Deformations of the Galilean algebra

José M. Figueroa-O'Farrill
Institute for Theoretical Physics, State University of New York at Stony Brook, Stony Brook, New York 11794-3840

(Received 17 March 1989; accepted for publication 9 August 1989)

All the infinitesimal deformations of the Galilean algebra with and without central extension are computed, as well as their integrability properties. Among the four-parameter family of infinitesimal deformations of the unextended algebra is found the Newton algebras, the Euclidean algebra E(4), the Poincaré algebra, the de Sitter algebras, and SO(5). For the centrally extended algebra there is found, in particular, an infinitesimal deformation containing a Poincaré subalgebra (although the embedding is not the natural one), and centrally extended versions of the Newton algebras.

I. INTRODUCTION

The theory of Lie algebra deformations[1-3] provides us with a systematic procedure that is an inverse to the more common Lie algebra contractions.[4,5] In practice, contractions are associated with physical parameters, which enter in the commutation relations of a representation of a Lie algebra, tending to some physically meaningful limit. For example, the low velocity limit ($c \to \infty$) of the Poincaré algebra yields[6,7] the Galilean algebra, and the Poincaré algebra is itself the flat space limit ($\kappa \to 0$) of the de Sitter algebras. It is more useful, however, to look at the inverse problem. Suppose that we are given a Lie algebra which, to the best of our empirical knowledge, is an exact symmetry of the physical system under consideration. It may be, however, that this symmetry is only approximate and hence—either to discover new symmetry principles or to suggest empirical tests which can probe the exactness of the symmetry—one would like to know all the possible algebras that are "close" in some sense to the given one. These are precisely those algebras to which the given algebra can be deformed. The advantage of the deformation approach lies in that deformations can be searched for systematically by computing Lie algebra cohomology groups. Equivalently, one could classify all isomorphism classes of Lie algebras of a given dimension and then compute all the possible contractions, but as the dimension grows, this problem becomes computationally untractable, whereas the computation of cohomology groups is still feasible, due in great part to theorems like the one of Hochschild and Serre,[8] which allows one to exploit the semisimple part of the algebra in question to simplify the calculations tremendously.

The possible deformations of the Poincaré algebra have been known for some time. It was proven by Levy-Nahas in Ref. 3 that the only algebras to which the Poincaré algebra can be deformed are the de Sitter algebras SO(4,1) and SO(3,2). It had been proven previously by Sharp in Ref. 9 that this was the case among the semisimple Lie algebras.

In this paper, we analyze the question starting from the Galilean algebra. Since both the Galilean algebra as well as its centrally extended version are physically interesting, we find all the Lie algebras to which these algebras can be deformed. Whereas our results for the centrally extended algebra are new, the ones for the unextended case can be read from the work of Bacry and Lévy-Leblond[7] and of Bacry and Nuyts[10] who classify all the (3 + 1 dimensional) kinematic Lie algebras under the constraint of space isotropy. Kinematic Lie algebras are those real Lie algebras generated by the ten elements $\{M_{ij},K_i,P_i,P_0\}$ where i,j run from 1 to 3 and $M_{ij} = \frac{1}{2}\epsilon_{ijk}g^{kl}J_l$, and the constraint of space isotropy merely fixes the transformation laws of the generators under space rotations; K_i, J_i, P_i are vectors and P_0 is a scalar. Although we make no isotropy assumption on the deformations, we notice that as a consequence of the semisimplicity of the rotation subalgebra (and hence its rigidity under deformations) there are no deformations of the Galilean algebra that are not isotropic.

In summary, we find that there are four infinitesimal deformations of the unextended Galilean algebra yielding, among many other algebras, the Newton algebras, the Euclidean algebra E(4), the Poincaré algebra, as well as diverse real forms of B_2: SO(3,2), SO(4,1), and SO(5). For the centrally extended algebra there are three infinitesimal deformations, one of which corresponds to centrally extended versions of the Newton algebras, and one of which contains a Poincaré subalgebra, although the embedding is not the natural one. We also investigate the integrability properties of these infinitesimal deformations.

This paper is organized as follows. In Sec. II we review the basic facts about deformations of Lie algebras. Lack of space prohibits a more detailed account, but the reader is urged to look at the beautiful treatment to be found in Ref. 2. In Sec. III we discuss the factorization theorem of Hochschild and Serre, which simplifies many of the calculations. We also give two brief applications of this theorem: the determination of the possible central extensions of the Galilean algebra, and the theorem of Levy-Nahas on the deformations of the Poincaré algebra. Finally, in Sec. IV we determine the infinitesimal deformations of the Galilean algebra with and without central extension and discuss their integrability domains.

II. DEFORMATIONS OF LIE ALGEBRAS

Let $\mathfrak{g}$ be a finite-dimensional real Lie algebra and $\mathfrak{g}[[t]]$ the space of formal power series in t with coefficients in $\mathfrak{g}$. By a deformation of we $\mathfrak{g}$ we mean a skew-symmetric bilinear

 0022-2488/89/122735-05$02.50

map $\mathfrak{g}\times\mathfrak{g}\to \mathfrak{g}[[t]]$ that satisfies the Jacobi identity formally[11] order by order in t. In other words, we can think of a deformation as a new bracket $[\ ,\ ]_t$ defined by

$$[X,Y]_t = [X,Y] + \sum_{n=0}^{\infty} t^n C_n(X,Y), \tag{2.1}$$

for all X, $Y\in \mathfrak{g}$ and where the C_n are cochains in $C^2(\mathfrak{g};\mathfrak{g})$. The Jacobi identity imposes certain conditions on these cochains. In particular, the linear therm C_1 has to be a cocycle. Conversely, if C_1 is any cocycle in $C^2(\mathfrak{g};\mathfrak{g})$, we can begin to define a deformation by

$$[X,Y]_t = [X,Y] + tC_1(X,Y). \tag{2.2}$$

The fact that C_1 is a cocycle guarantees that the Jacobi identity is satisfied up to terms of order t^2. Therefore we call C_1 an *infinitesimal deformation*. In general, not every infinitesimal deformation is the linear term of a deformation. Those that are, are called *integrable*. To see what stands in the way of an infinitesimal deformation giving rise to a deformation, we look at the terms of order t^2 in the Jacobi identity for the above infinitesimal deformation. One finds the following:

$$C_1(X, C_1(Y,Z)) + C_1(Y, C_1(Z,X)) + C_1(Z, C_1(X,Y)), \tag{2.3}$$

which, since C_1 is a cocycle, can be seen to be a cocycle in $C^3(\mathfrak{g};\mathfrak{g})$. If and only if it is also a coboundary, say $-dC_2$, we can continue the deformation as follows:

$$[X,Y]_t = [X,Y] + t\,C_1(X,Y) + t^2\,C_2(X,Y), \tag{2.4}$$

guaranteeing that the Jacobi identity is satisfied up to term of order t^3 and higher. Looking at the terms of order t^3, we again find a cocycle in $C^3(\mathfrak{g};\mathfrak{g})$ and so on. Hence the obstruction to the integrability of an infinitesimal deformation is an infinite sequence of cocycles in $C^3(\mathfrak{g};\mathfrak{g})$ whose cohomology classes all have to vanish. These classes appear very naturally within the framework of Nijenhuis and Richardson,[2] who define the structure of a graded Lie algebra on the cohomology $H(\mathfrak{g};\mathfrak{g})$. We refer the reader to their paper for the details.

On the other hand, not all infinitesimal deformations are "essential." Suppose that C_1 is a coboundary. That is, $C_1 = -dB_1$, for some B_1 in $C^1(\mathfrak{g};\mathfrak{g})$. Then we define the map T by $T(X) = X + tB_1(X)$. It is then trivial to verify that up to terms of order t^2,

$$[T(X), T(Y)]_t = T([X,Y]). \tag{2.5}$$

Conversely such a map T exists only if C_1 is a coboundary. Hence infinitesimal deformations such that C_1 is a coboundary will be called *trivial*. This allows us to introduce an equivalence relation on the set of infinitesimal deformations. Two infinitesimal deformations are considered equivalent if their difference is trivial. Hence the equivalence classes of infinitesimal deformations are in bijective correspondence with the cohomology group $H^2(\mathfrak{g};\mathfrak{g})$.

Hence we see that there are two crucial cohomology groups in the theory of deformations of Lie algebras: $H^2(\mathfrak{g};\mathfrak{g})$, which contains the nontrivial infinitesimal deformations, and $H^3(\mathfrak{g};\mathfrak{g})$, which contains the obstructions to the integrability of the infinitesimal deformations. In general, unless one can show that it vanishes, it is neither necessary nor useful to compute $H^3(\mathfrak{g};\mathfrak{g})$ since only certain classes have to be checked. We do, however, need to compute $H^2(\mathfrak{g};\mathfrak{g})$, and in the next section we will describe a method attributable to Hochschild and Serre[8] that makes the computations rather straightforward.

Notice that a semisimple algebra is *rigid* in the sense that it admits no nontrivial deformations.

III. THE FACTORIZATION THEOREM OF HOCHSCHILD AND SERRE

In Ref. 8 Hochschild and Serre proved a factorization theorem that in many cases simplifies the calculation of Lie algebra cohomology groups. Leg $\mathfrak{g}$ be a finite-dimensional real Lie algebra and $\mathfrak{h}$ be an ideal such that the quotient Lie algebra $\mathfrak{s} = \mathfrak{g}/\mathfrak{h}$ is a semisimple. Let $\mathfrak{m}$ denote a $\mathfrak{g}$ module. Then the ideal $\mathfrak{h}$ defines a filtration of the cochains $C(\mathfrak{g};\mathfrak{m})$ whose spectral sequence degenerates at the E_2 term yielding the following isomorphism:

$$H^n(\mathfrak{g};\mathfrak{m}) \cong \bigoplus_{i=0}^{n} H^{n-i}(\mathfrak{s};\mathbb{R}) \otimes H^i(\mathfrak{h};\mathfrak{m})^{\mathfrak{s}}, \tag{3.1}$$

where $^{\mathfrak{s}}$ denotes $\mathfrak{s}$ invariants. Since $\mathfrak{s}$ is semisimple, it acts reducibly on the cochains $C(\mathfrak{h};\mathfrak{m})$, and hence the invariant cohomology can be computed from the invariant cochains.

Moreover, using the Whitehead lemmas, we know that $H^1(\mathfrak{s};\mathbb{R}) = H^2(\mathfrak{s};\mathbb{R}) = 0$. If, in addition, $\mathfrak{s}$ is simple, then $H^3(\mathfrak{s};\mathbb{R}) \cong \mathbb{R}$. Hence for $\mathfrak{s}$ simple, the first few $H(\mathfrak{g};\mathfrak{g})$'s are as follows:

$$H^0(\mathfrak{g};\mathfrak{g}) \cong Z(\mathfrak{g}),$$
$$H^1(\mathfrak{g};\mathfrak{g}) \cong H^1(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}},$$
$$H^2(\mathfrak{g};\mathfrak{g}) \cong H^2(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}},$$
$$H^3(\mathfrak{g};\mathfrak{g}) \cong H^3(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}} \oplus Z(g),$$

where $Z(\mathfrak{g})$ denotes the center of $\mathfrak{g}$.

A. Central extensions of the Galilean algebra

As a trivial application of the factorization theorem, we compute the central extensions of the Gililean algebra; that is, $H^2(\mathfrak{g};\mathbb{R})$. The Galilean algebra is a real Lie algebra generated by $\{M_{ij}, K_i, P_i, P_0\}$ where i, j run from 1 to 3 and $M_{ij} = -M_{ji}$. The Lie bracket is given by

$$[M_{ij}, M_{kl}] = g_{jk}M_{il} + g_{il}M_{jk} - g_{ik}M_{jl} - g_{il}M_{ik},$$
$$[M_{ij}, K_k] = g_{jk}K_i - g_{ik}K_j,$$
$$[M_{ij}, P_k] = g_{jk}P_i - g_{ik}P_j,$$
$$[K_i, P_0] = P_i,$$

and all other brackets are zero. Let $\mathfrak{h}$ denote the ideal generated by $\{K_i, P_i, P_0\}$. Then $\mathfrak{s}$ is the subalgebra generated by $\{M_{ij}\}$. By the factorization theorem, $H^2(\mathfrak{g};\mathbb{R}) \cong H^2(\mathfrak{h};\mathbb{R})^{\mathfrak{s}}$. The space of $\mathfrak{s}$ invariant two-cochains is one-dimensional spanned by $K^{i*} \wedge P^{j*} g_{ij}$, where K^{i*} is the canonical dual vector to K_i and the same for P^{i*}. It is clearly a cocycle and not a coboundary since the only $\mathfrak{s}$ invariant one-cochain is P^{0*}, which is a cocycle. Therefore there is only one nontrivial central extension. Let us denote its generator by c. Then the extended Galilean algebra is supplemented with the extra term $[K_i, P_j] = g_{ij}c$, as is well known.

