\Phi\in\mathbf{H}$,

$$\Phi_\pm = W_\pm^{-1}W_{\pm,q}\Phi, \tag{4}$$

where $W_\pm = W_{\pm,0}$.

The center of $\widetilde{\mathbf{G}}$ is isomorphic to $\mathbf{Z}\times\mathbf{Z}_2$ and the generator ζ of the $\mathbf{Z}$ factor acts on $\widetilde{\mathbf{M}}$ by

$$\zeta(\tau,u) = (\tau+\pi,-u).$$

In particular, ζ maps C_- onto C_+ and

$$U_q(\zeta) = W_{+,q}^{-1}W_{-,q}. \tag{5}$$

This gives rise to a close relationship between ζ and scattering. Equation (4) implies that $S = W_+^{-1}W_{+,q}W_{-,q}^{-1}W_-$. Using (5) and the fact that $U(\zeta) = -I$, it follows that

$$S = (W_-^{-1}W_{-,q})(-U_q(\zeta^{-1}))(W_-^{-1}W_{-,q})^{-1}. \tag{6}$$

Thus S is conjugate by a diffeomorphism to $-U_q(\zeta^{-1})$.

For $g\in\widetilde{\mathbf{P}}$,

$$(U_q(g)\Phi)\pm = U(g)\Phi_\pm. \tag{7}$$

It follows that $U(g)S = SU(g)$ for all $g\in\widetilde{\mathbf{P}}$, which expresses the $\widetilde{\mathbf{P}}$ invariance of the scattering operator.

III. STATEMENT OF RESULTS

Recall that given a Banach space $\mathbf{X}$, a sequence $\{x_n\}_{n\in\mathbf{Z}}$ in $\mathbf{X}$ is said to be *almost periodic* if the closure in $L^\infty(\mathbf{Z},\mathbf{X})$ of the set of translates of the sequence $\{x_n\}$ is compact. (For equivalent definitions, see Ref. 16.) Given a homeomorphism $F:\mathbf{X}\to\mathbf{X}$, the point $x\in\mathbf{X}$ is said to be *almost periodic* for F if the sequence $\{F^nx\}$ is almost periodic and *weakly almost periodic* for F if $\{\lambda(F^nx)\}$ is almost periodic for every $\lambda\in\mathbf{X}^*$. A point $x\in\mathbf{X}$ is said to be *periodic of order n* for F if n is the least positive integer such that $F^n(x) = x$. Also, we say that two elements of $\mathbf{H}$ are *equivalent* if for some $g\in\widetilde{\mathbf{P}}$, $U(g)$ maps one into the other. We shall prove the following theorem.

Theorem 1: Assume $q\neq 0$. Then there exists a family $\Psi_a\in\mathbf{H}$, $a\geqslant 0$, such that the following holds. (i) If $a\neq b$, then Ψ_a and Ψ_b are inequivalent. (ii) For each $n\geqslant 1$ there are infinitely many values of a such that Ψ_a is a periodic point of order n for S. (iii) For uncountably many values of a, Ψ_a is almost periodic, but not periodic for S. (iv) For infinitely many values of a, every neighborhood of Ψ_a in $\mathbf{H}$ contains a point that is not weakly almost periodic for S.

Next we recall some facts from Ref. 15 about the geometry of $\mathbf{H}$. Given a non-negative integer n, let $\mathbf{H}_n$ denote the space of C^n vectors for the action U of $\widetilde{\mathbf{G}}$ on $\mathbf{H}$. Recall that $\mathbf{H}_n$ has a natural Banach space structure and is densely embedded in $\mathbf{H}$. In fact, the elements of $\mathbf{H}_n$ are precisely the C^n vectors for U_q regardless of the value of q and for any q the action U_q restricts to an action of $\widetilde{\mathbf{G}}$ as diffeomorphisms of $\mathbf{H}_n$. The space $\mathbf{H}_n$ has a symplectic form

$$\omega(\Phi,\Psi) = \int_{S^3} \Phi_1\cdot\Psi_2 - \Phi_2\cdot\Psi_1,$$

where Φ, Ψ are tangent vectors identified with vectors in $\mathbf{H}_n$. [We shall use the definitions of symplectic forms, Riemann metrics, and Kähler metrics that require only weak nondegeneracy, where a bilinear form A is *weakly nondegenerate* if for every $u\neq 0$ there exists v such that $A(u,v)\neq 0$.] The symplectic form ω is invariant under the action U_q of $\widetilde{\mathbf{G}}$ for any q. By the following theorem ω has infinitely many distinct extensions to a $\widetilde{\mathbf{P}}$-invariant Kähler metric, but no extensions to a $\widetilde{\mathbf{G}}$-invariant Kähler metric.

Theorem 2: Assume $q\neq 0$ and let n be a non-negative integer. Let N be any $U_q(\zeta)$- invariant open neighborhood of $0\in\mathbf{H}_n$. Then there does not exist a Riemann metric on N invariant under the action of $U_q(\zeta)$. There exist infinitely many distinct Kähler metrics on $\mathbf{H}_n$ which are flat, have an imaginary part equal to ω, and are invariant under the action U_q of $\widetilde{\mathbf{P}}$.

Last, the Minkowski time evolution for Eq. (1) is completely integrable in the following sense. Recall that if $(\mathbf{X},\omega)$ is a symplectic manifold, the subspace $L\subset T_x\mathbf{X}$ is *isotropic* if the restriction of ω to L vanishes.

Theorem 3: There exists a sequence of real-analytic functions $F_i:\mathbf{H}\to\mathbf{R}$ such that for all i and all t, $U_q(\exp(t\mathbf{T}_0))^*F_i = F_i$. There exist real-analytic vector fields v_i on $\mathbf{H}$ such that for all i,

$$dF_i = \omega(v_i,\cdot).$$

Moreover, (i) for all i and j, $[v_i,v_j] = 0$ and (ii) generically, i.e., except for Φ in a set of first category in $\mathbf{H}$, the subspace $L = \{\Psi\in T_\Phi\mathbf{H}:\forall i\ dF_i(\Psi) = 0\}$ is isotropic in $T_\Phi\mathbf{H}$.

In short, the functions F_i are a complete set of conserved quantities for Eq. (1), with pairwise vanishing Poisson brackets.

IV. SPACE-INDEPENDENT SOLUTIONS ON $\widetilde{\mathbf{M}}$

Topological arguments have been used to prove the existence of periodic solutions of a general class of wave equations of the form $(\Box + 1)\varphi + f'(\varphi) = 0$ on $\widetilde{\mathbf{M}}$; unfortunately, the case of f quartic represents a "critical exponent" to which these methods do not thus far apply.[17] Instead of taking an abstract approach, we will exploit the properties of explicit "space-independent" periodic solutions of $(\Box + 1)\varphi + q'(\varphi) = 0$ on $\widetilde{\mathbf{M}}$ to prove Theorems 1 and 2.

We recall that the one-parameter *Einstein time* evolution subgroup of $\widetilde{\mathbf{G}}$, with the generator $\mathbf{X}_0$, acts on $\widetilde{\mathbf{M}}$ by

$$\exp(t\mathbf{X}_0)(\tau,u) = (\tau + t,u),$$

where $(\tau,u)\in\mathbf{R}\times S^3$. Thus ζ equals $\xi\exp(\pi\mathbf{X}_0)$, where $\xi\in\widetilde{\mathbf{G}}$ acts on $\widetilde{\mathbf{M}}$ by $\xi(\tau,u) = (\tau, -u)$.

Proof of Theorem 1: We first prove Theorem 1 in the special case $V = \mathbf{R}$, $q(\varphi) = \frac{1}{4}\lambda\varphi^4$, where $\lambda > 0$. If the function g_a is defined as the solution of $\ddot{g} + g + \lambda g^3 = 0$ with $g_a(0) = a > 0$ and $\dot{g}_a(0) = 0$, then the function $\varphi_a:\widetilde{\mathbf{M}}\to\mathbf{R}$ given by $\varphi_a(\tau,u) = g_a(\tau)$ satisfies $(\Box + 1)\varphi_a + \lambda\varphi_a^3 = 0$. The function g_a is easily seen to have the period

$$P(a) = 4\int_0^1 \frac{dx}{\sqrt{(1-x^2) + \lambda a^2(1-x^4)/2}}.$$

The period $P(a)$ is continuous and strictly decreasing in a, with $P(0) = 2\pi$ and $P(a)\to 0$ as $a\to +\infty$. Let $\Phi_a = (a,0)\in\mathbf{H}$ and let

$$\Psi_a = (\Phi_a)_-.$$

To prove statement (i) of Theorem 1, note that if $U(g)\Psi_a = \Psi_b$ for $g\in\widetilde{\mathbf{P}}$, then $U_q(g)\Phi_a = \Phi_b$ by Eq. (7). This is impossible if $a\neq b$, as can be seen by inspection of the action of the group $\widetilde{\mathbf{G}}$ on functions of the form φ_a as given in Eq. (3).

From (6) and the fact that $U_q(\xi)\Phi_a = \Phi_a$ it follows that

$$\begin{aligned} S^n\Psi_a &= ((-U_q(\zeta^{-1}))^n\Phi_a)_- \\ &= ((-U_q(\exp(-\pi\mathbf{X}_0)))^n\Phi_a)_-. \end{aligned} \tag{8}$$

Thus $S^n\Psi_a = \Psi_a$ if and only if $(-U_q(\exp(-\pi\mathbf{X}_0)))^n\Phi_a = \Phi_a$ and Ψ_a is almost periodic for S if and only if Φ_a is almost periodic for $-U_q(\exp(-\pi\mathbf{X}_0))$. Since for all $a > 0$ the function $U_q(\exp(-t\mathbf{X}_0))\Phi_a$ is periodic in t, it follows that Φ_a is almost periodic for $-U_q(\exp(-\pi\mathbf{X}_0))$. For n even, $(-U_q(\exp(-\pi\mathbf{X}_0)))^n\Phi_a = \Phi_a$ if and only if $kP(a) = n\pi$ for some integer k. For n odd, $(-U_q(\exp(-\pi\mathbf{X}_0)))^n\Phi_a = \Phi_a$ if and only if $(k+\frac{1}{2})P(a) = n\pi$ for some integer k. Using the fact that $P(a)$ is continuous and strictly decreasing, it follows that for any n there are infinitely many values of a for which Ψ_a is a periodic point of order n for S and that there are uncountably many values of a for which Ψ_a is almost periodic, but not periodic for S. Thus statements (ii) and (iii) of Theorem 1 have been shown.

To prove statement (iv) of Theorem 1, we consider small perturbations of the space-independent solutions.

Lemma 1: Suppose that $a > 0$ and $P(a)/\pi$ is rational. Then $U_q(\exp(-2n\pi\mathbf{X}_0))\Phi_a = \Phi_a$ for some n and the differential of $U_q(\exp(-2n\pi\mathbf{X}_0))$ is a bounded operator on the tangent space of $\mathbf{H}$ at Φ_a. This operator is of the form $I + K$, with K compact, and decomposes as the direct sum of linear operators T_i on two-dimensional subspaces of the tangent space of $\mathbf{H}$ at Φ_a. For infinitely many such a one of the summands T_i has a real eigenvalue of modulus less than 1.