B. Deformations of the Poincaré algebra

As a final trivial application of the factorization theorem, we determine the deformations of the Poincaré algebra. This was first done by Levy-Nahas in Ref. 3. The Poincaré algebra is spanned by $\{M_{ab}, P_a\}$, where a, b run from 1 to 4 and $M_{ab} = -M_{ba}$. The Lie bracket is given by

$$[M_{ab}, M_{cd}] = g_{bc}M_{ad} + g_{ad}M_{bc} - g_{ac}M_{bd} - g_{bd}M_{ac},$$
$$[M_{ab}, P_c] = g_{bc}P_a - g_{ac}P_b,$$

all other brackets being zero. From this it already follows that $Z(\mathfrak{g}) = 0$. We choose $\mathfrak{h}$ to be a translation ideal. Then $\mathfrak{s}$ can be identified with the Lorentz subalgebra. There are no Lorentz invariant elements in $\mathfrak{h}$. The only Lorentz invariant cochain in $C^1(\mathfrak{h}; \mathfrak{g})$ is $P^{a*} \otimes P_a$ which is clearly a cocycle but not a coboundary by the previous remark. There are only two linearly independent Lorentz invariant cochains in $C^2(\mathfrak{h}; \mathfrak{g})$. They are $P^{a*} \wedge P^{b*} \otimes M_{ab}$ and $P^{a*} \wedge P^{b*} \otimes \epsilon_{abcd} g^{ce} M_{ef}$. The first one is a cocycle but the second one is not. Therefore there is a unique nontrivial infinitesimal deformation of the Poincaré algebra. There is only one linearly independent Lorentz invariant cochain in $C^3(\mathfrak{h}; \mathfrak{g})$. It is $P^{a*} \wedge P^{b*} \wedge P^{c*} \oplus \epsilon_{abcd} g^{de} P_e$. It is not just a cocycle but also a coboundary. Hence there are no obstructions to integrability and the unique nontrivial infinitesimal deformation is integrable. It turns out that the deformed algebra needs no terms of order t^2 since the obstruction cocycle vanishes identically, and hence we are left with the deformed algebra

$$[M_{ab}, M_{cd}] = g_{bc}M_{ad} + g_{ad}M_{bc} - g_{ac}M_{bd} - g_{bd}M_{ac},$$
$$[M_{ab}, P_c] = g_{bc}P_a - g_{ac}P_b,$$
$$[P_a, P_b] = tM_{ab}.$$

Notice that by rescaling P_a we can always reduce t down to a sign, but without complexifying we cannot reabsorb the sign. These algebras correspond to the de Sitter and anti-de Sitter Lie algebras depending on the sign of t. Both of these algebras are simple and hence rigid, admitting no further deformations.

IV. DEFORMATIONS OF THE GALILEAN ALGEBRA

In this section, we look at the deformations of the Galilean algebra with and without central extension.

TABLE I. Results of cohomology calculations for the unextended Galilean algebra.

Space	Dimension
$C^0(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	1
$Z^0(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	0
$H^0(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	0
$B^1(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	1
$C^1(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	7
$Z^1(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	3
$H^1(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	2
$B^2(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	4
$C^2(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	16
$Z^2(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	8
$H^2(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	4
$B^3(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	8

A. Unextended Galilean algebra

The factorization theorem tells us that in order to compute $H^2(\mathfrak{g};\mathfrak{g})$ we merely need to compute $H^2(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$, The method is straightforward. We first isolate the $\mathfrak{s}$-invariant cochains and then determine which of these are cocycles and coboundaries. This then gives us the dimension of the cohomology group, as well as representative cocycles from each class. We let $Z(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$ and $B(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$ determine the $\mathfrak{s}$-invariant cocycles and coboundaries, respectively; Table I summarizes the results.

In particular, there is a four-parameter family of nontrivial infinitesimal deformations generated by the following cocycles:

$$C_1^1 = K^{i*} \wedge K^{j*} \otimes \epsilon_{ijk} g^{kl} P_l,$$
$$C_1^2 = \tfrac{1}{2} K^{i*} \wedge K^{j*} \otimes M_{ij} - P^{i*} \wedge K^{j*} \otimes g_{ij} P_0,$$
$$C_1^3 = P^{0*} \wedge P^{i*} \otimes K_i,$$
$$C_1^4 = P^{0*} \wedge P^{i*} \otimes P_i + P^{0*} \wedge K^{i*} \otimes K_i.$$

The most general nontrivial infinitesimal deformation is therefore a linear combination $\Sigma_{a=1}^4 t_a C_1^a$. To investigate the integrability of the infinitesimal deformations, we must first compute the obstruction cocycles and then determine which ones are coboundaries. The first obstruction is the three-cocycle $J_2 = \Sigma_{a,b=1}^4 t_a t_b J_2^{ab}$ where

$$J_2^{ab}(X,Y,Z) = C_1^a(X, C_1^b(Y,Z)) + \text{cyclic permutations.} \tag{4.1}$$

A straightforward calculation yields

$$\begin{aligned} J_2 = {} & t_1 t_3 (\tfrac{1}{2} P^{0*} \wedge K^{i*} \wedge K^{j*} \otimes \epsilon_{ijk} g^{kl} K_l - P^{0*} \wedge K^{i*} \wedge P^{j*} \otimes \epsilon_{ijk} g^{kl} P_l) + t_1 t_2 K^{i*} \wedge K^{j*} \wedge K^{k*} \otimes \epsilon_{ijk} P_0 \\ & - \tfrac{1}{2} t_1 t_4 P^{0*} \wedge K^{i*} \wedge K^{j*} \otimes \epsilon_{ijk} g^{kl} P_l + t_2 t_3 (P^{i*} \wedge P^{j*} \wedge K^{k*} \otimes g_{jk} K_i - P^{0*} \wedge P^{i*} \wedge K^{j*} \otimes M_{ij}) \\ & + t_2 t_4 (2P^{0*} d \wedge P^{i*} \wedge K^{j*} \otimes g_{ij} P_0 - P^{0*} \wedge K^{i*} \wedge K^{j*} \otimes M_{ij} + P^{i*} \wedge P^{j*} \wedge K^{k*} \otimes g_{jk} P_i + P^{i*} \wedge K^{j*} K^{k*} \otimes g_{ij} K_k). \end{aligned} \tag{4.2}$$

All terms, except for those proportional to $t_1 t_2$ and $t_2 t_4$, are coboundaries. Hence this infinitesimal deformation is not integrable unless $t_1 = t_4 = 0$ or $t_2 = 0$.

In the first case, $t_1 = t_4 = 0$, we have that $J_2 = -dC_2$ where

$$C_2 = \tfrac{1}{2} t_2 t_3 P^{i*} \wedge P^{j*} \otimes M_{ij}. \tag{4.3}$$

Computing the obstruction to the integrability of this second-order infinitesimal deformation, we find that $J_3 = 0$ and hence this is already a deformation. Reabsorbing the deformation parameter t into the t_a's, we have the following deformed algebra:

$$[M_{ij}, M_{kl}] = g_{jk} M_{il} + g_{il} M_{jk} - g_{ik} M_{jl} - g_{jl} M_{ik},$$

$$
\begin{aligned}
&[M_{ij},K_k] = g_{jk}K_i - g_{ik}K_j,\\
&[M_{ij},P_k] = g_{jk}P_i - g_{ik}P_j,\\
&[K_i,P_0] = P_i,\\
&[K_i,K_j] = t_2M_{ij},\\
&[P_0,P_i] = t_3K_i,\\
&[P_i,P_j] = t_2t_3M_{ij},\\
&[P_i,K_j] = -t_2g_{ij}P_0,
\end{aligned}
$$

for any value of t_2 and t_3. For $t_3 = 0$ we obtain either the Euclidean algebra E(4) or the Poincaré algebra depending on the sign of t_2, which via rescaling can be reduced to a sign itself. For t_2 and t_3 both nonzero we get, depending on their sign, SO(5), SO(4,1), or SO(3,2). The correspondence is the usual one: $K_i \to M_{0i}$, $P_i \to M_{5i}$, and $P_0 \to M_{50}$. Then substituting these into the commutation relations, we see that after some rescaling we can identify t_2 with $-g_{00}$ and t_2t_3 with $-g_{55}$. Finally, when $t_2 = 0$ we obtain the Newton algebras which, depending on the sign of t_3, we call N_+ and N_-. For $t_3 > 0$ and after some rescaling we get the algebra N_+ which is defined by

$$
\begin{aligned}
&[M_{ij},M_{kl}] = g_{jk}M_{ik} + g_{il}M_{jk} - g_{ik}M_{jl} - g_{jl}M_{ik},\\
&[M_{ij},K_k] = g_{jk}K_i - g_{ik}K_j,\\
&[M_{ij},P_k] = g_{jk}P_i - g_{ik}P_j,\\
&[P_0,K_i] = -P_i,\\
&[P_0,P_i] = K_i.
\end{aligned}
$$

For $t_3 < 0$ after some rescaling and rotating P_i and K_i we obtain the algebra N_- defined by

$$
\begin{aligned}
&[M_{ij},M_{kl}] = g_{jk}M_{il} + g_{il}M_{jk} - g_{ik}M_{jl} - g_{jl}M_{ik},\\
&[M_{ij},K_k] = g_{jk}K_i - g_{ik}K_j,\\
&[M_{ij},P_k] = g_{jk}P_i - g_{ik}P_j,\\
&[P_0,K_i] = K_i,\\
&[P_0,P_i] = -P_i.
\end{aligned}
$$

These results are summarized graphically in Fig. 1, which represents the $t_1 = t_4 = 0$ plane in the parameter space of infinitesimal deformations.

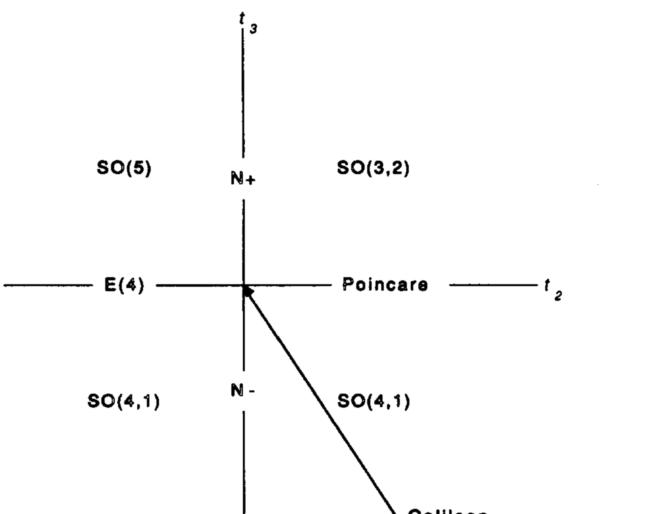


FIG. 1. The $t_1 = t_4 = 0$ plane in the parameter space of infinitesimal deformations of the unextended Galilean algebra.

In the second case, $t_2 = 0$, we have that J_2 is a coboundary, but the obstruction cocycle at level 3 to which this second-order deformation leads is not integrable unless $t_1 = 0$ or $t_3 = 0$. In the first case, $t_1 = 0$, we see that J_2 is automatically zero so that this is already a deformation. The deformed algebra is given by

$$
\begin{aligned}
&[M_{ij},M_{kl}] = g_{jk}M_{il} + g_{il}M_{jk} - g_{ik}M_{jl} - g_{jl}M_{ik},\\
&[M_{ij},K_k] = g_{jk}K_i - g_{ik}K_j,\\
&[M_{ij},P_k] = g_{jk}P_i - g_{ik}P_j,\\
&[P_0,K_i] = -P_i + t_4K_i,\\
&[P_0,P_i] = t_4P_I + t_3K_i.
\end{aligned}
$$

This is nonsemisimple Lie algebra that does not seem particularly interesting. In the second case, $t_3 = 0$, a long calculation yields an obstruction cocycle at level 5 that is not a coboundary unless $t_1 = 0$ or $t_4 = 0$. In any of these cases $J_2 = 0$ to begin with and we already have deformations. The case $t_1 = 0$ yields (after some rescaling) the algebra

$$
\begin{aligned}
&[M_{ij},M_{kl}] = g_{jk}M_{ik} + g_{il}M_{jk} - g_{ik}M_{jl} - g_{jl}M_{ik},\\
&[M_{ij},K_k] = g_{jk}K_i - g_{ik}K_j,\\
&[M_{ij},P_k] = g_{jk}P_i - g_{ik}P_j,\\
&[P_0K_i] = -P_i + K_i,\\
&[P_0,P_i] = P_i,
\end{aligned}
$$

which is a contraction of the previously found algebra; whereas in the second case, $t_4 = 0$, we find, after some rescaling, the deformed algebra

$$
\begin{aligned}
&[M_{ij},M_{kl}] = g_{jk}M_{il} + g_{il}M_{jk} - g_{ik}M_{jk} - g_{jl}M_{ik},\\
&[M_{ij},K_k] = g_{jk}K_i - g_{ik}K_j,\\
&[M_{ij},P_k] = g_{jk}P_i - g_{ik}P_j,\\
&[K_i,K_j] = \epsilon_{ijk}g^{kl}P_l,\\
&[P_0,K_i] = -P_i.
\end{aligned}
$$

In brief, we have a four parameter family $\{t_a\}$ of nontrivial infinitesimal deformations. The "domain of integrability" of these infinitesimal deformations, i.e., the subset of $\mathbb{R}^4$ corresponding to those values of $\{t_a\}$ for which the infinitesimal deformations are integrable, is given by the (3,4)-plane, the (2,3)-plane, and the 1-axis. The interesting deformations seem to be in the (2,3)-plane, as Fig. 1 shows.