Proof: The differential of the Einstein time evolution operator $U_q(\exp(-t\mathbf{X}_0))$ can be computed by solving the variational differential equation as in Ref. 15. Thus if Ψ is a tangent vector at the point $\Phi_a\in\mathbf{H}$ and $\Psi(t) = dU_q(\exp(-t\mathbf{X}_0))\Psi$, then $\Psi(t)$, under its canonical identification with a vector in $\mathbf{H}$, satisfies the integral equation

$$\Psi(t) = e^{tA}\Psi + \int_0^t e^{(t-s)A}B(s)\Psi(s)ds, \tag{9}$$

where the skew-adjoint operator A on $\mathbf{H}$ is given by $A(f_1,f_2) = (f_2, -(\Delta+1)f_1)$ and the smooth maps $B(t):\mathbf{H}\to\mathbf{H}$ are given by $B(t)(f_1,f_2) = (0, -3\lambda g_a(t)^2 f_1)$.

One can solve (9) using a time-ordered exponential. Letting $\Psi_{\rm int}(t) = e^{-tA}\Psi(t)$, it follows that

$$\Psi_{\rm int}(t) = \Big(I + \sum_{n\geqslant 1}\int_{0<t_1<\cdots<t_n<t} C(t_n)\cdots C(t_1)dt_1\cdots dt_n\Big)\Psi,$$

where $C(t) = e^{-tA}B(t)e^{tA}$ and the sum is norm convergent. In fact, since $C(t)$ is compact for all t and norm continuous as a function of t, the operator $\Psi\mapsto\Psi_{\rm int}(t)$ is of the form $I+K$, with K compact. The spectrum of $(\Delta+1)$ on $L_2(S^3)$ consists of the eigenvalues N^2, $N\geqslant 1$, with finite multiplicity. Thus $e^{2\pi nA} = I$, so that $\Psi(2\pi n) = \Psi_{\rm int}(2\pi n)$. It follows that the operator $\Psi\mapsto\Psi(2\pi n)$ is of the form $I+K$, with K compact.

Let f_i be an orthonormal basis of eigenfunctions of $(\Delta+1)$ on $L_2(S^3)$. If $\Psi = (\alpha_1 f_i, \alpha_2 f_i)$ where $\alpha_1, \alpha_2\in\mathbf{R}$ and $(\Delta+1)f_i = N^2 f_i$, then (9) implies that $\Psi(t) = (y(t)f_i, \dot y(t)f_i)$, where $y:\mathbf{R}\to\mathbf{R}$ satisfies

$$\partial_t^2 y(t) + (N^2 + 3\lambda g_a(t)^2)y(t) = 0, \qquad (10)$$

with $y(0) = \alpha_1$ and $\dot y(0) = \alpha_2$. Let $T_i:(\alpha_1,\alpha_2)\mapsto(y(0),\dot y(0))$. Then the operator $\Psi\mapsto\Psi(2\pi n)$ is a direct sum of the 2×2 matrices T_i.

Next we use some results from Floquet theory.[18] First, in order to show that one of the matrices T_i has a real eigenvalue of modulus less than 1, it suffices to show that for some integer $N>1$ Eq. (10) has unbounded solutions, or that N^2 lies in one of the "intervals of instability" $(\lambda_{2j-1},\lambda_{2j})$. We use an explicit formula for the function g_a above to show that these intervals of instability are nonempty. By the definition of the Jacobi elliptic function $\mathrm{cn}(t\,|m)$ as the solution of the differential equation

$$\partial_t^2\,\mathrm{cn}(t\,|m) + (1-2m)\mathrm{cn}(t\,|m) + 2m\,\mathrm{cn}^3(t\,|m) = 0,$$

with the initial conditions $\mathrm{cn}(t\,|m) = 1$ and $\partial_t\,\mathrm{cn}(t\,|m) = 0$, it follows that $g_a(t) = a\,\mathrm{cn}(bt\,|m)$, where

$$b = \sqrt{c+1}, \quad m = c/2(c+1), \quad c = \lambda a^2.$$

By the identity $\mathrm{sn}^2(t\,|m) + \mathrm{cn}^2(t\,|m) = 1$, Eq. (10) can be seen to be equivalent to the Lamé equation

$$\partial_s^2 y(s) + (k^2 - p(p+1)m^2\,\mathrm{sn}^2(s|m))y(s) = 0,$$

where $k^2 = (N^2+3c)/(c+1)$, $p(p+1) = 12(1+c^{-1})$, and $s = bt$. The Lamé equation is of intrinsic interest in Floquet theory and appears in the inverse spectral approach to space-periodic solutions of the Kortweg–deVries equation.[9] It is known that none of the intervals of instability of the Lamé equation is empty unless p is an integer. By the definition of p above, this does not occur if $c\geqslant 1$.

We will treat the endpoints of the intervals of instability for (10) as functions of a. The above remarks imply that the endpoints of the $2j$th interval of instability of (10), $(\lambda_{2j-1}(a),\lambda_{2j}(a))$, satisfy $\lambda_{2j-1}(a) < \lambda_{2j}(a)$ if $c\geqslant 1$. It follows from Floquet theory that the $\lambda_j(a)$ are continuous functions of a. Moreover, the following bounds hold:

$$\Big|\sqrt{\lambda_{2j-1}(a)} - \frac{4\pi j}{P(a)}\Big|, \Big|\sqrt{\lambda_{2j}(a)} - \frac{4\pi j}{P(a)}\Big| \leqslant \epsilon_j(a),$$

where $\epsilon_j(a)\to 0$ as $j\to\infty$, uniformly in a for $a>0$ in any compact interval.

For any sufficiently small $P>0$, there exist a_1, a_2 with $2P(a_1) = P(a_2) = P$; $a_2 > a_1$; and λa_1^2, $\lambda a_2^2 \geqslant 1$. Choose j large enough such that $\epsilon_j(a)\leqslant 1$ for all $a\in[a_1,a_2]$. Thus we have

$$\sqrt{\lambda_{2j}(a_1)} \leqslant \frac{2\pi j}{P} + 1, \quad \sqrt{\lambda_{2j-1}(a_2)} \geqslant \frac{4\pi j}{P} - 1.$$

By the intermediate value theorem this implies that for P sufficiently small, some integer N lies in $(\sqrt{\lambda_{2j-1}(a)}, \sqrt{\lambda_{2j}(a)})$ for all a in some nonempty open subinterval of $[a_1,a_2]$. Thus infinitely many $a\in[a_1,a_2]$ have $P(a)/\pi$ rational and $N^2\in(\lambda_{2j-1}(a),\lambda_{2j}(a))$. □

Lemma 2: For infinitely many values of a the following holds. $S^{2n}\Psi_a = \Psi_a$ for some n; the differential of S^{2n} at Ψ_a is of the form $I+K$, with K compact, and has a real eigenvalue of modulus less than 1.

Proof: If $P(a)/\pi$ is rational, then $S^{2n}\Psi_a = \Psi_a$ for some n. By (8) the differential of S^{2n} at Ψ_a is conjugate to the differential of $U_q(\exp(-2n\pi X_0))$ at Φ_a. By Lemma 1 this implies that the differential of S^{2n} at Ψ_a is of the form $I+K$, with K compact, and is conjugate to the direct sum of the matrices T_i. Choosing a as in Lemma 1, it follows that the differential of S^{2n} at Φ_a has a real eigenvalue of modulus less than 1. □

To prove statement (iv) of Theorem 1 in the special case we are considering, choose a as in Lemma 2. By Lemma 2, for some constant $0<r<1$ the spectrum of $(dS^{2n})_{\Psi_a}$ lies off the circle $|z| = r$ and has nonempty intersection with $\{|z|<r\}$, i.e., Ψ_a is an r-pseudohyperbolic fixed point for S^{2n}. By the invariant manifold theorem for pseudohyperbolic fixed points,[19] in any neighborhood of Ψ_a there exists $\Psi'\neq\Psi_a$ such that

$$\lim_{n\to+\infty}\|S^{2n}\Psi' - \Psi_a\| = 0.$$

This implies that $\{S^n\Psi'\}$ cannot be weakly almost periodic since the only convergent weakly almost periodic sequences are constant.[16]

Last, we make the following observation. Given q on V, let V' be a one-dimensional subspace of V such that the maximum of q on the unit ball of V occurs at the intersection of V' with the unit sphere. Let e be a unit vector in V' and let $\lambda = 4q(e)$. If $\varphi:\widetilde{\mathbf{M}}\to\mathbf{R}$ is a finite-Einstein-energy solution of $(\Box+1)\varphi + \lambda\varphi^3 = 0$, then $e\varphi:\widetilde{\mathbf{M}}\to V$ is a finite-Einstein-energy solution of (2). This embedding is easily seen to reduce the proof of Theorem 1 to the special case we have considered. □

Corollary: Assume that $q\neq 0$ and n is a non-negative integer. Then there exists a family $\Phi_a\in\mathbf{H}_n$, $a\geqslant 0$, such that the following holds. (i) For each $m\geqslant 1$ there are infinitely many values of a such that Φ_a is a periodic point of order m for $U_q(\zeta)$. (ii) For uncountably many values of a, Φ_a is almost periodic, but not periodic for $U_q(\zeta)$. (iii) For infinitely many values of a, every neighborhood of Φ_a in $\mathbf{H}_n$ contains a point that is not weakly almost periodic for $U_q(\zeta)$.

Proof: This follows from the proof of Theorem 1. □

Proof of Theorem 2: For the first statement of theorem 2 it suffices to show that there is no Riemann metric g on N invariant under $U_q(\zeta^{2n}) = U_q(\exp(2n\pi \mathbf{X}_0))$. If there were, the metric on N given by

$$d(x,y) = \inf\left\{ \int_0^1 g(f'(t),f'(t))^{1/2}\,dt: \right.$$

$$\left. f:[0,1]\to N \text{ is smooth}, f(0)=x, f(1)=y \right\}$$

would be continuous in the $\mathbf{H}_n$ topology and invariant under $U_q(\exp(2n\pi\mathbf{X}_0))$. However, by the following lemma this would contradict the properties of the family $\{\Phi_a\}$ established in the proof of Theorem 1.

Lemma 3: Let $\mathbf{X}$ be a topological space with a continuous metric d (not necessarily inducing the topology of $\mathbf{X}$) and let U_t be a continuous action of $\mathbf{R}$ on $\mathbf{X}$. Suppose that for integer values of t the map U_t preserves d. Then there cannot exist a family $\Phi_a \in \mathbf{X}$, $a \geqslant 0$ such that (i) Φ_a depends continuously on a, (ii) Φ_a is periodic for U_t with period $P(a)$, and (iii) $P(a)$ is continuous and strictly monotone.

Proof: Choose $a \geqslant 0$ such that $P(a)$ is rational and nonzero. For some n, $U_n(\Phi_a) = \Phi_a$. Let $D = \sup_{t\in\mathbf{R}} d(\Phi_a, U_t\Phi_a)$; clearly, $0 < D < \infty$. Choose b such that $P(b)$ is irrational and close enough to a such that $d(\Phi_a,\Phi_b) \leqslant D/4$ and $d(\Phi_b, U_t\Phi_b) > 3D/4$ for some $t\in[0,P(b)]$. By the Kronecker theorem, for any $\delta > 0$ there exists an integer m such that $|nm \bmod P(b) - t| \leqslant \delta$. Choosing δ sufficiently small, one obtains m such that $d(\Phi_b, U_{nm}\Phi_b) \geqslant 3D/4$. Thus we have

$$\begin{aligned} d(U_{nm}\Phi_a, U_{nm}\Phi_b) &= d(\Phi_a, U_{nm}\Phi_b) \\ &\geqslant d(\Phi_b, U_{nm}\Phi_b) - d(\Phi_a,\Phi_b) \\ &\geqslant D/2, \end{aligned}$$

contradicting the assumption that U_{nm} preserves d, which implies

$$d(U_{nm}\Phi_a, U_{nm}\Phi_b) = d(\Phi_a,\Phi_b) \leqslant D/4. \qquad \square$$

The existence of a flat Kähler structure g^- on $\mathbf{H}_n$ with an imaginary part equal to ω and invariant under the action U_q of $\widetilde{\mathbf{P}}$ is shown in Ref. 15 in the special case of a one-component field φ and the proof generalizes straightforwardly. Since $\zeta\in\widetilde{\mathbf{G}}$ is central it follows that all the flat Kähler structures $U_q(\zeta^k)^* g^-$ are similarly $\widetilde{\mathbf{P}}$ invariant and since $U_q(\zeta)$ is symplectic they all have an imaginary part equal to ω; since $U_q(\zeta^{2n})$ preserves no Riemann metric they are all distinct. □

V. COMPLETE INTEGRABILITY

We begin by introducing a representation of the space $\mathbf{H}$ in terms of Fourier transforms of finite-Einstein-energy solutions of $\Box f = 0$. Let V^c denote the complexification of V and let $\mathscr{S}(\mathbf{R}^3,V)$ denote the Schwartz space of V-valued functions on $\mathbf{R}^3$.