B. Centrally extended Galilean algebra

The summary of the calculations for the centrally extended Galilean algebra is shown in Table II.

In particular, there is a three-parameter family of nontrivial infinitesimal deformations induced by the following cocycles:

$$
\begin{aligned}
C_1^1 &= P^{0*} \wedge P^{i*} \otimes K_i,\\
C_1^2 &= P^{0*} \wedge P^{i*} \otimes P_i + P^{0*} \wedge K^{i*} \otimes K_i - 2c^* \wedge P^{0*} \otimes c,\\
C_1^3 &= \tfrac{1}{2}K^{i*} \wedge K^{j*} \otimes M_{ij} + c^* \wedge K^{i*} \otimes P_i.
\end{aligned}
$$

Again the most general nontrivial infinitesimal deformation is $\Sigma_{a=1}^3 t_a C_1^a$. The first obstruction cocycle is given by

TABLE II. Results of the cohomology calculations for the extended Galilean algebra.

Space	Dimension
$C^0(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	2
$Z^0(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	1
$H^0(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	1
$B^1(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	1
$C^1(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	10
$Z^1(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	4
$H^1(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	3
$B^2(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	6
$C^2(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	25
$Z^2(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	9
$H^2(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	3
$B^3(\mathfrak{h};\mathfrak{g})^{\mathfrak{s}}$	16

$$J_2 = t_1 t_3 (P^{0*} \wedge c^* \wedge K^{i*} \otimes K_i - P^{0*} \wedge c^* \wedge P^{i*} \otimes P_i) + 2 t_2 t_3 P^{0*} \wedge c^* \wedge K^{i*} \otimes P_i. \tag{4.4}$$

This is a coboundary if and only if it vanishes identically. That is, $t_1 = t_2 = 0$ or $t_3 = 0$. In this latter case the deformed algebra looks like

$$\begin{aligned}
&[M_{ij},M_{kl}] = g_{jk}M_{il} + g_{il}M_{jk} - g_{ik}M_{jl} - g_{jl}M_{ik},\\
&[M_{ij},K_k] = g_{jk}K_i - g_{ik}K_j,\\
&[M_{ij},P_k] = g_{jk}P_i - g_{ik}P_j,\\
&[P_0,K_i] = -P_i + t_2 K_i,\\
&[P_0,P_i] = t_1 K_i + t_2 P_i,\\
&[P_i,K_j] = g_{ij}c,\\
&[c,P_0] = -t_2 c.
\end{aligned}$$

If $t_2 = 0$, and depending on the sign of t_1, we get centrally extended versions of the Newton algebras N_+ and N_- considered in the previous subsection. If $t_1 = 0$ we get, after some rescaling,

$$\begin{aligned}
&[M_{ij},M_{kl}] = g_{jk}M_{il} + g_{il}M_{jk} - g_{ik}M_{jl} - g_{jl}M_{ik},\\
&[M_{ij},K_k] = g_{jk}K_i - g_{ik}K_j,\\
&[M_{ij},P_k] = g_{jk}P_i - g_{ik}P_j,\\
&[P_0,K_i] = -P_i + K_i,\\
&[P_0,P_i] = P_i,\\
&[P_i,K_j] = g_{ij}c,\\
&[c,P_0] = -2c.
\end{aligned}$$

In the case that $t_1 = t_2 = 0$ we get the following deformed algebra:

$$\begin{aligned}
&[M_{ij},M_{kl}] = g_{jk}M_{il} + g_{il}M_{jk} - g_{ik}M_{jl} - g_{jl}M_{ik},\\
&[M_{ij},K_k] = g_{jk}K_i - g_{ik}K_j,\\
&[M_{ij},P_k] = g_{jk}P_i ! g_{ik}P_j,\\
&[K_i,P_0] = P_i,\\
&[K_i,K_j] = t_3 M_{ij},\\
&[c,K_i] = t_3 P_i,\\
&[P_i,K_j] = g_{ij}c.
\end{aligned}$$

Notice that depending on the sign of t_3, this algebra has a Poincaré of $E(4)$ subalgebra. However, this seems to be an accident since the role of the fourth momentum generator is played not by P_0 but by the central extension c. In fact, we get either Poincaré or $E(4)$ with an extra generator which we call D obeying $[D,M_{0i}] = -P_i$.

In brief, we have a three-parameter family of nontrivial infinitesimal deformations, whose integrability domain is the subset of $\mathbb{R}^3$ consisting of the (1,2)-plane and the three-axis.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge conversations with Francisco Figueirido, Takashi Kimura, and Eduardo Ramos. I am also indebted to the referee for making me aware of Refs. 7 and 10.

[1] M. Gerstenhaber, Ann. Math. 79, 59 (1964).
[2] A. Nijenhuis and R. W. Richardson, J. Math. Mech. 17, 89 (1967).
[3] M. Levy-Nahas, J. Math. Phys. 8, 1211 (1967).
[4] E. Inönu and E. P. Wigner, Proc. Natl. Acad. Sci. USA 39, 510 (1953).
[5] E. J. Saletan, J. Math. Phys. 2, 1 (1964).
[6] More precisely, as it is shown in Ref. 7, the limit $c \to \infty$ of the Poincaré algebra yields either the Galilean algebra (rescaling the boosts and the space translations) or the Carroll algebra (rescaling the boosts and the time translations).
[7] H. Bacry and J.-M. Lévy-Leblond, J. Math. Phys. 9, 1605 (1968).
[8] G. Hochschild and J.-P. Serre, Ann. Math. 57, 591 (1953).
[9] W. T. Sharp, Ph.D. thesis, Princeton University, 1960.
[10] H. Bacry and J. Nuyts, J. Math. Phys. 27, 2455 (1986).
[11] Notice that we do not impose any convergence properties on the series. We are therefore dealing with formal deformations. It will turn out, however, that the deformations found in this paper are all polynomial and, therefore, trivially convergent.

Multivariable Meixner, Krawtchouk, and Meixner–Pollaczek polynomials

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

(Received 17 May 1988; accepted for publication 28 June 1989)

A multivariable biorthogonal generalization of the Meixner, Krawtchouk, and Meixner–Pollaczek polynomials is presented. It is shown that these are orthogonal with respect to subspaces of lower degree and biorthogonal within a given subspace. The weight function associated with the Krawtchouk polynomials is the multivariate binomial distribution.

I. INTRODUCTION

The Meixner polynomials[1-3] of a single variable appear in the Askey tableau[4] of hypergeometric orthogonal polynomials under family (j) and are defined by the following hypergeometric representation:

$$m_n(x;\beta,c) = (\beta)_n\, {}_2F_1(-n,-x;\beta;1-c^{-1}),$$
$$(\beta)_n \equiv \Gamma(n+\beta)/\Gamma(\beta), \tag{1.1}$$

where β and c are complex parameters, with $0<|c|<1$ and n a non-negative integer. These can also be expressed as a Rodrigues-type formula,

$$m_n(x;\beta,c) = \frac{\Gamma(x+1)}{\Gamma(x+\beta)}\, c^{-n-x} D^n\left[\frac{c^x\Gamma(x+\beta)}{\Gamma(x-n+1)}\right], \tag{1.2}$$

where D is a finite difference operator defined as

$$Df(x) \equiv f(x+1) - f(x). \tag{1.3}$$

They satisfy a discrete orthogonality relation,

$$\sum_{x=0}^{\infty} m_n(x) m_{n'}(x) w(x) = (\beta)_n n! c^{-n}(1-c)^{-\beta}\delta_{nn'}, \tag{1.4}$$

where the weight function is given by

$$w(x) = \frac{c^x}{x!}\frac{\Gamma(x+\beta)}{\Gamma(\beta)}, \tag{1.5}$$

and the sum is over all non-negative integers $x = 0,1,2,\ldots,\infty$.

A special case of the Meixner polynomials is associated with the binomial distribution; these were introduced earlier by Krawtchouk[5] and appear as family (i) in the Askey tableau. They are defined

$$k_n(x) = (q^n/n!)\, m_n(x;-\Delta,-q/(1-q)),$$
$$n = 0,1,2,\ldots,\Delta, \tag{1.6}$$

where Δ is a non-negative integer and q is a real parameter in the range $0<q<1$. The corresponding Rodrigues formula is

$$k_n(x) = \frac{(-1)^n}{n!}\frac{\Gamma(x+1)\Gamma(\Delta-x+1)}{q^x(1-q)^{\Delta-x}} \times D^n\left[\frac{q^x(1-q)^{\Delta-x+n}}{\Gamma(x-n+1)\Gamma(\Delta-x+1)}\right], \tag{1.7}$$

while the orthogonality relation is written as

$$\sum_{x=0}^{\Delta} k_n(x) k_m(x) w(x) = \binom{\Delta}{n}(1-q)^n q^n \delta_{nm}, \quad n,m = 0,1,2,\ldots,\Delta, \tag{1.8}$$

where the weight function is the binomial distribution

$$w(x) = \binom{\Delta}{x} q^x(1-q)^{\Delta-x}, \tag{1.9}$$

and the sum is over the finite range of non-negative integers $x = 0,1,2,\ldots,\Delta$.

As shown in the Askey tableau, a limit case of the Krawtchouk polynomials are the so-called Charlier[2,3] polynomials designated as family (l). If we put $q = a/\Delta$, then for fixed a, x, and n,

$$\lim_{\Delta\to\infty} k_n(x) = (1/n!)\, C_n^{(a)}(x), \tag{1.10}$$

where the Charlier polynomials are defined

$$C_n^{(a)}(x) = \sum_{j=0}^{n}\binom{n}{j}\binom{x}{j} j!(-a)^{n-j}, \tag{1.11}$$

and these satisfy the orthogonality relation

$$\sum_{x=0}^{\infty} C_n^{(a)}(x) C_m^{(a)}(x)\frac{a^x}{x!}\exp\{-a\} = a^n n! \delta_{nm}. \tag{1.12}$$

Under (g) in the Askey tableau are found the Meixner–Pollaczek polynomials which satisfy a continuous orthogonality relation. The hypergeometric representation is

$$P_n^\lambda(x;\phi) = (\Gamma(n+2\lambda)/n!\Gamma(2\lambda))\exp\{in\phi\} \times {}_2F_1(-n,\lambda+ix;2\lambda;1-\exp\{-2i\phi\}), \tag{1.13}$$

where λ and ϕ are real parameters with $0<\phi<\pi$ and $\lambda>0$. They also have a Rodrigues-type representation,

$$P_n^\lambda(x;\phi) = \frac{(-1)^n}{n!}\frac{\Gamma(1-\lambda+ix)}{\Gamma(\lambda+ix)}\exp\{-2\phi x\}\delta^n \times\left[\frac{\Gamma(\lambda+\frac{1}{2}n+ix)}{\Gamma(1-\lambda-\frac{1}{2}n+ix)}\exp\{2\phi x\}\right], \tag{1.14}$$

where δ is another finite difference operator defined as

$$\delta f(x) \equiv f(x+\tfrac{1}{2}i) - f(x-\tfrac{1}{2}i). \tag{1.15}$$

These are orthogonal on the infinite real line

$$\int_{-\infty}^{\infty} P_n^\lambda(x;\phi) P_m^\lambda(x;\phi) w^\lambda(x;\phi)\,dx = (2\pi)\frac{\Gamma(n+2\lambda)}{n!(2\sin\phi)^{2\lambda}}\delta_{nm}, \tag{1.16}$$

where the weight function is given by

$$w(x) = \exp\{(2\phi-\pi)x\}|\Gamma(\lambda+ix)|^2. \tag{1.17}$$

These are a limit case of the continuous Hahn polynomials, family (f) in the tableau, and also of the continuous dual Hahn polynomials, family (c).

Several of the families appearing in the Askey tableau have been generalized to multivariable biorthogonal polynomials; these include the Jacobi,[6] continuous Hahn,[7] and discrete Hahn[8] polynomials. In this paper, we present an analogous multivariable extension of the Meixner, Krawtchouk, and Meixner–Pollaczek polynomials. A limit of the Krawtchouk family then yields a trivial multivariable generalization of the Charlier polynomials. The biorthogonality relations are proved by introducing a set of auxiliary variables and then expressing the inner products as differentiations on these variables. Alternatively, the Krawtchouk and at least partly the Meixner–Pollaczek families can be obtained as limits of the multivariable discrete[8] and continuous[7] Hahn polynomials, respectively.