Lemma 4: Given $\Phi\in\mathbf{H}$, let φ be the solution of $(\Box + 1)\varphi = 0$ on $\widetilde{\mathbf{M}}$ with $(\varphi,\dot\varphi)|_{\tau=0} = \Phi$ and let f be the solution of $\Box f = 0$ on $\mathbf{M}_0$ given by $f(x) = p(x)\varphi(\iota(x))$. Given $\vec{k}\in\mathbf{R}^3$ let $k_0 = \|\vec{k}\|$. Then there is a unique function $\widehat{\Phi}\in L^2(\mathbf{R}^3, k_0^{-1}\,d^3k, V^c)$ such that

$$f(x) = (2\pi)^{-3/2}\,\mathrm{Re}\int \widehat{\Phi}(\vec{k})e^{i(k_0x_0 - \vec{k}\cdot\vec{x})}k_0^{-1}\,d^3k, \qquad (11)$$

The map $\Phi\mapsto\widehat{\Phi}$ is one-to-one and continuous from $\mathbf{H}$ to $L^2(\mathbf{R}^3,k_0^{-1}\,d^3k,V^c)$; denote its range as $\widehat{\mathbf{H}}$ and give $\widehat{\mathbf{H}}$ the norm such that $\|\widehat{\Phi}\|_{\widehat{\mathbf{H}}} = \|\Phi\|_{\mathbf{H}}$. There is a continuous embedding of $C_0^\infty(\mathbf{R}^3 - 0, V^c)$ in $\widehat{\mathbf{H}}$.

Proof: Suppose that $\Phi\in\mathbf{H}$, f is defined as above, and $\widehat{\Phi}$ is defined by

$$\widehat{\Phi}(\vec{k}) = \int (k_0 f(0,\vec{x}) - i\dot f(0,\vec{x}))e^{i\vec{k}\cdot\vec{x}}\,d^3x \qquad (12)$$

(in the distributional sense). By results of Ref. 20,

$$\begin{aligned} \int |\widehat{\Phi}|^2 k_0^{-1}\,d^3k &= \int_{x_0=0} ((\Delta^{1/4}f)^2 + (\Delta^{-1/4}\dot f)^2)\,d^3x \\ &\leqslant \int \left\{ \left(1+\frac{r^2}{4}\right)((\nabla f)^2 + \dot f^2) - \frac{1}{2}f^2 \right\} d^3x \\ &= \|\Phi\|^2, \end{aligned}$$

so that $\Phi\mapsto\widehat{\Phi}$ is continuous from $\mathbf{H}$ to $L^2(\mathbf{R}^3,k_0^{-1}\,d^3k,V^c)$. Taking Fourier transforms, (11) follows, which implies that the map $\Phi\mapsto\widehat{\Phi}$ is one-to-one.

If f is a solution of $\Box f = 0$ with $(f,\dot f)|_{x_0=0}\in\mathscr{S}(\mathbf{R}^3,V)\oplus\mathscr{S}(\mathbf{R}^3,V)$, then f is a finite-Einstein-energy solution and the map $\mathscr{S}(\mathbf{R}^3,V)\oplus\mathscr{S}(\mathbf{R}^3,V)\to\mathbf{H}$ is continuous. By (11) this implies that there is a continuous embedding of $C_0^\infty(\mathbf{R}^3 - 0, V^c)$ in $\widehat{\mathbf{H}}$. □

Lemma 5: If Ψ_1, Ψ_2 are tangent vectors at a point of $\mathbf{H}$, identified with vectors in $\mathbf{H}$, we have

$$\omega(\Psi_1,\Psi_2) = \mathrm{Im}\int \widehat{\Psi}_1\cdot\widehat{\Psi}_2 k_0^{-1}\,d^3k.$$

Proof: This is well known; see, for example, Ref. 20. □

Proof of Theorem 3: By Lemma 4, for any $h\in C_0^\infty(\mathbf{R}^3,V^c)$ the function $F_h:\mathbf{H}\to\mathbf{R}$ given by

$$F_h(\Phi) = \int |h\cdot\widehat{\Phi}|^2 k_0^{-1}\,d^3k$$

is real analytic and for all t,

$$U(\exp(t\mathbf{T}_0))^* F_h = F_h,$$

i.e., F_h is a conserved quantity for the free Minkowski time evolution on $\mathbf{H}$. We obtain analogous conserved quantities for the interacting Minkowski time evolution as follows. Define $F_{h,q}:\mathbf{H}\to\mathbf{R}$ by

$$F_{h,q}(\Phi) = F_h(\Phi_-),$$

where $\Phi_- = W_-^{-1}W_{-,q}\Phi$ is the "in field." Since F_h is real analytic and the map $\Phi\mapsto\Phi_-$ is real analytic,[21] it follows that $F_{h,q}$ is real analytic. Recalling that $U(g)\Phi_- = (U_q(g)\Phi)_-$ for all $\Phi\in\mathbf{H}$ and $g\in\widetilde{\mathbf{P}}$, it follows that for all t,

$$\begin{aligned} F_{h,q}(U_q(\exp(t\mathbf{T}_0))\Phi) &= F_h((U_q(\exp(t\mathbf{T}_0))\Phi)_-) \\ &= F_h(U(\exp(t\mathbf{T}_0))\Phi_-) \\ &= F_h(\Phi_-) \\ &= F_{h,q}(\Phi), \end{aligned}$$

so that $U_q(\exp(t\mathbf{T}_0))^* F_{h,q} = F_{h,q}$.

Let $\{e_\lambda\}$ be an orthornormal basis of V and let h_i be a sequence in $C_0^\infty(\mathbf{R}^3-0,V)$ forming an orthonormal basis of $L^2(\mathbf{R}^3,V^c)$ such that each h_i is of the form ke_λ for some λ and some real-valued $k\in C_0^\infty(\mathbf{R}^3-0)$. Defining $F_i=F_{h_i,q}$, it follows from the above remarks that the F_i are real analytic and invariant under $U_q(\exp(t\mathbf{T}_0))$, as desired.

Next we construct the corresponding vector fields v_i. Let $R\Phi=\widehat{\Phi_-}$. Since the map $\Phi\mapsto\Phi_-$ is a real-analytic diffeomorphism, the same is true of $R:\mathbf{H}\to\widehat{\mathbf{H}}$. A computation using the Fourier transform shows that the map $\widehat{\Phi}\mapsto 2i(h_i\cdot\widehat{\Phi})\bar{h}_i$ is a bounded linear transformation of $\widehat{\mathbf{H}}$. Let $v_i\in T_\Phi\mathbf{H}$ be the tangent vector such that $dR(v_i)=2i(h_i\cdot R\Phi)\bar{h}_i$. As a function of Φ, v_i defines a real-analytic vector field on $\mathbf{H}$.

Let $\Psi\in T_\Phi\mathbf{H}$. It follows from Lemma 5 and the fact that $\Phi\mapsto\Phi_-$ is symplectic[15] that

$$\begin{aligned}\omega(v_i,\Psi)&=\mathrm{Im}\int dR(v_i)\cdot\overline{dR(\Psi)}k_0^{-1}\,d^3k\\&=\mathrm{Im}\int 2i(h_i\cdot R\Phi)(\overline{h_i\cdot dR(\Psi)})k_0^{-1}\,d^3k.\end{aligned}$$

By definition,

$$F_i(\Phi)=\int|h_i\cdot R\Phi|^2k_0^{-1}\,d^3k,$$

so that

$$\begin{aligned}dF_i(\Psi)&=2\,\mathrm{Re}\int(h_i\cdot R\Phi)\,\overline{(h_i\cdot dR(\Psi))}k_0^{-1}\,d^3k\\&=\omega(v_i,\Psi),\end{aligned}$$

as claimed.

To prove statement (i) of Theorem 3 it suffices to show that $\omega(v_i,v_j)=0$, as follows:

$$\begin{aligned}\omega(v_i,v_j)&=\mathrm{Im}\int dR(v_i)\cdot\overline{dR(v_j)}k_0^{-1}\,d^3k\\&=4\,\mathrm{Im}\int(h_i\cdot R\Phi)(\bar{h}_i\cdot h_j)(\overline{h_j\cdot R\Phi})k_0^{-1}\,d^3k\\&=0,\end{aligned}$$

using the fact that h_i and h_j are of the form ke_λ with k real valued.

To prove statement (ii) of Theorem 3 we show that the space $L\subset T_\Phi\mathbf{H}$ is isotropic for $\Phi\in D$, where $D\in\mathbf{H}$ is the set of Φ such that $e_\lambda\cdot R\Phi$ is a.e. nonzero for all λ. (Here and below "almost everywhere" and all integrals are relative to the measure $k_0^{-1}\,d^3k$.) Suppose that $\Phi\in D$ and $\Psi\in T_\Phi\mathbf{H}$. Then $\Psi\in L$, that is, $dF_i(\Psi)=0$ for all i if and only if $(e_\lambda\cdot R\Phi)(\overline{e_\lambda\cdot dR(\Psi)})$ is a.e. imaginary for all λ. By the definition of D given Ψ_1, $\Psi_2\in L$, we may conclude that $(e_\lambda\cdot dR(\Psi_1))(\overline{e_\lambda\cdot dR(\Psi_2)})$ is a.e. real; hence

$$\omega(\Psi_1,\Psi_2)=\mathrm{Im}\int dR(\Psi_1)\cdot\overline{dR(\Psi_2)}k_0^{-1}\,d^3k=0.$$

Thus L is isotropic.

It remains to show that the complement of D is of first category. Since $\Phi\mapsto\Phi_-$ is a diffeomorphism, it suffices to show that for each λ, the complement of the set of $\Phi\in\mathbf{H}$ such that $e_\lambda\cdot\widehat{\Phi}$ is a.e. nonzero is of first category. Thus it is enough to show that in the case $V=\mathbf{R}$, the complement of the set D' of $\Phi\in\mathbf{H}$ such that $\widehat{\Phi}$ is a.e. nonzero is of first category. Note that $D'=\bigcap_{m,n}D_{m,n}$, where $D_{n,m}$ is the set of $\Phi\in\mathbf{H}$ such that $\widehat{\Phi}$ vanishes on a set of measure less than $1/n$ in the ball of radius m. Thus it suffices to show that for all n, m the complement $D_{n,m}^c$ is nowhere dense in $\mathbf{H}$.

First we show that the interior of $D_{n,m}^c$ is empty. Given any $\Phi\in\mathbf{H}$, by Lemma 4 there exists $\Psi\in\mathbf{H}$ of arbitrarily small norm such that $\widehat{\Phi}+\widehat{\Psi}$ vanishes on a set of measure less than $1/n$ in the ball of radius m, so that $\Phi+\Psi\in D_{n,m}$.

Next we show that $D_{n,m}^c$ is closed. Suppose that $\Phi_i\in D_{n,m}^c$ and $\Phi_i\to\Phi$, but $\Phi\in D_{n,m}$. Then for each i the set

$$A_i=\{\vec{k}\in\mathbf{R}^3:\|\vec{k}\|\leqslant m,|\widehat{\Phi}_i|=0\}$$

has measure greater than or equal to $1/n$, but for some constant $c>0$ the set

$$A=\{\vec{k}\in\mathbf{R}^3:\|\vec{k}\|\leqslant m,|\widehat{\Phi}|<c\}$$

has measure less than $1/n$ since the ball of radius m has finite measure. We obtain a contradiction as follows. On one hand $\widehat{\Phi}_i\to\widehat{\Phi}$ in L^2 by Lemma 5, so that

$$\int_{A_i}|\widehat{\Phi}|^2=\int_{A_i}|\widehat{\Phi}-\widehat{\Phi}_i|^2\to 0.$$

On the other hand,

$$\int_{A_i}|\widehat{\Phi}|^2\geqslant\int_{A_i-A}|\widehat{\Phi}|^2\geqslant c^2\int_{A_i-A}1$$

and the measure of A_i-A is bounded below by a positive constant. □

ACKNOWLEDGMENTS

The author is indebted to I. Segal and Z. Zhou for many discussions, especially concerning invariant Kähler metrics on solution manifolds. He would also like to thank J. Stafney for help with the category argument.

[1] I. Segal, in *Global Analysis*, Proceedings of the Symposium on Pure Mathematics 16, edited by S. Chern and S. Smale (A.M.S., Providence, RI, 1970), pp. 95–100.
[2] G. Mack, in *Nonperturbative Quantum Field Theory*, edited by G. t'Hooft (Plenum, New York, 1988).
[3] I. Segal, J. Math. Phys. **6**, 468 (1960).
[4] I. Segal, in *Topics in Functional Analysis*, Adv. Math. Suppl. Studies 3, edited by I. Gohberg and M. Kac (Academic, New York, 1978), pp. 321–343.
[5] J. Śniatycki, *Geometric Quantization and Quantum Mechanics*, Appl. Math. Sci. 30 (Springer, New York, 1980).
[6] E. Nikolaevskii and L. Shchur, Sov. Phys. JETP **58**, 1 (1983).
[7] G. K. Savvidy, Nucl. Phys. B **246**, 302 (1984).
[8] W.-H. Steeb, C. Villet, and A. Kunick, J. Phys. A **18**, 3269 (1985).
[9] S. Novikov, Ed., *Integrable Systems*, London Math. Soc. Lecture Note Series 60 (Cambridge U.P., London, 1981).
[10] J. Baez, Adv. Math., to appear.
[11] R. Penrose, Phys. Rev. Lett. **10**, 66 (1963).
[12] I. Segal, Bull. Am. Math. Soc. **77**, 958 (1971).
[13] R. Lüscher, Nucl. Phys. B **140**, 429 (1978).
[14] J. Baez, I. Segal, and Z. Zhou (unpublished).
[15] J. Baez, J. Funct. Anal. **83**, 317 (1989).
[16] S. Zaidman, *Almost-periodic Functions in Abstract Spaces*, Res. Notes Math. 126 (Pitman, Boston, MA, 1985).
[17] Z. Zhou, J. Funct. Anal. **80**, 332 (1988).
[18] W. Magnus and S. Winkler, *Hill's Equation* (Interscience, New York, 1966).
[19] M. Hirsch, C. Pugh, and M. Shub, *Invariant Manifolds*, Lecture Notes Math. 583 (Springer, New York, 1977).
[20] S. Paneitz and I. Segal, J. Funct. Anal. **47**, 78 (1982).
[21] J. Baez and Z. Zhou, Commun. Math. Phys. **124**, 9 (1989).

Localized bound states of fermions interacting via massive vector bosons

D. C. Ionescu, J. Reinhardt, B. Müller, and W. Greiner
Institut für Theoretische Physik, J. W. Goethe-Universität, Robert-Mayer-Strasse 8-10, D-6000 Frankfurt am Main, Federal Republic of Germany

G. Soff
Gesellschaft für Schwerionenforschung (GSI), Planckstrasse 1, Postfach 110 552, D-6100 Darmstadt, Federal Republic of Germany

(Received 5 April 1989; accepted for publication 1 November 1989)

A model for composite systems consisting of fermions with internal degrees of freedom interacting via intermediate vector bosons (IVB) is constructed. Highly localized, low-mass bound states are found in the Hartree–Fock approximation. The dependence of these states as function of the coupling constant and vector boson mass is investigated. In the limit of infinite vector boson mass the interaction is described by Fermi-type contact forces.

I. INTRODUCTION

In this paper, we perform an analysis of the properties of strongly bound states of interacting fermions. The Hartree–Fock approximation is employed. Nonlinear interactions of fermions were extensively studied on the classical level during the last years because the corresponding nonlinear field equations possess solitary wave solutions of finite energy and momentum. This property has been utilized to generate models for particles with internal structure.[1] In a recent paper,[2] we investigated bound states of particle–antiparticle clusters due to Fermi-type contact interactions. The main result found in Ref. 2 was that the values of the coupling constants required to support bound states are quite large.

Let us briefly review the approach followed in Ref. 2, since we shall have occasion to make repeated use of it in the present paper. We investigated the question whether clusters of two fermion–antifermion pairs can be bound by a nonlinear force, for which we considered scalar- and vector-type contact interactions of the form $\lambda_s(\overline{\Psi}\Psi)^n$ and $\lambda_v(\overline{\Psi}\gamma^\mu\Psi)^n$, respectively. Of course, these interactions must be considered as effective interactions since they are not renormalizable. The ground state of the system was assumed to contain two fermions (spin up, spin down) and two antifermions (spin up, spin down) in the same s state. After expanding the fermion field operators in a single-particle basis we derived a nonlinear single-particle equation from a variational principle. In addition, we found that at least for the two fermion-pairs system vector and scalar coupling are equivalent. This surprising result could be explained making use of a Fierz transformation.

In this paper we investigate properties of fermion fields interacting strongly through the exchange of vector mesons. It is well known that in field theories of Yang–Mills type the interaction of fermion fields with vector gauge fields plays a fundamental role. In Ref. 3 the interaction of a fermion field with one vector-meson field was investigated in an attempt to understand some aspects of the problem of quark confinement. Of course, field theories involving coupled fermion and vector-meson fields are interesting in themselves, apart from the points of view presented above.

II. PROBLEM DEFINITION

Our starting point is the following relativistically invariant classical Lagrange density:

$$\begin{aligned}\mathscr{L} &= \mathscr{L}_{\mathrm{F}} + \mathscr{L}_{\mathrm{B}} + \mathscr{L}_{\mathrm{FB}} \\ &= \overline{\Psi}(i\gamma^\mu\partial_\mu - m)\Psi - \tfrac{1}{4}\,\mathbf{G}_{\mu\nu}\cdot\mathbf{G}^{\mu\nu} + \tfrac{1}{2}M_V^2\mathbf{V}_\mu\cdot\mathbf{V}^\mu \\ &\quad - g_v\overline{\Psi}\gamma^\mu(\mathbf{T}/2)\cdot\mathbf{V}_\mu\Psi, \end{aligned} \tag{1}$$

where $\mathscr{L}_{\mathrm{F}}$, $\mathscr{L}_{\mathrm{B}}$, and $\mathscr{L}_{\mathrm{FF}}$ describe the free fermion field Ψ, the free non-Abelian vector-boson field $\mathbf{V}_\mu$ with mass M_V, and the coupling between them, respectively. The vector $\mathbf{T}$ denotes the $N\times N$ generators of SU(N) and the vector arrow on the gauge field $\mathbf{V}_\mu$ indicates that it is in the adjoint representation of SU(N). In addition, the SU(N) gauge invariant field strength tensor $\mathbf{G}_{\mu\nu}$ reads

$$\mathbf{G}_{\mu\nu} = \partial_\mu\mathbf{V}_\nu - \partial_\nu\mathbf{V}_\mu + g_v\mathbf{V}_\mu\times\mathbf{V}_\nu. \tag{2}$$

The fermion field is assumed to be in the fundamental representation of SU(N). The mass of the vector field $\mathbf{V}_\mu$ is to be regarded as having a dynamical origin through the coupling of the massless vector field to a Higgs field $\Phi(x)$. At this point we make the assumption that the influence of the nonlinear gauge field terms can be absorbed into effective values for the (running) gauge coupling constant g_v, the vector boson mass M_V, and the fermion mass m. Since we are mainly interested in the case of large vector boson masses this should be a reasonable approximation due to the short range e^{-Mr}/r behavior of the vector field propagator. As a consequence, for large M_V the boson–boson interactions are of minor importance, in contrast to the case of massless gauge bosons where these terms are presumably responsible for confinement. In this Abelian dominance approximation the boson–boson interactions are absent and we have to deal with (N^2-1) uncoupled Proca fields. Furthermore we assume that the gauge fields are time-independent and we consider only their longitudinal components. We would like to remark that the quadratic terms in the definition (2) of the strength tensor $\mathbf{G}_{\mu\nu}$ vanishes in this case due to the antisymmetry of the SU(N) structure constants. With these assump-

 0022-2488/90/030763-08$03.00

tions one obtains from the Lagrangian (1) the following equations of motion:

$$[\boldsymbol{\alpha}\cdot\mathbf{p}+\beta m+g_v(\mathbf{T}/2)\cdot\mathbf{V}_0(\mathbf{x})]\psi(\mathbf{x})=\omega\psi(\mathbf{x}), \tag{3}$$

$$(-\Delta+M_V^2)\mathbf{V}_0(\mathbf{x})=g_v\psi^+(\mathbf{x})(\mathbf{T}/2)\psi(\mathbf{x}), \tag{4}$$

where we assumed a stationary time dependence of the fermion field:

$$\Psi(t,\mathbf{x})=\psi(\mathbf{x})e^{-i\omega t}. \tag{5}$$

To facilitate the quantization of the Dirac field we eliminate from Eq. (4) the vector fields by employing the massive spin-1 propagator in Coulomb gauge

$$\mathbf{V}_0(\mathbf{x})=g_v\int d^3\mathbf{y}\, G(\mathbf{x},\mathbf{y}))\psi^+(\mathbf{y})\frac{\mathbf{T}}{2}\psi(\mathbf{y}), \tag{6}$$

where $G(\mathbf{x},\mathbf{y})$ is given by

$$G(\mathbf{x},\mathbf{y})=G(\,|\mathbf{x}-\mathbf{y}|\,)=(1/4\pi)(e^{-M_V|\mathbf{x}-\mathbf{y}|}/|\mathbf{x}-\mathbf{y}|). \tag{7}$$

The Hamiltonian of the system, defined as the timelike component of the energy-momentum tensor $T^{\mu\nu}$ derived from the Lagrangian (1) is now expressed in the form

$$\begin{aligned}H[\psi,V_0(\psi)]&=\int d^3\mathbf{x}\, T^{00}\\&=\int d^3\mathbf{x}\,\psi^+(\boldsymbol{\alpha}\cdot\mathbf{p}+\beta m)\psi\\&\quad+g_v\int d^3\mathbf{x}\,\psi^+\frac{\mathbf{T}}{2}\psi\cdot\mathbf{V}_0\\&\quad-\sum_{i=1}^{N^2-1}\int d^3\mathbf{x}[(\nabla V_0^i)^2+M_v^2V_0^{i^2}].\end{aligned} \tag{8}$$