II. MULTIVARIABLE MEIXNER POLYNOMIALS

The extension to p variables $x_1,x_2,...,x_p$ is given in terms of multivariable hypergeometric series as

$$M^{c_1c_2\cdots c_p}_{n_1n_2\cdots n_p}(x_1,x_2,...,x_p;\beta) = (X+\beta)_N \times F^{0:2;...;2}_{1:0;...;0}\left(\begin{matrix}—: -n_1,-x_1;...;-n_p,-x_p \\ -N-X-\beta+1: —;...;—\end{matrix}; c_1^{-1}\cdots c_p^{-1}\right), \tag{2.1}$$

and also the distinct biorthogonal family

$$\overline{M}^{c_1c_2\cdots c_p}_{n_1n_2\cdots n_p}(x_1,x_2,...,x_p;\beta) = (\beta)_N F^{0:2;...;2}_{1:0;...;0}\left(\begin{matrix}—: -n_1,-x_1;...;-n_p,-x_p \\ \beta: —;...;—\end{matrix}; (C-1)c_1^{-1}\cdots(C-1)c_p^{-1}\right), \tag{2.2}$$

with the corresponding multivariable weight function

$$w^{c_1c_2\cdots c_p}(x_1,x_2,...,x_p;\beta) = \left[\prod_{k=1}^{p}\frac{c_k^{x_k}}{x_k!}\right]\frac{\Gamma(X+\beta)}{\Gamma(\beta)}. \tag{2.3}$$

$F^{q:l_1;...;l_p}_{r:v_1;...;v_p}$ is the generalized Kampé de Fériet hypergeometric series[9] defined as

$$F^{q:l_1;...;l_p}_{r:v_1;...;v_p}\left(\begin{matrix}\alpha_1,...,\alpha_q: \beta_1^{(1)},...,\beta_{l_1}^{(1)};...;\beta_1^{(p)},...,\beta_{l_p}^{(p)}; \\ \gamma_1,...,\gamma_r: \xi_1^{(1)},...,\xi_{v_1}^{(1)};...;\xi_1^{(p)},...,\xi_{v_p}^{(p)};\end{matrix} z_1,z_2,...,z_p\right) = \sum_{\{j_k\}}\frac{\Pi_{i=1}^{q}(\alpha_i)_J\Pi_{i=1}^{l_1}(\beta_i^{(1)})_{j_1}\cdots\Pi_{i=1}^{l_p}(\beta_i^{(p)})_{j_p}}{\Pi_{i=1}^{r}(\gamma_i)_J\Pi_{i=1}^{v_1}(\xi_i^{(1)})_{j_1}\cdots\Pi_{i=1}^{v_p}(\xi_i^{(p)})_{j_p}}\frac{z_1^{j_1}}{j_1!}\frac{z_2^{j_2}}{j_2!}\cdots\frac{z_p^{j_p}}{j_p!}, \tag{2.4}$$

where we are using the following shorthand notation,

$$X\equiv\sum_{k=1}^{p}x_k,\quad N\equiv\sum_{k=1}^{p}n_k,$$
$$J\equiv\sum_{k=1}^{p}j_k,\quad C\equiv\sum_{k=1}^{p}c_k, \tag{2.5}$$

and $\{j_k\}$ denotes summation indices $j_1j_2,...,j_p$, which run over all non-negative integers with the convention that $1/\Gamma(-m)=0, m=0,1,2,...$. This multivariable extension is nontrivial in that the polynomials and weight function do not factor with respect to the independent variables. The overbar in (2.2) denotes the biorthogonal family and should not be confused with complex conjugation. The $p+1$ complex parameters $c_1,c_2,...,c_p$, β are restricted by

$$c_k\neq 0,\quad \sum_{k=1}^{p}|c_k|<1, \tag{2.6}$$

and identify a particular family of polynomials and their weight function. The variable β is arbitrary, including negative integers. In the latter case, (2.1)–(2.3) are still well defined through use of the identity.

$$\frac{\Gamma(\beta+l)}{\Gamma(\beta)} = (-1)^l\frac{\Gamma(-\beta+1)}{\Gamma(-\beta-l+1)},\quad l=0,1,2,... . \tag{2.7}$$

The set of p non-negative integers $n_1,n_2,...,n_p$ label the members of a given family and the degree of a polynomial is simply given by N. When no ambiguity arises, we simply write $M_n(x)$,$\overline{M}_n(x)$ and $w(x)$ for the polynomials and weight function, respectively. In the special case of a single variable, $M_n(x)$ and $\overline{M}_n(x)$ both reduce to the familiar Meixner polynomials, but in general they are distinct.

The polynomials $M_n(x)$ have a Rodrigues-type formula that is a natural extension of the single variable expression

$$M_n(x) = \frac{1}{\Gamma(X+\beta)}\left[\prod_{k=1}^{p}\Gamma(x_k+1)c_k^{-n_k-x_k}D_k^{n_k}\right] \times\left[\Gamma(X+\beta)\prod_{k=1}^{p}\frac{c_k^{x_k}}{\Gamma(x_k-n_k+1)}\right], \tag{2.8}$$

where D_k is the multivariable generalization of (1.3) defined as

$$D_kf(x_1\cdots x_k\cdots x_p)\equiv f(x_1\cdots x_k+1\cdots x_p) - f(x_1\cdots x_k\cdots x_p), \tag{2.9}$$

whereas the analogous expression for $\overline{M}_n(x)$ is more obscure; here Y is kept fixed during the differenciating and is set equal to X afterwards,

$$\overline{M}_n(x) = (1-C)^{N+X}\left[\prod_{k=1}^{p}\Gamma(x_k+1)c_k^{-n_k-x_k}D_k^{n_k}\right] \times\left[(1-C)^{-X}\frac{\Gamma(N+\beta)}{\Gamma(Y-X+N+\beta)} \times\prod_{k=1}^{p}\frac{c_k^{x_k}}{\Gamma(x_k-n_k+1)}\right]. \tag{2.10}$$

To verify these representations, one substitutes the identity

$$D_k^{n_k} f(x_1 \cdots x_k \cdots x_p) = \sum_{j_k=0}^{n_k} \binom{n_k}{j_k} (-1)^{j_k} f(x_1 \cdots x_k + n_k - j_k \cdots x_p) \quad (2.11)$$

in (2.8) and (2.10), which immediately yields (2.1) and (2.2), respectively. Identity (2.11) is in turn easily proved by induction on n_k.

Next we calculate the norm of the weight function

$$\sum_{\{x_k\}} w(x) = \sum_{\{x_k\}} \left[\prod_{k=1}^{p} \frac{c_k^{x_k}}{x_k!} \right] \frac{\Gamma(X+\beta)}{\Gamma(\beta)}, \quad (2.12)$$

where the sum is over all non-negative integers $x_k = 0,1,2,...,\infty$, $k = 1,2,...,p$. Using identity (2.7), this can be written in the form

$$\sum_{\{x_k\}} \left[\prod_{k=1}^{p} \frac{c_k^{x_k}}{x_k!} \right] \frac{\Gamma(X+\beta)}{\Gamma(\beta)} = \sum_{\{x_k\}} \left[\prod_{k=1}^{p} \frac{(-c_k)^{x_k}}{x_k!} \right] \frac{\Gamma(-\beta+1)}{\Gamma(-\beta-X+1)}, \quad (2.13)$$

and if we then isolate one of the summations, say the x_1 sum

$$\sum_{\{x_2 \cdots x_p\}} \left[\prod_{k=2}^{p} \frac{(-c_k)^{x_k}}{x_k!} \right] \frac{\Gamma(-\beta+1)}{\Gamma(-\beta-x_2 \cdots -x_p+1)} \times \sum_{x_1=0}^{\infty} \binom{-\beta-x_2 \cdots -x_p}{x_1} (-c_1)^{x_1}, \quad (2.14)$$

we find just the binomial series. By (2.6), $|c_1| < 1$, so the series converges and we obtain

$$\sum_{\{x_2 \cdots x_p\}} \left[\prod_{k=2}^{p} \frac{(-c_k)^{x_k}}{x_k!} \right] \times \frac{\Gamma(-\beta+1)}{\Gamma(-\beta-x_2 \cdots -x_p+1)} (1-c_1)^{-\beta-x_2 \cdots -x_p}, \quad (2.15)$$

and if we then isolate the x_2 sum

$$\sum_{\{x_3 \cdots x_p\}} \left[\prod_{k=3}^{p} \frac{(-c_k)^{x_k}}{x_k!} \right] \frac{\Gamma(-\beta+1)}{\Gamma(-\beta-x_3 \cdots -x_p+1)} \times (1-c_1)^{-\beta-x_3 \cdots -x_p} \times \sum_{x_2=0}^{\infty} \binom{-\beta-x_3 \cdots -x_p}{x_2} \left(\frac{-c_2}{1-c_1} \right)^{x_2}, \quad (2.16)$$

we again find the binomial series and by (2.6) we have once more

$$\left| \frac{-c_2}{1-c_1} \right| < 1, \quad (2.17)$$

so this series converges, also giving

$$\sum_{\{x_3 \cdots x_p\}} \left[\prod_{k=3}^{p} \frac{(-c_k)^{x_k}}{x_k!} \right] \times \frac{\Gamma(-\beta+1)}{\Gamma(-\beta-x_3 \cdots -x_p+1)} (1-c_1-c_2)^{-\beta-x_3 \cdots -x_p}. \quad (2.18)$$

The remaining sums are then performed by induction, with (2.6) ensuring that the argument of each binomial series has a modulus less than one. The final result is

$$\sum_{\{x_k\}} \left[\prod_{k=1}^{p} \frac{c_k^{x_k}}{x_k!} \right] \frac{\Gamma(X+\beta)}{\Gamma(\beta)} = (1-C)^{-\beta}. \quad (2.19)$$

In the remainder of this section, we discuss the orthogonality and biorthogonality properties of the multivariable Meixner polynomials. First we demonstrate that the inner product of $M_n(x)$ with another polynomial of the same family $M_{n'}(x)$ vanishes if $N \neq N'$, that is, any two polynomials of differing degree are orthogonal. Consider the inner product

$$\sum_{\{x_k\}} M_n(x) M_{n'}(x) w(x), \quad N \neq N', \quad (2.20)$$

where without loss of generality we assume $N > N'$. The polynomial $M_{n'}(x)$ can be expanded (as can an arbitrary polynomial) in the following manner:

$$M_{n'}(x) = \sum_{\{j'_k\}} a(j'_1 j'_2,...,j'_p) \prod_{k=1}^{p} \frac{\Gamma(x_k+1)}{\Gamma(x_k - j'_k + 1)}, \quad (2.21)$$

where $a(j'_1 j'_2,...j'_p)$ are some constants which we need not evaluate, and the $\{j'_k\}$ sums run over non-negative integers such that $0 \leqslant J' \leqslant N'$. Then referring to (2.20) we consider the following inner product:

$$I \equiv \sum_{\{x_k\}} M_n(x) \left[\prod_{k=1}^{p} \frac{\Gamma(x_k+1)}{\Gamma(x_k - j'_k + 1)} \right] w(x), \quad N > J', \quad (2.22)$$

which upon substituting (2.1) and (2.3) becomes

$$I = \sum_{\{j_k\}} \left[\prod_{k=1}^{p} \binom{n_k}{j_k} (-c_k)^{-j_k} \right] \times \sum_{\{x_k\}} \left[\prod_{k=1}^{p} \frac{c_k^{x_k} x_k!}{(x_k - j_k)!(x_k - j'_k)!} \right] \times \frac{\Gamma(N-J+\beta+X)}{\Gamma(\beta)}. \quad (2.23)$$

To evaluate the $\{x_k\}$ sum, we introduce a set of auxiliary real variables t_k, $k = 1,2,...,p$ confined to the domain

$$0 < t_k < \left(\sum_{k=1}^{p} |c_k| \right)^{-1}, \quad k = 1,2,...,p, \quad (2.24)$$

and write I in terms of differentiations on these variables evaluated at $t_k = 1$,

$$I = \sum_{\{j_k\}} \left[\prod_{k=1}^{p} \binom{n_k}{j_k} (-c_k)^{-j_k} \right] \times \sum_{\{x_k\}} \left[\prod_{k=1}^{p} \left(\frac{\partial}{\partial t_k} \right)_{t_k=1}^{j'_k} \frac{(t_k c_k)^{x_k}}{(x_k - j_k)!} \right] \times \frac{\Gamma(N-J+X+\beta)}{\Gamma(\beta)}, \quad (2.25)$$

and then changing the summation indices x_k to $x'_k = x_k - j_k$, one obtains (after dropping the primes)

$$I=\sum_{\{j_k\}}\left[\prod_{k=1}^{p}\binom{n_k}{j_k}(-1)^{j_k}\right]$$
$$\times\sum_{\{x_k\}}\left[\prod_{k=1}^{p}\left(\frac{\partial}{\partial t_k}\right)^{j'_k}_{t_k=1} t_k^{j_k}\frac{(t_kc_k)^{x_k}}{x_k!}\right]$$
$$\times\frac{\Gamma(N+X+\beta)}{\Gamma(\beta)}, \tag{2.26}$$

and from (2.24) we have

$$\sum_{k=1}^{p}|t_kc_k|<1. \tag{2.27}$$

This sum is uniformly convergent on the parameter range (2.27) so we can interchange the order of the differentiations with the summations, and then substituting (2.19) (with t_kc_k replacing c_k and $N+\beta$ replacing β), one obtains

$$I=(\beta)_N\left[\prod_{k=1}^{p}\left(\frac{\partial}{\partial t_k}\right)^{j'_k}_{t_k=1}\right]\sum_{\{j_k\}}\left[\prod_{k=1}^{p}\binom{n_k}{j_k}(-t_k)^{j_k}\right]$$
$$\times\left(1-\sum_{k=1}^{p}t_kc_k\right)^{-N-\beta}, \tag{2.28}$$

and now one notices that the $\{j_k\}$ sum is simply the product of binomial expansions for integer powers. Summing these, we find

$$I=(\beta)_N\left[\prod_{k=1}^{p}\left(\frac{\partial}{\partial t_k}\right)^{j'_k}_{t_k=1}(1-t_k)^{n_k}\right]$$
$$\times\left(1-\sum_{k=1}^{p}t_kc_k\right)^{-N-\beta}, \tag{2.29}$$

and it is obvious that if any factor of $(1-t_k)$ survives the differentiations, it will vanish upon setting $t_k=1$. The total order of differentiations is J' so for $N>J'$ at least one factor of $(1-t_k)$ for some k will survive in every term after the differentiations and then will vanish. Thus recalling the definition of I, we have shown

$$\sum_{\{x_k\}}M_n(x)\left[\prod_{k=1}^{p}\frac{\Gamma(x_k+1)}{\Gamma(x_k-j'_k+1)}\right]w(x)=0,\quad\text{if } N>J', \tag{2.30}$$

and then referring back to (2.20) and (2.21), we immediately deduce that

$$\sum_{\{x_k\}}M_n(x)M_{n'}(x)w=0,\quad\text{if } N\neq N', \tag{2.31}$$

so $M_n(x)$ are orthogonal with respect to subspaces labelled by N, which however says nothing about polynomials of the same degree.