By employing the field equation (4) and relation (6) the corresponding second quantized Hamilton operator reads

$$\begin{aligned}\hat{H}&=\int d^3\mathbf{x}:\hat{\psi}^+(\mathbf{x})(\boldsymbol{\alpha}\cdot\mathbf{p}+\beta m)\hat{\psi}(\mathbf{x}):\\&\quad+\frac{g_v^2}{2}\int d^3\mathbf{x}\,d^3\mathbf{y}:\hat{\psi}^+(\mathbf{x})\\&\quad\times\frac{\mathbf{T}}{2}\hat{\psi}(\mathbf{x})G(\mathbf{x},\mathbf{y})\hat{\psi}^+(\mathbf{y})\frac{\mathbf{T}}{2}\hat{\psi}(\mathbf{y}):,\end{aligned} \tag{9}$$

where we have introduced normal ordering of the fermion fields in order to substract the vacuum energy. In the next step we expand the field operators $\hat{\psi}$ and $\hat{\psi}^+$ in a complete orthonormal basis, following the notation used in Ref. 2:

$$\begin{aligned}\hat{\psi}(\mathbf{x})&=\sum_{p>F}\hat{b}_p\varphi_p(\mathbf{x})+\sum_{a<F}\hat{d}_a^+\Phi_a(\mathbf{x}),\\\hat{\psi}^+(\mathbf{x})&=\sum_{p>F}\hat{b}_p^+\varphi_p^+(\mathbf{x})+\sum_{a<F}\hat{d}_a\Phi_a^+(\mathbf{x}),\end{aligned} \tag{10}$$

where the sums in these expansions run over a complete set of states. The $\hat{b}_p$, $\hat{b}_p^+$ ($\hat{d}_a$, $\hat{d}_a^+$) are one-particle (antiparticle) annihilation and creation operators, respectively. The summation indices p and a represent the quantum numbers of the different states. In a spherically symmetric basis, these are a radial quantum number ν, angular momentum quantum number l, spin projection quantum number σ, and internal SU(N) quantum number τ running from 1 to N while F denotes the Fermi level. In the next step we specify the variational ground state $|\mathrm{GS}\rangle$ of the system by the action of a product of one-particle operators acting on the vacuum state:

$$|\mathrm{GS}\rangle=\sum_{p=1}^{n}\hat{b}_p^+\,|0\rangle. \tag{11}$$

The physical picture we adopt is that there are only particles in the system and no antiparticles. To be specific, we considered a composite object consisting of $2N$ fermions in the ground state, i.e., the index p in Eq. (11) takes on the values $n=2N$

$$\begin{aligned}&p=\{\nu=1,\ l=0;\sigma,\tau\},\\&\sigma=\pm\tfrac{1}{2},\quad 1\leqslant\tau\leqslant N,\end{aligned} \tag{12}$$

and thus Eq. (11) contains $2N$ creation operators for particles. These operators generate the effective orbitals representing the Hartree–Fock ground state. In order to find the yet unknown orbitals $\varphi_{\bar{p}}$ we start with the following variational equation:

$$\frac{\delta}{\delta\varphi_{\bar{p}}^+(\mathbf{y})}\Big\{\langle\mathrm{GS}|\hat{H}|\mathrm{GS}\rangle-\sum_{p=1}^{n}\omega_p\int d^3\mathbf{x}\,\varphi_p^+(\mathbf{x})\varphi_p(\mathbf{x})\Big\}=0, \tag{13}$$

where the ω_p's are Lagrange multipliers which enforce the orthonormality of the single-particle orbitals φ_p. Evidently one has to evaluate the matrix element of $\hat{H}$ with respect to the ground state. Following Ref. 2 we insert the expansions (10) into the normal ordered Hamiltonian (9) and employ the action of the particle–hole operators on the ground state $|\mathrm{GS}\rangle$. Thus we obtain

$$\begin{aligned}\langle\mathrm{GS}|\hat{H}|\mathrm{GS}\rangle&=\sum_p\int d^3\mathbf{x}\,\varphi_p^+(\mathbf{x})(\boldsymbol{\alpha}\cdot\mathbf{p}+\beta m)\varphi_p(\mathbf{x})+\frac{g_v^2}{2}\int d^3\mathbf{x}\,d^3\mathbf{y}\Big\{\sum_{p,p'}\Big[\varphi_p^+(\mathbf{x})\frac{\mathbf{T}}{2}\varphi_p(\mathbf{x})\Big]\\&\quad\cdot G(\mathbf{x},\mathbf{y})\cdot\Big[\varphi_{p'}^+(\mathbf{y})\frac{\mathbf{T}}{2}\varphi_{p'}(\mathbf{y})\Big]-\sum_{p,p'}\Big[\varphi_p^+(\mathbf{x})\frac{\mathbf{T}}{2}\varphi_{p'}(\mathbf{x})\Big]\cdot G(\mathbf{x},\mathbf{y})\cdot\Big[\varphi_{p'}^+(\mathbf{y})\frac{\mathbf{T}}{2}\varphi_p(\mathbf{y})\Big]+\cdots\Big\},\end{aligned} \tag{14}$$

where the dots at the end in this expression represent terms containing at least one hole operator ($\hat{d}_a$, $\hat{d}_a^+$) which vanish since the ground state (11) contains no antiparticles. We note that the above summations run over occupied states only, i.e., over the $2N$ fermion states. One can identify the first and second term in the above equation with the direct and exchange term, respectively. Since the interaction is spin-independent the summations over spin projections σ and SU(N) quantum number τ can be performed separately. For the direct (D) and exchange (E) term we obtain, respectively:

$$D=\frac{1}{4}\sum_{\sigma,\sigma'}[\varphi_\sigma^+(\mathbf{x})\varphi_\sigma(\mathbf{x})]G(\mathbf{x},\mathbf{y})[\varphi_{\sigma'}^+(\mathbf{y})\varphi_{\sigma'}(\mathbf{y})]S_D, \tag{15}$$

with

$$S_D=\sum_{\tau,\tau'=1}^{N}(u_\tau^+\mathbf{T}u_\tau)\cdot(u_{\tau'}^+\mathbf{T}u_{\tau'})=\sum_{i=1}^{N^2-1}(\mathrm{Tr}\,T_i)^2 \tag{16}$$

and

$$E=\frac{1}{4}\sum_{\sigma,\sigma'}[\varphi_\sigma^+(\mathbf{x})\varphi_{\sigma'}(\mathbf{x})]G(\mathbf{x},\mathbf{y})[\varphi_{\sigma'}^+(\mathbf{y})\varphi_\sigma(\mathbf{y})]S_E, \tag{17}$$

with

$$S_E=\sum_{\tau,\tau'=1}^{N}(u_\tau^+\mathbf{T}u_{\tau'})\cdot(u_{\tau'}^+\mathbf{T}u_\tau)=\sum_{i=1}^{N^2-1}\mathrm{Tr}\,T_i^2. \tag{18}$$

In the above expressions the φ_σ's denote the fermion wave functions in coordinate representation, while the u_τ's are N-component fundamental basis vectors of SU(N). After performing the summations in (16) and (18) we obtain for S_D and S_E

$$S_D=0,\quad S_E=2(N^2-1). \tag{19}$$

Thus we see that the direct term (D) vanishes, while the exchange term (E) is different from zero only for parallel spin orientations, i.e., the summations over spin projections σ,σ' in Eq. (17) provide a factor 2. With these ingredients the energy matrix element (14) reads

$$\langle\mathrm{GS}|\hat{H}|\mathrm{GS}\rangle = 2N\int d^3\mathbf{x}\,\varphi_p^+(\mathbf{x})(\boldsymbol{\alpha}\cdot\mathbf{p}+\beta m)\varphi_p(\mathbf{x}) - [(N^2-1)/2]g_v^2\int d^3\mathbf{x}\,d^3\mathbf{y}[\varphi_p^+(\mathbf{x})\varphi_p(\mathbf{x})]G(\mathbf{x},\mathbf{y})[\varphi_p^+(\mathbf{y})\varphi_p(\mathbf{y})]. \tag{20}$$

This is just the total energy of the $2N$-fermion system, i.e., the Hartree–Fock energy E_{HF}. After inserting the above matrix element in the variational equation (13) we obtain the following single-particle equation for one of the $2N$ fermions:

$$\Big\{\boldsymbol{\alpha}\cdot\mathbf{p}+\beta m-\frac{N^2-1}{2N}g_v^2\times\int d^3\mathbf{y}\,\varphi_{\bar p}^+(\mathbf{y})\varphi_{\bar p}(\mathbf{y})G(\mathbf{x},\mathbf{y})\Big\}\varphi_{\bar p}(\mathbf{x})=\omega_{\bar p}\varphi_{\bar p}(\mathbf{x}), \tag{21}$$

where $\bar p$ represents the quantum numbers of the specific fermion state. Next we discuss and interpret the last results diagrammatically. To do this, we first rewrite the total energy E_{HF} from (20) by employing the single-particle equation (21) and get

$$\begin{aligned}E_{\mathrm{HF}}&=2N\omega_{\bar p}\int d^3\mathbf{x}\,\varphi_{\bar p}^+(\mathbf{x})\varphi_{\bar p}(\mathbf{x})+\frac{N^2-1}{2}g_v^2\\&\quad\times\int d^3\mathbf{x}\,d^3\mathbf{y}\,\varphi_{\bar p}^+(\mathbf{y})\varphi_{\bar p}(\mathbf{y})\\&\quad\times G(\mathbf{x},\mathbf{y})\varphi_{\bar p}^+(\mathbf{x})\varphi_{\bar p}(\mathbf{x})\\&=2N\omega_{\bar p}+\frac{N^2-1}{2}g_v^2\int d^3\mathbf{x}\,d^3\mathbf{y}\,\varphi_{\bar p}^+(\mathbf{y})\varphi_{\bar p}(\mathbf{y})\\&\quad\times G(\mathbf{x},\mathbf{y})\varphi_{\bar p}^+(\mathbf{x})\varphi_{\bar p}(\mathbf{x}).\end{aligned} \tag{22}$$

In the last step of the above derivation we made use of the normalization condition of the fermion wave function. Consequently, the total energy consists of two terms: the first representing the sum of all single-particle energies $\omega_{\bar p}$, while the second denotes the exchange energy. Since the Hartree–Fock approximation is just the first order correction to the exact ground-state energy, the effective interaction is described by the two many-body diagrams with one wiggly line of Fig. 1. The bubble diagram [Fig. 1(a)] is proportional to the trace of the SU(N) generators and thus vanishes. In other words since this diagram includes one closed fermion line which represents a SU(N) singlet there is no interaction possible. In the case of the open-oyster diagram [Fig. 1(b)] there are exactly (N^2-1) independent interactions possible because the gauge field has (N^2-1) components in the adjoint representation of the SU(N) group. Thus it becomes

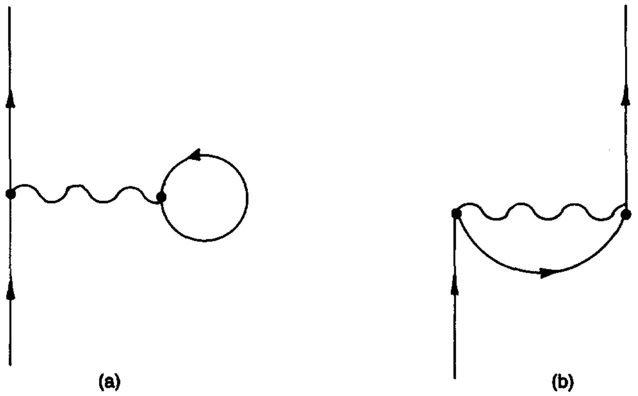

FIG. 1. Many-body diagrams describing the effective interaction in the Hartree–Fock approximation: (a) direct interaction (bubble-diagram) and (b) exchange interaction (open-oyster diagram).

clear without any calculation that the direct interaction provides no correction to the ground state energy while the exchange interaction contributes with a factor (N^2-1).