Now we demonstrate the analogous result for the biorthogonal family $\overline{M}_n(x)$ in essentially the same manner. As before, we expand $\overline{M}_{n'}(x)$ as

$$\overline{M}_{n'}(x)=\sum_{\{j'_k\}}\bar{a}(j'_1,j'_2,\ldots,j'_p)\prod_{k=1}^{p}\frac{\Gamma(x_k+1)}{\Gamma(x_k-j'_k+1)}, \tag{2.32}$$

where $\bar{a}(j'_1,j'_2,\ldots,j'_p)$ are some constants and the $\{j'_k\}$ sums are over non-negative integers such that $0\leqslant J'\leqslant N'$. Then we consider the inner product

$$\overline{I}\equiv\sum_{\{x_k\}}\overline{M}_n(x)\left[\prod_{k=1}^{p}\frac{\Gamma(x_k+1)}{\Gamma(x_k-j'_k+1)}\right]w(x),\quad N>J', \tag{2.33}$$

which upon substituting (2.2) and (2.3) gives

$$\overline{I}=(\beta)_N\sum_{\{j_k\}}\left[\prod_{k=1}^{p}\binom{n_k}{j_k}c_k^{-j_k}\right](C-1)^J$$
$$\times\sum_{\{x_k\}}\left[\prod_{k=1}^{p}\frac{c_k^{x_k}x_k!}{(x_k-j_k)!(x_k-j'_k)!}\right]\frac{\Gamma(X+\beta)}{\Gamma(J+\beta)}, \tag{2.34}$$

and again introducing the parameters t_k, $k=1,2,\ldots p$ confined to the domain (2.24),

$$\overline{I}=(\beta)_N\sum_{\{j_k\}}\left[\prod_{k=1}^{p}\binom{n_k}{j_k}c_k^{-j_k}\right](C-1)^J$$
$$\times\sum_{\{x_k\}}\left[\prod_{k=1}^{p}\left(\frac{\partial}{\partial t_k}\right)^{j'_k}_{t_k=1}\frac{(t_kc_k)^{x_k}}{(x_k-j_k)!}\right]\frac{\Gamma(X+\beta)}{\Gamma(J+\beta)}, \tag{2.35}$$

and then changing the summation indices x_k to $x'_k=x_k-j_k$, we obtain (after dropping the primes)

$$\overline{I}=(\beta)_N\sum_{\{j_k\}}\left[\prod_{k=1}^{p}\binom{n_k}{j_k}\right](C-1)^J$$
$$\times\sum_{\{x_k\}}\left[\prod_{k=1}^{p}\left(\frac{\partial}{\partial t_k}\right)^{j'_k}_{t_k=1}\frac{t_k^{j_k}(t_kc_k)^{x_k}}{x_k!}\right]$$
$$\times\frac{\Gamma(X+J+\beta)}{\Gamma(J+\beta)}. \tag{2.36}$$

As before, we bring the differentiations outside the sums and then substitute (2.19) (with t_kc_k replacing c_k and $J+\beta$ replacing β), which yields

$$\overline{I}=(\beta)_N\left[\prod_{k=1}^{p}\left(\frac{\partial}{\partial t_k}\right)^{j'_k}_{t_k=1}\right]$$
$$\times\sum_{\{j_k\}}\left[\prod_{k=1}^{p}\binom{n_k}{j_k}\left(1-\sum_{k=1}^{p}t_kc_k\right)^{n_k-j_k}\right.$$
$$\left.\times\{-(1-C)t_k\}^{j_k}\right]\left(1-\sum_{k=1}^{p}t_kc_k\right)^{-N-\beta}, \tag{2.37}$$

and once more $\{j_k\}$ sums are simply binomial expansions for integer powers. Summing these, we obtain

$$\overline{I}=(\beta)_N\left[\prod_{k=1}^{p}\left(\frac{\partial}{\partial t_k}\right)^{j'_k}_{t_k=1}\left\{1-\sum_{k=1}^{p}t_kc_k-(1-C)t_k\right\}^{n_k}\right]$$
$$\times\left(1-\sum_{k=1}^{p}t_kc_k\right)^{-N-\beta}, \tag{2.38}$$

and again it is obvious that if any factor of $1-\Sigma_{k=1}^{p}t_kc_k-(1-C)t_k$ survives the differentiations, it will vanish upon setting $t_k=1$. The total order of differentiations is J' so for $N>J'$ at least one factor of $1-\Sigma_{k=1}^{p}t_kc_k-(1-C)t_k$ for some k will survive in every term after the differentiations and then will vanish. Thus recalling the definition of $\overline{I}$, we have shown

$$\sum_{\{x_k\}} \overline{M}_n(x)\left[\prod_{k=1}^{p} \frac{\Gamma(x_k+1)}{\Gamma(x_k-j'_k+1)}\right]w(x)=0, \quad \text{if } N>J', \tag{2.39}$$

and then referring to (2.32) we immediately deduce that

$$\sum_{\{x_k\}} \overline{M}_n(x)\overline{M}_{n'}(x)w(x)=0, \quad \text{if } N\neq N', \tag{2.40}$$

and so $\overline{M}_n(x)$ are also orthogonal with respect to subspaces labelled by N.

Though the polynomial familes $M_n(x)$ and $\overline{M}_n(x)$ are each orthogonal among themselves with respect to degree N, they are not so for different polynomials of the same degree. We now demonstrate that these two families form a biorthogonal system; consider the inner product

$$\sum_{\{x_k\}} M_n(x)\overline{M}_{n'}(x)w(x). \tag{2.41}$$

If $N>N'$ we expand $\overline{M}_{n'}(x)$ as in (2.32) and then by (2.30) this inner product vanishes, whereas if $N<N'$ we expand $M_n(x)$ as in (2.21) and by (2.39) it again vanishes. So this inner product is zero unless $N=N'$, which we now assume. Substituting (2.2) into (2.41) gives

$$\sum_{\{j'_k\}}\left[\prod_{k=1}^{p}\binom{n'_k}{j'_k}c_k^{-j'_k}\right]\frac{\Gamma(N'+\beta)}{\Gamma(J'+\beta)}(C-1)^{J'} \times\sum_{\{x_k\}} M_n(x)\left[\prod_{k=1}^{p}\frac{x_k!}{(x_k-j'_k)!}\right]w(x), \tag{2.42}$$

and noting that the $\{x_k\}$ sum is just the inner product I which we already calculated in (2.29), this becomes

$$(\beta)_N\sum_{\{j'_k\}}\left[\prod_{k=1}^{p}\binom{n'_k}{j'_k}c_k^{-j'_k}\left(\frac{\partial}{\partial t_k}\right)^{j'_k}_{t_k=1}(1-t_k)^{n_k}\right] \times\frac{\Gamma(N'+\beta)}{\Gamma(J'+\beta)}(C-1)^{J'}\left(1-\sum_{k=1}^{p}t_kc_k\right)^{-N-\beta}. \tag{2.43}$$

As we discussed with the previous cases, if the order of the differentiations J' is less than the degree of the factors $(1-t_k)^{n_k}$ which is N, then one obtains zero. So the only term in the $\{j'_k\}$ sum which does not vanish is that for which $J'=N'=N$, that is, for $j'_k=n'_k$ $k=1,2,\ldots,p$, and then the inner product simplifies to

$$(\beta)_N(C-1)^N\left[\prod_{k=1}^{p}c_k^{-n'_k}\left(\frac{\partial}{\partial t_k}\right)^{n'_k}_{t_k=1}(1-t_k)^{n_k}\right] \times\left(1-\sum_{k=1}^{p}t_kc_k\right)^{-N-\beta}. \tag{2.44}$$

Now if $n'_k>n_k$ for some k, and recalling that $N=N'$, then $n'_{k'}<n_{k'}$ for some other k'. In that case, a factor of $(1-t_{k'})$ will survive in every term after the differentiations and then will vanish upon setting $t_{k'}=1$. So (2.44) is zero unless $n'_k=n_k$ for every k, and in the nonvanishing case it simply becomes

$$(\beta)_N\left[\prod_{k=1}^{p}n_k!c_k^{-n_k}\right](1-C)^{-\beta}, \tag{2.45}$$

so the inner product (2.41) is simply

$$\sum_{\{x_k\}} M_n(x)\overline{M}_{n'}(x)w(x) =(\beta)_N\left[\prod_{k=1}^{p}n_k!c_k^{-n_k}\delta_{n_kn'_k}\right](1-C)^{-\beta}. \tag{2.46}$$

III. MULTIVARIABLE KRAWTCHOUK POLYNOMIALS

A special case of the previously discussed Meixner family are the multivariable Krawtchouk polynomials. These are obtained for

$$\beta=-\Delta, \quad c_k=\frac{q_k}{(Q-1)}, \quad k=1,2,\ldots,p, \tag{3.1}$$

where Δ is a non-negative integer and q_k, $k=1,2,\ldots,p$ are real parameters satisfying

$$q_k>0, \quad 0<Q<1, \quad Q\equiv\sum_{k=1}^{p}q_k. \tag{3.2}$$

As we already mentioned, (2.1), (2.2), and (2.3) are well defined for β a negative integer through use of identity (2.7). Also making a change in normalization analogous to (1.6), we define the multivariable Krawtchouk polynomials as

$$K_n(x)=(-\Delta+X)_N\left[\prod_{k=1}^{p}q_k^{n_k}/n_k!\right]F^{0:2;\ldots;2}_{1:0;\ldots;0}\left(\begin{matrix}\text{—}:-n_1,-x_1;\ldots;-n_p,-x_p:\\ \Delta-N-X+1:\text{—};\ldots;\text{—}\end{matrix};(Q-1)q_1^{-1}\cdots(Q-1)q_p^{-1}\right), \tag{3.3}$$

and their biorthogonal counterparts

$$\overline{K}_n(x)=(-\Delta)_N\left[\prod_{k=1}^{p}q_k^{n_k}/n_k!\right]F^{0:2;\ldots;2}_{1:0;\ldots;0}\left(\begin{matrix}\text{—}:-n_1,x_1;\ldots;-n_p,-x_p:\\ -\Delta:\text{—};\ldots;\text{—}\end{matrix};q_1^{-1}\cdots q_p^{-1}\right), \tag{3.4}$$

where for given Δ the indices $n_1,n_2,\ldots,n_p$ are confined by $0\leqslant N\leqslant\Delta$. The corresponding Rodrigues formulas as deduced from (2.8) and (2.10) are

$$K_n(x)=(-1)^N\frac{\Gamma(\Delta-X+1)}{(1-Q)^{\Delta-X}}\left[\prod_{k=1}^{p}\frac{\Gamma(x_k+1)}{n_k!q_k^{x_k}}D_k^{n_k}\right]\left[\frac{(1-Q)^{\Delta+N-X}}{\Gamma(\Delta-X+1)}\prod_{k=1}^{p}\frac{q_k^{x_k}}{\Gamma(x_k-n_k+1)}\right], \tag{3.5}$$

and

$$\overline{K}_n(x)=(-1)^N\left[\prod_{k=1}^{p}\frac{\Gamma(x_k+1)}{n_k!q_k^{x_k}}D_k^{n_k}\right]\left[\frac{\Gamma(\Delta-N+X-Y+1)}{\Gamma(\Delta-N+1)}\prod_{k=1}^{p}\frac{q_k^{x_k}}{\Gamma(x_k-n_k+1)}\right], \tag{3.6}$$

where in (3.6) Y is kept fixed during the differenciating and then is set equal to X afterwards. The weight function is the multivariate binomial distribution

$$w(x) = \frac{\Delta!}{(\Delta - X)!}\left[\prod_{k=1}^{p} \frac{q_k^{x_k}}{x_k!}\right](1-Q)^{\Delta - X}, \tag{3.7}$$

and the $\{x_k\}$ sum defining the inner product is over the finite set of non-negative integers lying on the discrete simplex $0 \leqslant X \leqslant \Delta$. In this respect the infinite $\{x_k\}$ sums have been truncated by β being a negative integer.