III. RADIAL EQUATIONS AND NUMERICAL ANALYSIS

In the following we first shall rewrite the single-particle equation (21) in a more convenient form by transforming the corresponding integrodifferential equation in terms of two coupled differential equations. We define a new function $u(\mathbf{x})$ according to

$$u(\mathbf{x}) = -g_v \int d^3y\, \varphi_{\bar p}^+(\mathbf{y})\varphi_{\bar p}(\mathbf{y})G(\mathbf{x},\mathbf{y}). \tag{23}$$

With this definition we obtain the following coupled equations for the two functions $\varphi_{\bar p}(\mathbf{x})$ and $u(\mathbf{x})$:

$$[\boldsymbol{\alpha}\cdot\mathbf{p} + \beta m + [(N^2-1)/(2N)]g_v u(\mathbf{x})]\varphi_{\bar p}(\mathbf{x}) = \omega_{\bar p}\varphi_{\bar p}(\mathbf{x}), \tag{24}$$

$$(\Delta - M_V^2)u(\mathbf{x}) = g_v \varphi_{\bar p}^+(\mathbf{x})\varphi_{\bar p}(\mathbf{x}). \tag{25}$$

The last equation results by multiplication of Eq. (23) with the operator $(\Delta - M_V^2)$ from the left. It is interesting to compare the field equations (24) and (25) with the corresponding stationary field equations which follow from the classical field theory defined by

$$\mathscr{L} = \bar\chi(i\gamma_\mu\partial^\mu - m)\chi - \frac{1}{2}[(\partial_\mu A_\nu - \partial_\nu A_\mu)\partial^\mu A^\nu - M^2 A_\mu A^\mu] - \lambda\bar\chi\gamma^\mu\chi A_\mu, \tag{26}$$

describing the coupling of the Dirac field χ with an Abelian Proca field A_μ. In the case of a vector field containing the longitudinal component only, i.e., $A_\mu = (A_0,0)$ and assuming $\partial_t A_0 = 0$ there follow from (26) the field equations

$$[\boldsymbol{\alpha}\cdot\mathbf{p} + \beta m + \lambda A_0(\mathbf{x})]\chi(\mathbf{x}) = \epsilon\chi(\mathbf{x}), \tag{27}$$

$$(\Delta - M^2)A_0(\mathbf{x}) = -\lambda\chi^+(\mathbf{x})\chi(\mathbf{x}). \tag{28}$$

By comparing the last two equations with Eqs. (24) and (25) and identifying $\varphi_{\bar p}$ and u with χ and A_0, respectively, it is evident that they are identical except for the negative sign on the right-hand side of Eq. (28). Note that the factor $(N^2-1)/(2N)$ in Eq. (24) can be formally eliminated by making the redefinitions

$$g_v \to \sqrt{(N^2-1)/2N}\, g_v \quad\text{and}\quad u(\mathbf{x}) \to \sqrt{(N^2-1)/2N}\, u(\mathbf{x}).$$

The reason for the opposite signs in Eqs. (25) and (28) is to be understood as a direct consequence of the internal fermionic SU(N) degrees of freedom which are responsible for the fact that the resulting effective interaction becomes attractive. This is analogous to the different possible signs of the nucleon–nucleon interaction depending on the total isospin of the nucleon–nucleon system. The attraction becomes maximal for the closed-shell configuration [SU(N)-singlet] which we are considering. As a consequence of the fact that the coupling constant λ from (27), (28) appears with a different relative sign in these equations in contrast to the corresponding g_v from (24), (25), Eqs. (27), (28) do not admit bound-state solutions with $-m \leqslant \epsilon \leqslant m$ for real values of λ and M. We shall prove this statement later explicitly in the limit $M_V \to \infty$.

Returning to Eqs. (24) and (25) we derive the corresponding radial equations by employing the general form of s-wave functions $(l=0)$ with spin up $(\sigma = 1/2)$ and spin down $(\sigma = -1/2)$, respectively:

$$\varphi_{1/2}(\mathbf{x}) = \frac{1}{\sqrt{4\pi}}\begin{bmatrix} ig(r) \\ 0 \\ f(r)\cos\vartheta \\ f(r)\sin\vartheta e^{i\varphi}\end{bmatrix}, \tag{29}$$

and

$$\varphi_{-1/2}(\mathbf{x}) = \frac{1}{\sqrt{4\pi}}\begin{bmatrix} 0 \\ ig(r) \\ f(r)\sin\vartheta e^{-i\varphi} \\ -f(r)\cos\vartheta\end{bmatrix}. \tag{30}$$

Next we perform the transformations $G = r\cdot g$, $F = r\cdot f$ and $U = r\cdot u$ and after inserting (29), (30) into Eqs. (24), (25) we obtain the following coupled equations for the radial functions G, F and U:

$$\frac{d}{dr}G = \frac{1}{r}G + \left(\omega + m - \frac{N^2-1}{2N}g_v\frac{U}{r}\right)F, \tag{31}$$

$$\frac{d}{dr}F = -\frac{1}{r}F - \left(\omega - m - \frac{N^2-1}{2N}g_v\frac{U}{r}\right)G, \tag{32}$$

$$\frac{d^2}{dr^2}U = M_V^2 U + \frac{1}{4\pi}g_v\frac{G^2+F^2}{r}. \tag{33}$$

In addition, we have the normalization condition for the fermion wave function, i.e.,

$$\int d^3x\, \varphi_{1/2}^+\varphi_{1/2} = \int_0^\infty dr(G^2+F^2) = 1. \tag{34}$$

For the numerical analysis we have used the general purpose computer code COLSYS.[4] The above eigenvalue problem has been converted into a system of ordinary differential equations with boundary conditions expressed at the two end points by defining a new function $N(r)$

$$N(r) = \int_0^r dr'(G^2+F^2), \tag{35}$$

which, employing the normalization condition (36), leads to the additional differential equation

$$\frac{d}{dr}N = G^2 + F^2, \tag{36}$$

with the boundary conditions

$$N(0) = 0 \quad\text{and}\quad N(\infty) = 1. \tag{37}$$

Furthermore we regarded the coupling constant g_v as an additional independent function, satisfying

$$\frac{d}{dr}g_v = 0, \tag{38}$$

which is to be determined for a given eigenvalue ω. In this way we have converted the eigenvalue problem (31)–(33) into a boundary value problem which is defined by the five differential equations (31)–(33), (36), and (38) with the corresponding boundary conditions (34) together with the integrability conditions $G(0) = F(\infty) = U(\infty) = 0$.

IV. THE LIMIT $M_V \to \infty$

In passing to the limit of a heavy boson it is useful to consider the integro-differential Eq. (21) and to express the propagator in momentum space. We consider the interaction term from (21), hereafter denoted to as M_I

$$
\begin{aligned}
M_I &= -\frac{N^2-1}{2N} g_v^2 \int d^3\mathbf{y}\, G(\mathbf{x},\mathbf{y}) \varphi^+_{1/2}(\mathbf{y}) \varphi_{1/2}(\mathbf{y}) \\
&= -\frac{N^2-1}{2N} g_v^2 \int d^3\mathbf{y} \left[\int \frac{d^3\mathbf{k}}{(2\pi)^{3/2}} \widetilde{G}(\mathbf{k}) \cdot e^{-i\mathbf{k}\cdot(\mathbf{x}-\mathbf{y})} \right] \varphi^+_{1/2}(\mathbf{y}) \varphi_{1/2}(\mathbf{y}) \\
&\to -[(N^2-1)/(2N)](g_v/M_V)^2 [\varphi^+_{1/2}(\mathbf{x})\varphi_{1/2}(\mathbf{x})] \quad \text{(for large } M_V\text{)},
\end{aligned} \tag{39}
$$

where we inserted the momentum-space propagator $\widetilde{G}(\mathbf{k}) \sim (|\mathbf{k}|^2 + M_V^2)^{-1} \to 1/M_V^2$, for $M_V \gg |\mathbf{k}|$. The last expression in the above equation is valid only in the case that the momentum transfer is negligibly small compared with the mass of the gauge boson M_V. The index 1/2 of the fermion wave function indicates the considered fermion with spin up. Thus in this limiting case Eq. (21) becomes

$$
\left\{ \alpha\cdot\mathbf{p} + \beta m - \frac{N^2-1}{2N} \left(\frac{g_v}{M_V}\right)^2 \varphi^+_{1/2}(\mathbf{x})\varphi_{1/2}(\mathbf{x}) \right\} \varphi_{1/2}(\mathbf{x}) = \omega_{1/2}\varphi_{1/2}(\mathbf{x}). \tag{40}
$$

This is just the nonlinear Dirac-equation with vector coupling belonging to the Fermi-type effective Lagrangian

$$
\mathscr{L} = \overline{\varphi}(i\gamma_\mu \partial^\mu - m)\varphi + g_1(\overline{\varphi}\gamma^\mu\varphi)^2, \tag{41}
$$

considering only the $\mu = 0$ component of the current and identifying $(g_v/M_V)^2 = 4N/(N^2-1)g_1$.

The nonlinear Dirac equation resulting from (41) has been extensively studied by several authors (see, for example, Refs. 5 and 6) and it is well known that it admits bound-state solutions for $g_1 > 0$. To be more precise we remark that the interaction Lagrangian $\sim g_1(\overline{\varphi}\gamma^\mu\varphi)^2$ which is positive definite for $g_1 > 0$ can be interpreted as an attractive potential $V = -g_1(\overline{\varphi}\gamma^\mu\varphi)^2$ which is responsible for the existence of bound states in this theory. It also becomes now clear why the classical field theory defined by the Lagrangian (26) does not have bound state solutions, due to the relative sign of the coupling constant in Eqs. (27), (28). The corresponding interaction term in Eq. (39) would have an opposite sign and thus one would have to identify $-(g_v/M_V)^2 \sim g_1 > 0$, which is not possible for real values of g_v and M_V.

V. NUMERICAL SOLUTIONS AND DISCUSSION OF THE RESULTS

We now turn the discussion to the solutions of the coupled equations (24), (25) for the wave functions and the single-particle energies ω associated with fermion bound states. We have analyzed numerically the radial equations (31)–(33), (36), and (38), searching for those combinations of parameters (g_v, M_V) that provide a particular fermion energy $\omega \in [-m, m]$. In addition, we also solved the coupled equations corresponding to the Lagrangian (41), in order to analyze the limiting case $M_V \to \infty$, i.e.,

$$
\frac{d}{dr}\begin{pmatrix} w \\ v \end{pmatrix} = \begin{pmatrix} 1/r & \omega + m + g_1(w^2+v^2)/r^2 \\ -\omega + m - g_1(w^2+v^2)/r^2 & -1/r \end{pmatrix} \cdot \begin{pmatrix} w \\ v \end{pmatrix}, \tag{42}
$$

where $w(r)$ and $v(r)$ are radial $s_{1/2}$-wave functions. Table I indicates the values of the coupling constant g_1 from Eq. (42) required to support bound states in this system.