Notice that the c_k parameters as defined by (3.1) and (3.2) do not satisfy the second restriction in (2.6). That condition was necessary so that the infinite binomial series would converge. Now, however, since we are dealing with finite sums, this restriction is no longer necessary.

The orthogonality and biorthogonality proofs follow as before with the only difference being that the summations are now finite. Then corresponding to (2.31) and (2.40) we have that the multivariable Krawtchouk polynomials are orthogonal with respect to subspaces labelled by $N(0 \leqslant N \leqslant \Delta)$

$$\begin{aligned} &\sum_{\{x_k\}} K_n(x) K_{n'}(x) w(x) = 0, \quad \text{if } N \neq N', \\ &\sum_{\{x_k\}} \overline{K}_n(x) \overline{K}_{n'}(x) w(x) = 0, \quad \text{if } N \neq N', \end{aligned} \tag{3.8}$$

while in general these two families are biorthogonal, corresponding to (2.46),

$$\sum_{\{x_k\}} K_n(x) \overline{K}_{n'}(x) w(x) = \frac{\Delta!}{(\Delta - N)!}\left[\prod_{k=1}^{p} \frac{q_k^{n_k}}{n_k!}\delta_{n_k n_k'}\right](1-Q)^N \tag{3.9}$$

where we recall that the $\{x_k\}$ sum is over the discrete simplex $0 \leqslant X \leqslant \Delta$.

The Askey tableau shows that the single variable Krawtchouk polynomials can be obtained as a limit of the Hahn polynomials; this still holds true for the multivariable biorthogonal families. In the multivariable Hahn[8] polynomials $G_n^{\alpha\beta}(x)$ and $H_n^{\alpha\beta}(x)$ set $\alpha_k + 1 = \gamma q_k$, $k = 1,2,\dots,p$, $\beta + 1 = \gamma(1-Q)$ and then take the limit $\gamma \to \infty$. This yields

$$\begin{aligned} \lim_{\gamma\to\infty} G_n^{\alpha\beta}(x) &= \frac{1}{(-\Delta)_N}\left[\prod_{k=1}^{p} n_k! q_k^{-n_k}\right] K_n(x), \\ \lim_{\gamma\to\infty} H_n^{\alpha\beta}(x) &= \frac{1}{(-\Delta)_N}\left[\prod_{k=1}^{p} n_k! q_k^{-n_k}\right] \overline{K}_n(x), \end{aligned} \tag{3.10}$$

while from the Hahn weight function $w^{\alpha\beta}(x)$ one obtains the Krawtchouk weight

$$\begin{aligned} &\lim_{\gamma\to\infty} \Delta! \gamma^{-\Delta}\left[\Gamma(\beta+1)\prod_{k=1}^{p}\Gamma(\alpha_k+1)\right]^{-1} w^{\alpha\beta}(x) \\ &= \frac{\Delta!}{(\Delta - X)!}\left[\prod_{k=1}^{p} \frac{q_k^{x_k}}{x_k!}\right](1-Q)^{\Delta - X}. \end{aligned} \tag{3.11}$$

The orthogonality (3.8) and biorthogonality (3.9) relations also follow in this limit.

The Askey tableau also shows that the single variable Charlier polynomials are a limit case of the Krawtchouk polynomials; let us consider this limit in the multivariable case. Set $q_k = a_k/\Delta$ and then take the limit $\Delta \to \infty$ with a_k, x_k, and n_k fixed. One finds

$$\lim_{\Delta\to\infty} K_n(x) = \lim_{\Delta\to\infty} \overline{K}_n(x) = \prod_{k=1}^{p} \frac{1}{n_k!} C_{n_k}^{(a_k)}(x_k), \tag{3.12}$$

that is, both families of Krawtchouk polynomials reduce to just a product of single variable Charlier polynomials. Furthermore, the Krawtchouk weight function (3.7) in this limit reduces to just a product of single variable Charlier weights.

IV. MULTIVARIABLE MEIXNER–POLLACZEK POLYNOMIALS

The corresponding orthogonality relations for the Meixner–Pollaczek family are deduced in a similar manner. The extension to p variables $x_1, x_2, \dots, x_p$ can be expressed in terms of multivariable hypergeometric series as follows:

$$\begin{aligned} P_{n_1 n_2 \cdots n_p}^{\lambda_1 \lambda_2 \cdots \lambda_{p+1}}(x_1,x_2,\dots,x_p;\phi_1,\phi_2\dots,\phi_p) &= (\lambda_{p+1} - iX)_N\left[\prod_{k=1}^{p} \exp\{in_k\phi_k\}/n_k!\right] \\ &\times F_{1:0;\dots;0}^{0:2;\dots;2}\left(\begin{matrix} —: -n_1,\lambda_1 + ix_1;\dots; -n_p,\lambda_p + ix_p \\ -N - \lambda_{p+1} + iX + 1: —;\dots;— \end{matrix}; \exp\{-2i\phi_1\}\cdots\exp\{-2i\phi_p\}\right), \end{aligned} \tag{4.1}$$

and the distinct biorthogonal family

$$\begin{aligned} \overline{P}_{n_1 n_2 \cdots n_p}^{\lambda_1 \lambda_2 \cdots \lambda_{p+1}}(x_1,x_2,\dots,x_p;\phi_1,\phi_2,\dots,\phi_p) &= (\lambda)_N\left[\prod_{k=1}^{p} \exp\{in_k\phi_k\}/n_k!\right] \\ &\times F_{1:0;\dots;0}^{0:2;\dots;2}\left(\begin{matrix} —: -n_1,\lambda_1 + ix_1;\dots; -n_p,\lambda_p + ix_p \\ \lambda: —;\dots;— \end{matrix}; -f_1(\phi)\cdots -f_p(\phi)\right), \end{aligned} \tag{4.2}$$

$$f_k(\phi)\equiv\Big(1-\sum_{j=1}^{p}\exp\{2i\phi_j\}\Big)\exp\{-2i\phi_k\},$$

with the multivariable weight function

$$w^{\lambda_1\cdots\lambda_{p+1}}(x_1\cdots x_p;\phi_1\cdots\phi_p)$$
$$=\Big[\prod_{k=1}^{p}\exp\{(2\phi_k-\pi)x_k\}\Gamma(\lambda_k+ix_k)\Big]$$
$$\times\Gamma(\lambda_{p+1}-iX), \qquad (4.3)$$

where the $p+1$ lambda parameters $\lambda_1,\lambda_2,\ldots,\lambda_{p+1}$ are complex and the p angle parameters $\phi_1,\phi_2,\ldots,\phi_p$ are real. We are using the same shorthand notation as in previous sections, and in addition we have defined $\lambda\equiv\Sigma_{k=1}^{p+1}\lambda_k$; we simply write $P_n(x),\overline{P}_n(x)$ and $w(x)$ for the polynomials and weight function, respectively. In the special case of a single variable, $P_n(x)$ and $\overline{P}_n(x)$ both reduce to the familiar single variable Meixner–Pollaczek polynomials, but in general they are distinct.

The angle parameters ϕ_k are confined to

$$0<\phi_k<\pi,\quad k=1,2,\ldots,p, \qquad (4.4)$$

and if the real parts of the complex parameters $\lambda_1,\lambda_2,\ldots,\lambda_{p+1}$ are greater than zero, the p integration contours are simply the p real axes. For more general values of these lambda parameters, the contours are deformed to separate the increasing sequences of poles of the weight function from the decreasing sequences, which is possible whenever the two sets are disjoint. In the remainder of this section, we simply write the contours as the real axes, but one should interpret them as above if necessary.

These polynomials also have Rodrigues-type formulas

$$P_n(x)=\frac{i^N}{\Gamma(\lambda_{p+1}-iX)}\Big[\prod_{k=1}^{p}\frac{\exp\{-(2\phi_k-\pi)x_k\}}{n_k!\Gamma(\lambda_k+ix_k)}\delta_k^{n_k}\Big]\Big[\Gamma(\lambda_{p+1}+N/2-iX)\prod_{k=1}^{p}\Gamma(\lambda_k+n_k/2+ix_k)$$
$$\times\exp\{(2\phi_k-\pi)x_k\}\Big], \qquad (4.5)$$
$$\overline{P}_n(x)=i^N\Big(1-\sum_{k=1}^{p}\exp\{2i\phi_k\}\Big)^{N/2-iX}\Big[\prod_{k=1}^{p}\frac{\exp\{-(2\phi_k-\pi)x_k\}}{n_k!\Gamma(\lambda_k+ix_k)}\delta_k^{n_k}\Big],$$
$$\times\Big[\Big(1-\sum_{k=1}^{p}\exp\{2i\phi_k\}\Big)^{iX}\frac{\Gamma(N+\lambda)}{\Gamma(N/2+\lambda+iX-iY)}\prod_{k=1}^{p}\Gamma(\lambda_k+n_k/2+ix_k)\exp\{(2\phi_k-\pi)x_k\}\Big],$$

where δ_k is the multivariable generalization of (1.15),

$$\delta_k f(x_1\cdots x_k\cdots x_p)\equiv f(x_1\cdots x_k+i/2\cdots x_p)$$
$$-f(x_1\cdots x_k-i/2\cdots x_p), \qquad (4.6)$$

and as before in the second Rodrigues formula, Y is kept fixed during the differenciating and then is set equal to X afterwards. To verify these representations, one substitutes the identity

$$\delta_k^{n_k}f(x_1\cdots x_k\cdots x_p)$$
$$=\sum_{j_k=0}^{n_k}\binom{n_k}{j_k}(-1)^{j_k}f(x_1\cdots x_k+in_k/2-ij_k\cdots x_p) \qquad (4.7)$$

in (4.5) and (4.6), which immediately yields (4.1) and (4.2), respectively. Identity (4.7) is in turn easily proved by induction on n_k.

Next consider the following multiple integral,

$$\int_{-\infty}^{\infty}dx_1\cdots\int_{-\infty}^{\infty}dx_p$$
$$\times\Big[\prod_{k=1}^{p}\Gamma(\lambda_k+ix_k)z_k^{ix_k}\Big]\Gamma(\lambda_{p+1}-iX),$$
$$-\pi<\arg(z_k)<\pi,\quad k=1,2,\ldots,p \qquad (4.8)$$

which for $z_k=\exp\{-i(2\phi_k-\pi)\}$ is the norm of the weight function. This is a multiple Mellin–Barnes type integral which can be calculated by induction and the following single integral formula,[10]

$$\int_{-\infty}^{\infty}dx\,\Gamma(\lambda_1+ix)\Gamma(\lambda_2-ix)z^{ix}$$
$$=(2\pi)\Gamma(\lambda_1+\lambda_2)z^{-\lambda_1}\Big(1+\frac{1}{z}\Big)^{-\lambda_1-\lambda_2}, \qquad (4.9)$$

where $-\pi<\arg(z)<\pi$, and in this manner one obtains

$$\int_{-\infty}^{\infty}dx_1\cdots\int_{-\infty}^{\infty}dx_p\Big[\prod_{k=1}^{p}\Gamma(\lambda_k+ix_k)z_k^{ix_k}\Big]\Gamma(\lambda_{p+1}-iX)$$
$$=(2\pi)^p\Gamma(\lambda)\Big[\prod_{k=1}^{p}z_k^{-\lambda_k}\Big]\Big[1+\sum_{k=1}^{p}\frac{1}{z_k}\Big]^{-\lambda}. \qquad (4.10)$$

In the remainder of this section, we derive the orthogonality and biorthogonality relations of these polynomials; we demonstrate that they satisfy continuous analogs of (2.31), (2.40), and (2.46). First we show that the inner product,

$$\int_{-\infty}^{\infty}dx_1\cdots\int_{-\infty}^{\infty}dx_p\,P_n(x)P_{n'}(x)w(x),\quad N\neq N', \qquad (4.11)$$

vanishes if $N\neq N'$. Without loss of generality we assume $N>N'$, then we expand $P_{n'}(x)$ as follows:

$$P_{n'}(x)=\sum_{\{j_k'\}}b(j_1'j_2',\ldots,j_p')\prod_{k=1}^{p}\frac{\Gamma(j_k'+\lambda_k+ix_k)}{\Gamma(\lambda_k+ix_k)}, \qquad (4.12)$$

where $b(j_1'j_2',\ldots,j_p')$ are some constants which we need not evaluate and the $\{j_k'\}$ sum is over the range $0\leqslant J'\leqslant N'$. Next we consider the inner product

$$I \equiv \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p\, P_n(x) \times \left[\prod_{k=1}^{p} \frac{\Gamma(j'_k + \lambda_k + ix_k)}{\Gamma(\lambda_k + ix_k)}\right] w(x), \quad N > J', \tag{4.13}$$

which upon substituting (4.1) and (4.3) becomes

$$I \equiv \sum_{\{j_k\}} \left[\prod_{k=1}^{p} \binom{n_k}{j_k} \frac{1}{n_k!} \exp\{i(n_k - 2j_k)\phi_k\}\right] \times \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p \left[\prod_{k=1}^{p} \Gamma(j_k + \lambda_k + ix_k) \times \frac{\Gamma(j'_k + \lambda_k + ix_k)}{\Gamma(\lambda_k + ix_k)} \times \exp\{(2\phi_k - \pi)x_k\}\right] \Gamma(N - J + \lambda_{p+1} - iX). \tag{4.14}$$

In analogy with the discrete case, we introduce a set of auxiliary real parameters t_k, $k = 1,2,...,p$, so that we may write

$$\frac{\Gamma(j'_k + \lambda_k + ix_k)}{\Gamma(\lambda_k + ix_k)} = \left(\frac{\partial}{\partial t_k}\right)^{j'_k}_{t_k=1} t_k^{j'_k + \lambda_k + ix_k - 1}, \tag{4.15}$$

and then substituting this into the expression for I, we obtain

$$I = \sum_{\{j_k\}} \left[\prod_{k=1}^{p} \binom{n_k}{j_k} \frac{1}{n_k!} \exp\{i(n_k - 2j_k)\phi_k\}\right] \times \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p \left[\prod_{k=1}^{p} \left(\frac{\partial}{\partial t_k}\right)^{j'_k}_{t_k=1} t_k^{j'_k + \lambda_k - 1}\right] \times \left[\prod_{k=1}^{p} \Gamma(j_k + \lambda_k + ix_k) z_k^{ix_k}\right] \Gamma(N - J + \lambda_{p+1} - iX), \tag{4.16}$$

where we have defined

$$z_k \equiv t_k \exp\{-i(2\phi_k - \pi)\}, \tag{4.17}$$

and by (4.4) these z_k parameters satisfy the conditions in (4.8). This integral is uniformly convergent on this parameter range, so we can interchange the order of integrations with differentiations, and then substituting (4.10) for the multiple integral, we obtain

$$I = (2\pi)^p \Gamma(N + \lambda) \left[\prod_{k=1}^{p} \frac{1}{n_k!} \exp\{i(n_k + 2\lambda_k)\phi_k - i\pi\lambda_k\}\right] \times \left[\prod_{k=1}^{p} \left(\frac{\partial}{\partial t_k}\right)^{j'_k}_{t_k=1} t_k^{j'_k - 1}\right] \sum_{\{j_k\}} \left[\prod_{k=1}^{p} \binom{n_k}{j_k} \left(-\frac{1}{t_k}\right)^{j_k}\right] \times \left(1 - \sum_{k=1}^{p} \frac{1}{t_k} \exp\{2i\phi_k\}\right)^{-N-\lambda}. \tag{4.18}$$

The $\{j_k\}$ sums are now simply binomial expansions for integer powers, which upon summing gives

$$I = (2\pi)^p \Gamma(N + \lambda) \left[\prod_{k=1}^{p} \frac{1}{n_k!} \exp\{i(n_k + 2\lambda_k)\phi_k - i\pi\lambda_k\}\right] \times \left[\prod_{k=1}^{p} \left(\frac{\partial}{\partial t_k}\right)^{j'_k}_{t_k=1} t^{j'_k - 1} \left(1 - \frac{1}{t_k}\right)^{n_k}\right] \times \left(1 - \sum_{k=1}^{p} \frac{1}{t_k} \exp\{2i\phi_k\}\right)^{-N-\lambda} \tag{4.19}$$

The interpretation of this expression is the same as with the discrete family. The total order of differentiations is J' so for $N > J'$ at least one factor of $(1 - 1/t_k)$ for some k will survive in every term after the differentiations and then will vanish upon setting $t_k = 1$. Thus recalling the definition of I, we have shown

$$\int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p\, P_n(x) \times \left[\prod_{k=1}^{p} \frac{\Gamma(j'_k + \lambda_k + ix_k)}{\Gamma(\lambda_k + ix_k)}\right] w(x) = 0, \quad \text{if} \quad N > J', \tag{4.20}$$

and then referring back to (4.11) and (4.12), we immediately deduce that

$$\int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p\, P_n(x) P_{n'}(x) w(x) = 0, \quad \text{if } N \neq N', \tag{4.21}$$

and so $P_n(x)$ are orthogonal with respect to subspaces labelled by N, which is the continuous analog of (2.31).

Now we demonstrate the analogous result for the biorthogonal family $\overline{P}_n(x)$. As before, we expand $\overline{P}_{n'}(x)$ as

$$\overline{P}_{n'}(x) = \sum_{\{j'_k\}} \bar{b}(j'_1 j'_2, ..., j'_p) \prod_{k=1}^{p} \frac{\Gamma(j'_k + \lambda_k + ix_k)}{\Gamma(\lambda_k + ix_k)}, \tag{4.22}$$

where $\bar{b}(j'_1 j'_2, ..., j'_p)$ are some constants and the $\{j'_k\}$ sum is over the range $0 \leqslant J' \leqslant N'$. Then we consider the inner product

$$\overline{I} \equiv \int_{-\infty}^{\infty} dx_1 \cdots \int_{-\infty}^{\infty} dx_p\, \overline{P}_n(x) \times \left[\prod_{k=1}^{p} \frac{\Gamma(j'_k + \lambda_k + ix_k)}{\Gamma(\lambda_k + ix_k)}\right] w(x), \tag{4.23}$$

which upon substitution (4.2) and (4.3) becomes

$$\overline{I} = \sum_{\{j_k\}} \left[\prod_{k=1}^{p} \binom{n_k}{j_k} \frac{1}{n_k!} \exp\{i(n_k - 2j_k)\phi_k\}\right] \times \left(1 - \sum_{k=1}^{p} \exp\{2i\phi_k\}\right)^{J} \frac{\Gamma(N + \lambda)}{\Gamma(J + \lambda)} \int_{-\infty}^{\infty} dx_1 \cdots \times \int_{-\infty}^{\infty} dx_p \left[\prod_{k=1}^{p} \Gamma(j_k + \lambda_k + ix_k) \frac{\Gamma(j'_k + \lambda_k + ix_k)}{\Gamma(\lambda_k + ix_k)} \exp\{(2\phi_k - \pi)x_k\}\right] \Gamma(\lambda_{p+1} - iX). \tag{4.24}$$

Proceeding as before, we substitute (4.15), interchange the order of integrations with differentiations, and then substitute (4.10) for the multiple integral. This gives

$$\overline{I} = (2\pi)^p \Gamma(N + \lambda) \left[\prod_{k=1}^{p} \frac{1}{n_k!} \exp\{i(n_k + 2\lambda_k)\phi_k - i\pi\lambda_k\}\right] \times \left[\prod_{k=1}^{p} \left(\frac{\partial}{\partial t_k}\right)^{j'_k}_{t_k=1} t_k^{j'_k - 1}\right] \sum_{\{j_k\}} \left[\prod_{k=1}^{p} \binom{n_k}{j_k} (-\xi_k)^{j_k}\right] \times \left(1 - \sum_{k=1}^{p} \frac{1}{t_k} \exp\{2i\phi_k\}\right)^{-\lambda}, \tag{4.25}$$

where we have defined

$$\xi_k(t_1\cdots t_p;\phi_1\cdots\phi_p)\equiv\frac{1}{t_k}\Big(1-\sum_{k=1}^{p}\exp\{2i\phi_k\}\Big)\times\Big[\Big(1-\sum_{k=1}^{p}\frac{1}{t_k}\exp\{2i\phi_k\}\Big)\Big]^{-1}. \tag{4.26}$$

Again the $\{j_k\}$ sums are simply binomial expansions for integer powers, which upon summation gives

$$\begin{aligned}\overline{I}=&(2\pi)^p\Gamma(N+\lambda)\\&\times\Big[\prod_{k=1}^{p}\frac{1}{n_k!}\exp\{i(n_k+2\lambda_k)\phi_k-i\pi\lambda_k\}\Big]\\&\times\Big[\prod_{k=1}^{p}\Big(\frac{\partial}{\partial t_k}\Big)^{j'_k}_{t_k=1}t_k^{j'_k-1}(1-\xi_k)^{n_k}\Big]\\&\times\Big(1-\sum_{k=1}^{p}\frac{1}{t_k}\exp\{2i\phi_k\}\Big)^{-\lambda}.\end{aligned} \tag{4.27}$$

From (4.26) we see that upon setting $t_k=1$, $k=1,2,\ldots,p$, the parameters ξ_k, $k=1,2,\ldots,p$ become equal to unity. Then it is obvious that if any factor of $(1-\xi_k)$ survives the differentiations, it will vanish upon setting $t_k=1$, $k=1,2,\ldots,p$ so the argument follows as before. The total order of differentiations is J' so for $N>J'$ at least one factor $(1-\xi_k)$ for some k will survive in every term after the differentiations and then will vanish. Thus recalling the definition of $\overline{I}$, we have shown

$$\int_{-\infty}^{\infty}dx_1\cdots\int_{-\infty}^{\infty}dx_p\,\overline{P}_n(x)\times\Big[\prod_{k=1}^{p}\frac{\Gamma(j'_k+\lambda_k+ix_k)}{\Gamma(\lambda_k+ix_k)}\Big]w(x)=0,\quad\text{if } N>J', \tag{4.28}$$

then referring back to (4.22) we immediately deduce that

$$\int_{-\infty}^{\infty}dx_1\cdots\int_{-\infty}^{\infty}dx_p\,\overline{P}_n(x)\overline{P}_{n'}(x)w(x)=0,\quad\text{if } N\neq N', \tag{4.29}$$

and so $\overline{P}_n(x)$ are also orthogonal with respect to subspaces labelled by N, which is the continuous analog of (2.40).

Though the polynomial familes $P_n(x)$ and $\overline{P}_n(x)$ are each orthogonal among themselves with respect to degree N, they are not so for different polynomials of the same degree. In analogy with discrete case, we demonstrate that these two families form a biorthogonal system; consider the inner product

$$\int_{-\infty}^{\infty}dx_1\cdots\int_{-\infty}^{\infty}dx_p\,P_n(x)\overline{P}_{n'}(x)w(x). \tag{4.30}$$

If $N>N'$ we expand $\overline{P}_{n'}(x)$ as in (4.22) and then by (4.20) this inner product vanishes whereas if $N<N'$ we expand $P_n(x)$ as in (4.12) and by (4.28) it again vanishes. So this inner product is zero unless $N=N'$, which we now assume. Substituting (4.2) into (4.30) gives

$$\begin{aligned}&\sum_{\{j'_k\}}\Big[\prod_{k=1}^{p}\binom{n'_k}{j'_k}\frac{1}{n'_k!}\exp\{i(n'_k-2j'_k)\phi_k\}\Big]\frac{\Gamma(N'+\lambda)}{\Gamma(J'+\lambda)}\\&\times\Big(1-\sum_{k=1}^{p}\exp\{2i\phi_k\}\Big)^{J'}\int_{-\infty}^{\infty}dx_1\cdots\int_{-\infty}^{\infty}dx_p\,P_n(x)\\&\times\Big[\prod_{k=1}^{p}\frac{\Gamma(j'_k+\lambda_k+ix_k)}{\Gamma(\lambda_k+ix_k)}\Big]w(x),\end{aligned} \tag{4.31}$$

and then substituting (4.19) for the multiple integral yields

$$\begin{aligned}&(2\pi)^p\Gamma(N+\lambda)\Big[\prod_{k=1}^{p}\frac{1}{n_k!}\exp\{i(n_k+2\lambda_k)\phi_k-i\pi\lambda_k\}\Big]\\&\times\sum_{\{j'_k\}}\Big[\prod_{k=1}^{p}\binom{n'_k}{j'_k}\frac{1}{n'_k!}\exp\{i(n'_k-2j'_k)\phi_k\}\Big]\frac{\Gamma(N'+\lambda)}{\Gamma(J'+\lambda)}\\&\times\Big(1-\sum_{k=1}^{p}\exp\{2i\phi_k\}\Big)^{J'}\Big[\prod_{k=1}^{p}\Big(\frac{\partial}{\partial t_k}\Big)^{j'_k}_{t_k=1}t_k^{j'_k-1}\\&\times\Big(1-\frac{1}{t_k}\Big)^{n_k}\Big]\Big(1-\sum_{k=1}^{p}\frac{1}{t_k}\exp\{2i\phi_k\}\Big)^{-N-\lambda}.\end{aligned} \tag{4.32}$$

As we discussed several times, if the order of the differentiations J' is less than the degree of the factors $(1-1/t_k)$ which is N, then one obtains zero. So the only term in the $\{j'_k\}$ sum which does not vanish is that for which $J'=N'=N$, that is, for $j'_k=n'_k$, $k=1,2,\ldots,p$, and then the inner product simplifies to

$$\begin{aligned}&(2\pi)^p\Gamma(N+\lambda)\Big[\prod_{k=1}^{p}\frac{1}{n_k!n'_k!}\\&\times\exp\{i(n_k-n'_k+2\lambda_k)\phi_k-i\pi\lambda_k\}\Big]\\&\times\Big(1-\sum_{k=1}^{p}\exp\{2i\phi_k\}\Big)^{N'}\Big[\prod_{k=1}^{p}\Big(\frac{\partial}{\partial t_k}\Big)^{n'_k}_{t_k=1}\\&\times t_k^{n'_k}\Big(1-\frac{1}{t_k}\Big)^{n_k}\Big]\Big(1-\sum_{k=1}^{p}\frac{1}{t_k}\exp\{2i\phi_k\}\Big)^{-N-\lambda}.\end{aligned} \tag{4.33}$$

As with the discrete family, if $n'_k>n_k$ for some k, and recalling that $N=N'$, then $n'_{k'}<n_{k'}$ for some other k'. In that case a factor of $(1-1/t_{k'})$ will survive in every term after the differentiations and then will vanish upon setting $t_{k'}=1$. So the inner product (4.33) is zero unless $n'_k=n_k$ for every k, and in the nonvanishing case it simply becomes

$$h_n\equiv(2\pi)^p\Gamma(N+\lambda)\Big[\prod_{k=1}^{p}\frac{1}{n_k!}\exp\{i(2\phi_k-\pi)\lambda_k\}\Big]\times\Big(1-\sum_{k=1}^{p}\exp\{2i\phi_k\}\Big)^{-\lambda}. \tag{4.34}$$

Thus the inner product (4.30) is simply

$$\int_{-\infty}^{\infty}dx_1\cdots\int_{-\infty}^{\infty}dx_p\,P_n(x)\overline{P}_{n'}(x)w(x)=h_n\prod_{k=1}^{p}\delta_{n_kn'_k}, \tag{4.35}$$

which is the continuous analog of (2.46).