First, we note that we could not determine stable bound states outside the energy range $0.944m \leqslant \omega \leqslant m$ for this system. For ω values lower than $\omega_{\text{crit}} = 0.944m$ we found that this system is not stable and thus there are no nodeless, normalizable solutions for $\omega < \omega_{\text{crit}}$. An analysis of stability problems of nonlinear Dirac equations is presented in Refs. 1, 7, and 8. In addition it is very interesting to remark that the largest g_1 values from Table I correspond to the weakest bound states. The minimal g_1 value required to support bound states is related to the minimal energy $\omega_{\text{crit}} = 0.944m$. We have found indications that at the critical point $(g_{\text{crit}}, \omega_{\text{crit}})$ the solutions merges with a second branch of nodeless localized solutions which for other values of the coupling constant is less deeply bound. Figure 2 indicates typical solutions of the fermion system interacting via IVB for masses $M_V \leqslant 10m$. For the depicted state with $\omega = 0.95m$ belonging to a four-fermion system ($N = 2$) the radial extension of the radial wave functions does not exceed

TABLE I. Values for the coupling constant g_1 required to form a bound state in the system with a $g_1(\varphi^+\varphi)^2$ interaction. Here, g_1 has the dimension $(\text{mass})^{-2}$.

ω $[m]$	g_1 $[m]^{-2}$
0.99	31.5
0.98	20.8
0.97	15.7
0.96	12.3
0.95	9.6
0.944	7.3

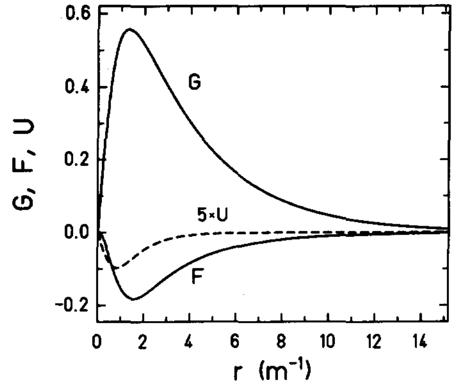

FIG. 2. Radial functions for the $s_{1/2}$ state of the four-fermion system in the case $M_V = 7m$, $\omega = 0.95m$, $g_v = 39.19$.

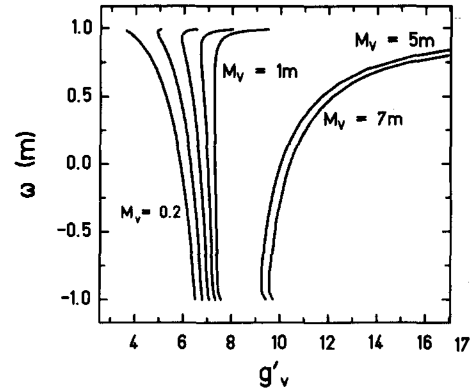

FIG. 4. The dependence of the fermion energy ω on the coupling constant $g_v' = \sqrt{(N^2-1)/2N} \cdot g_v$ for different masses $(M_V/m) = 0.2, 0.4, 0.6, 0.8, 1, 5, 7$ in the case of a four-fermion system.

$10m^{-1} - 15m^{-1}$. The function $U(r) = ru(r)$, where $u(r)$ is defined by Eq. (23) acts as an attractive potential allowing for bound states. However, we expect from the discussion in Sec. IV that this system is effectively described by the nonlinear Dirac equation which follows from the Lagrangian (41) for increasing values of M_V and low momentum transfer. In this case the functions $G(r)$ and $F(r)$ have to increase in magnitude such that the bound states are supported by the nonlinearity $(G^2 + F^2)/r^2$.

Figure 3 shows this effect for the two states with $M_V = 7m$ (dashed line) and $M_V = 52m$ belonging to the fermion energy $\omega = 0.95m$. It is clear from this figure that the vector density $\varphi^+\varphi \sim (G^2 + F^2)/r^2$ increases with increasing M_V. In addition, we note that the radial extension of these densities does not exceed about $3m^{-1}$.

In Fig. 4 we depict the dependence of the fermion eigenenergy ω on the coupling constant $g_v' = \sqrt{(N^2-1)/(2N)} \cdot g_v$ for different IVB masses $(M_V/m) = 0.2, 0.4, 0.6, 0.8, 1, 5, 7$. Evidently, there exist bound states in the whole interval $-m \leqslant \omega \leqslant m$. Concerning the curves with $M_V =$ const. we remark that there is a degeneracy of the coupling constant g_v with ω for $M_V \geqslant 0.4m$. We also note that the curves with $0.4m \leqslant M_V < 5m$ possess two turning points corresponding to Dirac eigenenergies $0.9m < \omega < m(g_{TP})$ and $-m < \omega < -0.9m(g_{tp})$ while the curves belonging to $M_V = 5m$ and $7m$ display only one turning point corresponding to fermion energies near the lower energy continuum and the curve $M_V = 0.2m$ with no turning point. Note that with increasing values of M_V the upper turning points g_{TP} move to lower energies.

Next we analyze the dependence of the fermion energy ω on the coupling constant g_v for increasing values of M_V, i.e., $M_V \gg 10m$. We expect that energy levels with $\omega > 0.944m$ asymptotically approach the energy values provided in Table I for the case of large M_V values, since for the system bound by contact forces only for $\omega > 0.944m$ we found bound states. Figure 5 illustrates this limiting process for the constant energy values $(\omega/m) = 0.95$, 0.97, and 0.99. In this figure we show the dependence of the effective coupling constant $g_{\text{eff}} = (g_v/M_V)^2 \cdot (N^2-1)/(4N)$ on the mass M_V for fixed value of the energy. In the previous section we demonstrated that the coupling constant g_{eff} tends to the coupling constant g_1 of the contact interaction in the case $M_V \gg |\mathbf{k}|$. In

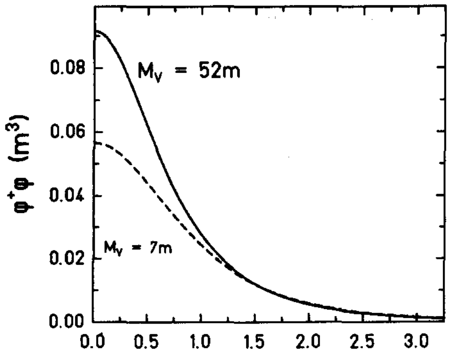

FIG. 3. Vector densities $\varphi^+\varphi = (G^2 + F^2)/(4\pi r^2)$ of the four-fermion system for the two IVB masses $M_V = 7m$ (dashed line) and $M_V = 52m$ (solid line) corresponding to the fermion energy $\omega = 0.95m$.

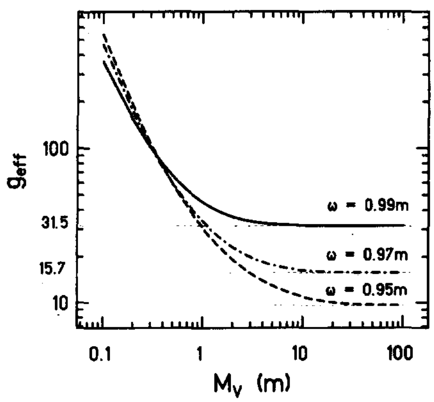

FIG. 5. The limiting process of increasing M_V values for $\omega > 0.944m$. The effective coupling constant $g_{\text{eff}} = (3/8) \cdot (g_v/M_V)^2$ tends to the coupling constant g_1 of the contact interaction from Table I with increasing values of M_V. Three typical situations in a four-fermion system are illustrated: $\omega = 0.99m$, $\omega = 0.97m$, $\omega = 0.95m$.

Fig. 5 the horizonal dashed lines represent the values of g_1 which provide the same energy eigenvalue ω. Obviously, the energy levels of the system interaction via IVB approach these energies for M_V values greater than $20m$.

It is very interesting to analyze the nature of this limiting process for ω values lower than $\omega_{\rm crit} = 0.944m$, since in this case we did not find any localized solutions for the system interacting via contact forces. Figure 6 illustrates typical features of such solutions with $\omega < \omega_{\rm crit}$ and large M_V. The shape of the radial functions is similar to that of the solutions belonging to $\omega \geqslant 0.944m$, in particular the curves are also nodeless. However, the radial extension of these solutions shrinks drastically, i.e., it becomes smaller by more than one order of magnitude, while the functions G and F take on very large values due to the normalization condition. In fact, Fig. 7 indicates that the corresponding radial extension of the vector density $\varphi^+\varphi \sim (G^2 + F^2)/r^2$ does not exceed $0.07m^{-1}$. A further peculiar feature of this new class of solutions, obtained for $\omega < \omega_{\rm crit} = 0.944m$ and large M_V, is the fact that for masses M_V larger than some critical value $12m < M_{\rm crit} < 13m$ the required coupling constant g_v is no more proportional to M_V. We found that for this class of solutions the g_v values become almost constant with increasing values of the corresponding M_V. Consequently, the effective coupling constant $g_{\rm eff} \sim (g_v/M_V)^2$ tends to zero in this case! The illustration of this effect is found in Fig. 8 for the $N = 2$ states of constant energy $\omega/m = 0.5, 0, -0.5$. First, with increasing values of M_V, g_v also increases up to $M_V \simeq M_{\rm crit}$ and then decreases. Asymptotically it tends to the constant value of about 12.35 while M_V increases further. We found that in this limiting process the coupling constant g_v tends to the same limiting value $g_\infty = 12.35$ for each solution in the range $-m < \omega < 0.944m$. These solutions become ω-independent. Thus for $\omega < \omega_{\rm crit} = 0.944m$ and $M_V > M_{\rm crit}$ there acts some mechanism which ultimately effects a collapse of the system.

It is of particular interest to analyze this type of extremely localized states belonging to fermion eigenenergies $-m \leqslant \omega < 0.944m$. In Sec. IV we demonstrated that only in the case that the momentum transfer is significantly lower that the boson mass, the considered field theory is effectively described by the Fermi contact theory. The solutions with $\omega < \omega_{\rm crit}$ describe the situation when the momentum transfer is of the same order of magnitude as the boson mass. As mentioned before, the transition to the limiting case of large IVB masses as discussed in Sec. IV is no longer valid. To achieve a better understanding of the mechanism which causes the shrinking of these solutions we perform the following scale transformations in Eqs. (31)–(34):

$$x = Mr, \quad \widetilde{G} = (1/\sqrt{M_V})G, \quad \widetilde{F} = (1/\sqrt{M_V})F, \qquad \widetilde{U} = (4\pi/g_v)U, \quad \tilde{g}_v = [(N^2-1)/8\pi N]g_v^2. \tag{43}$$

With the above definitions Eqs. (31)–(34) become

$$\frac{d}{dx}\widetilde{G} = \frac{1}{x}\widetilde{G} - \tilde{g}_v \frac{1}{x}\widetilde{U}\widetilde{F} + \frac{\omega + m}{M_V}\cdot\widetilde{F}, \tag{44}$$

$$\frac{d}{dx}\widetilde{F} = -\frac{1}{x}\widetilde{F} + \tilde{g}_v \frac{1}{x}\widetilde{U}\widetilde{G} - \frac{\omega - m}{M_V}\cdot\widetilde{G}, \tag{45}$$

$$\frac{d^2}{dx^2}\widetilde{U} = \widetilde{U} + \frac{1}{x}(\widetilde{G}^2 + \widetilde{F}^2), \tag{46}$$

with the normalization condition

$$\int_0^\infty dx(\widetilde{G}^2 + \widetilde{F}^2) = 1. \tag{47}$$

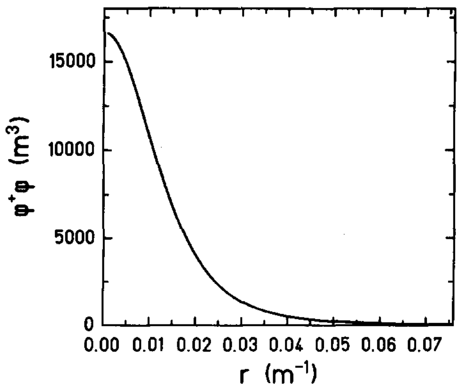

FIG. 7. The corresponding vector density $\varphi^+\varphi = (G^2 + F^2)/(4\pi r^2)$ of the solutions from Fig. 6. The parameters M_V, ω, and g_v are the same of Fig. 6.

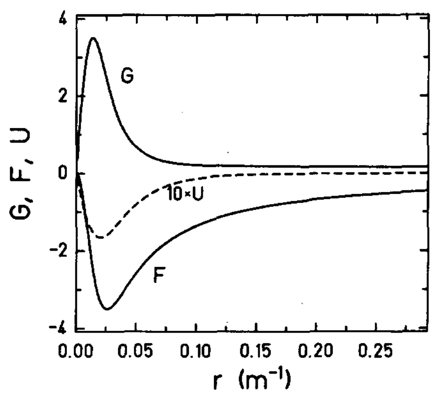

FIG. 6. Radial functions for the $s_{1/2}$ state of the four-fermion system belonging to the second class of solutions we have found for $\omega < \omega_{\rm crit} = 0.944m$. We have depicted a typical case with $M_V = 52m$, $\omega = 0.5m$, $g_v = 12.74$.

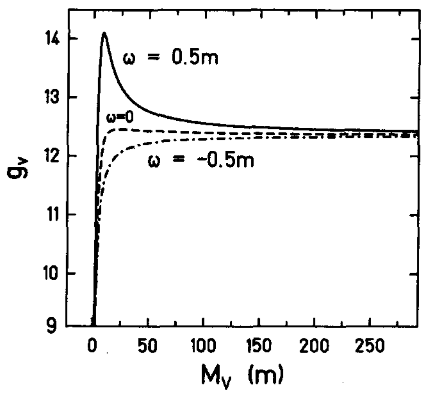

FIG. 8. The dependence of the coupling constant g_v on the IVB mass M_V for $\omega/m = 0.5, 0, -0.5$ in a four-fermion system which is characteristic for the second type of solutions we found, i.e., $\omega < 0.944m$.

We remark that the IVB mass M_V appears only on the right-hand side of Eqs. (44), (45) in the demonimator. The crucial point is to observe that the limiting process $M_V \to \infty$ is equivalent to the situation with M_V fixed while $m \to 0$ and $\omega \to 0$. Of course, since for bound states it is $-m \leqslant \omega \leqslant m$, there follows from $m \to 0$ that also $\omega \to 0$, i.e., the single-particle energy gap $[-m,m]$ shrinks to zero. By solving Eqs. (44)–(47) for $(m,\omega) = (0,0)$ we obtained a nodeless, stable solution corresponding to the coupling constant $\tilde{g}_v = 9.1$. Employing the relation between $\tilde{g}_v$ and g_v from Eq. (43) it is evident that this value of $\tilde{g}_v(\omega \to 0, m \to 0)$ perfectly agrees with the value of $g_v(M_V \to \infty)$, i.e., $g_\infty = 12.35$. Consequently, all solutions belonging to $-m \leqslant \omega < \omega_{\text{crit}}$ can be constructed from the solution of Eqs. (44)–(47) with m and ω equal zero by employing the scale transformations (43). Thus we found that for this class of states the IVB mass sets the scale in the system, i.e., different solutions of the unscaled Eqs. (31)–(34) belonging to different eigenenergies ω contain no additional information about the system since they represent a unique solution which is simply scaled according to Eq. (43).

Finally we consider the total energy of the system. In Sec. II we demonstrated that the total energy E_{HF} is given by the sum of the Dirac eigenenergies ω and a first order correction associated with the exchange interaction. Considering Eq. (22) the total energy reads

$$\begin{aligned} E_{\text{HF}} &= 2N\omega_{\bar{p}} + \frac{N^2-1}{2} g_v^2 \\ &\quad \times \int d^3\mathbf{x}\, d^3\mathbf{y}\, \varphi_{\bar{p}}^+(\mathbf{x}) \varphi_{\bar{p}}(\mathbf{x}) \\ &\quad \times G(\mathbf{x},\mathbf{y}) \varphi_{\bar{p}}^+(\mathbf{y}) \varphi_{\bar{p}}(\mathbf{y}) \\ &= 2N\omega_{\bar{p}} - \frac{N^2-1}{2} g_v \int d^3\mathbf{x}\, \varphi_{\bar{p}}^+(\mathbf{x}) \\ &\quad \times \varphi_{\bar{p}}(\mathbf{x}) u(\mathbf{x}), \end{aligned} \tag{48}$$

where we introduced the function $u(\mathbf{x})$ according to Eq. (23). Performing the above integration in spherical coordinates the integration over the angular variables can be carried out yielding the following final expression for the total energy:

$$E_{\text{HF}} = 2N\omega_{\bar{p}} - \frac{N^2-1}{2} g_v \int_0^\infty dr[(G^2+F^2)\cdot u], \tag{49}$$

where G, F, and u are the radial functions representing the solution of the coupled equations (31)–(33). Considering Eq. (23) which defines the function u, we see that the radial integral in the above expression is negative definite for positive values of the coupling constant g_v. Consequently, the total energy E_{HF} is shifted above the sum of single-particle energies $\omega_{\bar{p}}$. The first term in Eq. (49) represents a lower bound of E_{HF}. Furthermore, we found in all examples we have studied that the total energy becomes never negative, due to the absolute values of the second term in Eq. (49) which always exceeds possible negative contributions from the first term, i.e., for fermion eigenenergies in the range $-m \leqslant \omega_{\bar{p}} < 0$. Consequently, the total energy is positive definite and thus the neutral vacuum represents a stable groundstate of this field theory. Finally we note that with increasing IVB masses the contributions to the total energy coming from the radial integral in Eq. (49) become quite large such that the total energy lies above the threshold energy $2N\cdot m$ in the upper energy continuum. In this case the fermion system does not form a pure bound state. The solutions discussed above, however, possibly correspond to a resonance state of the system.

VI. CONCLUSIONS

Let us summarize the main results of our studies. We have investigated a quantum field theoretical model for a system consisting of $2N$ fermions with internal SU(N) degrees of freedom interacting via massive vector bosons which possesses stable bound states. We are not able to decide whether the energy of a pair of N-fermion clusters would have a lower energy than the considered $2N$-fermion system, since our energy relation (49) is valid only for closed shell systems. Consequently, we cannot decide whether these $2N$-fermion clusters are really stable against the decay into N-fermion clusters, nor can we make any assertions about the mass of such N-fermion clusters. However, in view of the high localization of our solutions we expect them to be at least quasistable against a decay into lighter clusters. For this second quantized system we derived a single-particle equation for fermions by employing the Hartree–Fock approximation. Furthermore we have examined the limiting process for increasing IVB masses, for which we demonstrated analytically as well as by numerical integration that the effective interaction is described by Fermi-type contact forces in the case that the IVB mass dominates the momentum transfer. By solving the corresponding radial equations we found two classes of solutions, namely one class which relates to bound states of the nonlinear Dirac equation in the limit $M_V \to \infty$ ($\omega \geqslant 0.944m$) while the second class of solutions ($\omega < 0.944m$) indicates a collapse of the system for large IVB masses. The second class of solutions represents just the situation in which the IVB mass and the momentum transfer are of the same order of magnitude. The total energy of the system was found to be positive in all examples we studied.

ACKNOWLEDGMENTS

We thank S. Schramm and M. Grabiak for stimulating and useful discussions.

[1]A. F. Ranada, in *Quantum Theory Groups—Fields and Particles*, edited by A. O. Barut (Reidel, Dordrecht, 1983).

[2]D. C. Ionescu, J. Reinhardt, B. Müller, W. Greiner, and G. Soff, J. Phys. G, Nucl. Phys. **14**, L 109 (1988).

[3]J. Rafelski and B. Müller, Phys. Rev. D **14**, 3532 (1976).

[4]U. Ascher, J. Christiansen, and R. D. Russel, Math. Comp. **33**, 659 (1979).

[5]D. C. Ionescu, J. Reinhardt, B. Müller, W. Greiner, and G. Soff, Phys. Rev. A **38**, 616 (1988).

[6]M. Soler, Phys. Rev. D **1**, 2766 (1970).

[7]A. Alvarez and M. Soler, Phys. Rev. Lett. **50**, 1230 (1983).

[8]W. A. Strauss and L. Vazquez, Phys. Rev. D **32**, 641 (1986).

ERRATUM

Erratum: Expressions for the zeta-function regularized Casimir energy [J. Math. Phys. 30, 1133 (1989)]

E. Elizalde and A. Romeo
Department of Structure and Constituents of Matter, Faculty of Physics, University of Barcelona, Diagonal 647, 08028 Barcelona, Spain

(Received 26 September 1989; accepted for publication 18 October 1989)

Formula (1.1) must be

$$H=\frac{1}{2}\sum_k \omega_k(a_k^\dagger a_k + a_k a_k^\dagger) = \sum_k \omega_k\left(n_k+\frac{1}{2}\right).$$

Expressions (2.8) and (2.9) ought to be written as

$$E(d,a,m) = -\frac{1}{2}\frac{L^{d-1}}{a^d}\pi^{-(d+1)/2}\left\{-\sqrt{\pi}\Gamma\left(-\frac{d}{2}\right)\left(\frac{am}{2}\right)^d + \left(\frac{am}{2}\right)^{d+1}\left[\Gamma\left(-\frac{d+1}{2}\right) + 4\sum_{n=1}^{\infty}\frac{K_{(d+1)/2}(amn)}{(amn/2)^{(d+1)/2}}\right]\right\}$$

and

$$\epsilon(d,a,m) = -2\frac{L^{d-1}}{a^d}\pi^{-(d+1)/2}\left(\frac{am}{2}\right)^{d+1} \times \sum_{n=1}^{\infty}\frac{K_{(d+1)/2}(amn)}{(amn/2)^{(d+1)/2}} \equiv -2\frac{L^{d-1}}{a^d}\pi^{-(d+1)/2}S(d,a,m),$$

respectively.

The equalities (B2) should read

$$S_B^{(\alpha)}[f,s] = \begin{cases} \displaystyle\sum_{a=0}^{\infty}\xi(s+1-\alpha a)f(a) - f\left(\frac{s}{\alpha}\right)\frac{\pi}{\alpha}\cot\frac{\pi s}{\alpha} - \Delta_B^{(\alpha)}[f,s], & \displaystyle\frac{s}{\alpha}\notin\mathbf{N}, \\ \displaystyle\sum_{\substack{a=0\\ a\neq s/\alpha}}^{\infty}\zeta(s+1-\alpha a)f(a) + \gamma f\left(\frac{s}{\alpha}\right) - \frac{1}{\alpha}f'\left(\frac{s}{\alpha}\right) - \Delta_B^{(\alpha)}[f,s], & \displaystyle\frac{s}{\alpha}\in\mathbf{N}. \end{cases}$$