The Askey tableau shows that the single variable Meixner–Pollaczek polynomials in the special case when $\phi=\pi/2$ can be obtained as a limit of the continuous Hahn polynomials. An analogous limit exists for the multivariable biorthogonal families and provides an independent check for the results of this section. In the multivariable continuous Hahn[7,8] polynomials $C_n(x)$ and $\overline{C}_n(x)$ set $a_k = \lambda_k, b_k = \gamma \exp\{i(2\phi_k - \pi)\}$ $k = 1,2,...,p,$ $c = \gamma,$ $d = \lambda_{p+1}$, and then take the limit $\gamma \to \infty$. One finds

$$\lim_{\gamma\to\infty} \gamma^{-N} C_n(x) = (-i)^N \left[\prod_{k=1}^{p} \exp\{in_k\phi_k\} \right] P_n(x),$$

$$\lim_{\gamma\to\infty} \gamma^{-N} \overline{C}_n(x) = (-i)^N \left[\prod_{k=1}^{p} \exp\{in_k\phi_k\} \right] \overline{P}_n(x), \tag{4.36}$$

while the continuous Hahn weight function $w^h(x)$ limits to the Meixner–Pollaczek weight

$$\lim_{\gamma\to\infty} \left[\Gamma(c) \prod_{k=1}^{p} \Gamma(b_k) \right]^{-1} w^h(x)$$
$$= \left[\prod_{k=1}^{p} \exp\{(2\phi_k - \pi)x_k\} \Gamma(\lambda_k + ix_k) \right]$$
$$\times \Gamma(\lambda_{p+1} - iX). \tag{4.37}$$

The orthogonality and biorthogonality relations (4.21), (4.29), and (4.35) also follow in the limit at least for values of $\phi_1,\phi_2,...,\phi_p$ satisfying $\mathrm{Re}(b_1,b_2,...b_p) > 0$ (which includes $\phi_1 = \phi_2 = \cdots = \phi_p = \pi/2$). Angle parameters not fulfilling this condition require separate consideration since then the integration contours must be deformed during the limit.

ACKNOWLEDGMENTS

I wish to thank the referee for many useful comments.

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

[1]J. Meixner, J. Lon. Math. Soc. **9**, 6 (1934).
[2]A. Erdelyi, A. Magnus, F. Oberhettinger, and F. Tricomi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1955), Vols. I–III.
[3]T. S. Chihara, *An Introduction to Orthogonal Polynomials* (Gordon and Breach, New York, 1978).
[4]R. Askey and J. A. Wilson, Memoirs Am. Math. Soc. **319**, 1–55 (1985); R. Askey, J. Phys. A **18**, L1017 (1985).
[5]M. Krawtchouk, C. R. Acad. Sci. Paris **189**, 620 (1929).
[6]C. S. Lam and M. V. Tratnik, Can. J. Phys. **63**, 1427 (1985).
[7]M. V. Tratnik, J. Math. Phys. **29**, 1529 (1988).
[8]M. V. Tratnik, J. Math. Phys. **30**, 627 (1989).
[9]H. M. Srivastava and P. W. Karlsson, *Multiple Gaussian Hypergeometric Series* (Horwood, New York, 1985), p. 38.
[10]E. T. Whittaker and G. N. Watson, *A Course of Modern Analysis* (Cambridge, U.P., London, 1950).

Algebras connected with the Z_n elliptic solution of the Yang–Baxter equation

Hou Bo-yu
Institute of Modern Physics, P.O. Box 105, Northwest University, Xian 710069, People's Republic of China
Wei Hua
Center of Theoretical Physics, CCAST (World Laboratory), Institute of Modern Physics, P.O. Box 105, Northwest University, Xian 710069, People's Republic of China

(Received 12 May 1989; accepted for publication 2 August 1989)

The quantum and classical algebras connected with the Z_n-symmetric elliptic solution of the Yang–Baxter equation are derived; their structure constants and the relations between the quantum algebra and the classical one are investigated in detail. Moreover, the trigonometric limit of these algebras is worked out.

I. INTRODUCTION

In recent years the Yang–Baxter equation (YBE)[1] and its exact solutions have been studied fruitfully.[2–6] This investigation is extended and associated with quantum algebra (quantum group),[7–9] conformal field theories, completely integrable models, and braid groups etc.[10–15] The solutions of YBE have been classified as rational, trigonometric, and elliptic cases, including their high-spin (fusion) representations.[5,6] The quantum groups for rational and trigonometric cases have been well studied.[7] For the elliptic case, Sklyanin investigated the quantum and classical algebras connected with the eight-vertex model in 1982.[8] In a short article on the representation,[9] Cherednik wrote down an expression for the quantum algebra of Belavin's Z_n-symmetric model in 1985. Since the elliptic case is related naturally to the Kac–Moody–Virasoro characters, and can be generated easily onto a high-genus Riemann surface, and its degeneration gives the trigonometric case, hence the elliptic quantum algebra is more interesting.

In Sec. II we first reduce the Yang–Baxter relation for the Z_n-symmetric elliptic solution (Z_n-SES) to a spectroparameter independent form by means of the Heisenberg group, and we give various explicit expressions for the structure constants of the algebras and their symmetric relations. In Sec III we deduce the quantum algebra. In Sec. IV we treat the classical YBE and find the classical algebra. In Sec. V we exhibit the correspondences between the quantum quantities as well as the equations and the classical ones. In Sec. VI we study the corresponding trigomometric algebras.

II. THE YANG–BAXTER RELATION

The Boltzmann weight in 2D statistical mechanics can be written as

$$R_{jk}(u)=\sum_{\alpha\in Z_n^2} W_\alpha(u) I_\alpha^{(j)} I_\alpha^{\dagger(k)}, \tag{1}$$

where $I_\alpha^{(j)}$ acts on the subspace of the jth site, $I_\alpha=h^{\alpha_1}g^{\alpha_2}$, h and g are $n\times n$ matrices with elements

$$h_{jk}=\delta_{j(\text{mod }n)}^{k+1},\quad g_{jk}=\omega^k\delta_{jk}, \tag{2}$$

and $\omega=e^{2\pi i/n}$, $\alpha=(\alpha_1,\alpha_2)$, $\alpha_j=0,1,\ldots,n-1$, $W_\alpha(u)$ is the Boltzmann coordinate. The YBE for the Boltzmann weights is

$$R_{12}(u-v)R_{13}(u)R_{23}(v)=R_{23}(v)R_{13}(u)R_{12}(u-v). \tag{3}$$

By means of

$$I_\alpha^{-1}=I_\alpha^\dagger=\omega^{\alpha_1\alpha_2}I_{-\alpha},\quad I_\alpha I_\beta=\omega^{\alpha_2\beta_1}I_{\alpha+\beta},$$
$$\operatorname{tr}(I_\alpha I_\beta^\dagger)=n\delta_\alpha^\beta,\quad \operatorname{tr}(A^{(j)}B^{(k)})=\operatorname{tr}A^{(j)}\operatorname{tr}B^{(k)}, \tag{4}$$

(3) is reduced to an equivalent YBE for the Boltzmann coordinates

$$\sum_{\gamma\in Z_n^2}(\omega^{\langle\gamma,\alpha\rangle}-\omega^{\langle\alpha,\gamma-\beta\rangle})\times W_{\gamma-\alpha}(u-v)W_{\alpha+\beta-\gamma}(u)\,W_\gamma(v)=0, \tag{5}$$

where $\langle a,\beta\rangle=a_1\beta_2-a_2\beta_1$. Taking $\gamma\to a+\beta-\gamma$ in the latter term, (5) turns to

$$\sum{}' \omega^{\langle\gamma,\alpha\rangle}W_{\alpha\beta\gamma}(u,v)=0, \tag{6}$$

where the summation Σ' means the restriction $\langle a, 2\gamma-\beta\rangle\neq 0$, and

$$W_{\alpha\beta\gamma}(u,v)=W_{\gamma-\alpha}(u-v)W_{\alpha+\beta-\gamma}(u)W_\gamma(v)-W_{\beta-\gamma}(u-v)W_\gamma(u)W_{\alpha+\beta-\gamma}(v). \tag{7}$$

For the Z_n-SES (Ref. 3) we have

$$W_\alpha(u)=\frac{\sigma_\alpha(u+\eta)}{\sigma_\alpha(\eta)}\frac{\sigma_0(\eta)}{\sigma_0(u+\eta)}, \tag{8}$$

with the Jacob theta function of rational characteristics $(1/2+\alpha_1/n,\ 1/2+\alpha_2/n)$,

$$\sigma_\alpha(u)=\sigma_\alpha(u,\tau)=\sum_{m=-\infty}^{\infty}\exp\left\{i\pi\tau\left(m+\frac{1}{2}+\frac{\alpha_1}{n}\right)+i2\pi\left(m+\frac{1}{2}+\frac{\alpha_1}{n}\right)\left(u+\frac{1}{2}+\frac{\alpha_2}{n}\right)\right\}.$$

Now we will construct a spectroparameter independent quantity $C_{\alpha\beta\gamma}$ by dividing (7) with a γ independent factor. Let

$$C_{\alpha\beta\gamma}(u,v)=\frac{\sigma_\beta(2\eta)}{\sigma_0(u-v)\sigma_\beta(u+2\eta)\sigma_\alpha(v)}\left[\frac{\sigma_{\gamma-\alpha}(u-v+\eta)}{\sigma_{\gamma-\alpha}(\eta)}\frac{\sigma_{\alpha+\beta-\gamma}(u+\eta)}{\sigma_{\alpha+\beta-\gamma}(\eta)}\frac{\sigma_\gamma(v+\eta)}{\sigma_\gamma(\eta)}-\frac{\sigma_{\beta-\gamma}(u-v+\eta)}{\sigma_{\beta-\gamma}(\eta)}\frac{\sigma_\gamma(u+\eta)}{\sigma_\gamma(\eta)}\frac{\sigma_{\alpha+\beta-\gamma}(v+\eta)}{\sigma_{\alpha+\beta-\gamma}(\eta)}\right]. \tag{9}$$

By means of the Heisenberg group, we have

$$C_{\alpha\beta\gamma}(u+1,v)=C_{\alpha\beta\gamma}(u+\tau,v)=C_{\alpha\beta\gamma}(u,v).$$

Hence $C_{\alpha\beta\gamma}(u,v)$ is a doubly periodic function with respect to (wrt) u and it has at most two poles on the lattice generated by 1 and τ. On account of the zero of its denominator, $u=v$, cancelling out one zero of its numerator, we confirm that $C_{\alpha\beta\gamma}(u,v)$ is an entire function wrt u, hence it is independent of u. A similar analysis for the spectroparameter v shows that $C_{\alpha\beta\gamma}(u,v)$ is also independent of v. We denote it as $C_{\alpha\beta\gamma}(\eta,\tau)$. Based on the spectroparameter independence of $C_{\alpha\beta\gamma}(\eta,\tau)$, we may get its various expressions by substituting appropriate values of u,v into (9). Taking $u=0$, $v=-\eta-(\gamma_1\tau+\gamma_2)/n$, we obtain

$$C_{\alpha\beta\gamma}(\eta,\tau)=\frac{\omega^{\gamma_1-\alpha_1}\sigma_{\alpha+\beta-2\gamma}(0)\sigma_\beta(2\eta)}{\sigma_{\gamma-\alpha}(\eta)\sigma_{\alpha+\beta-\gamma}(\eta)\sigma_\gamma(\eta)\sigma_{\beta-\gamma}(\eta)}. \tag{10}$$

We see that $C_{\alpha\beta\gamma}(\eta,\tau)$ is a four-order elliptic function of η with periods 1 and τ. Now we rewrite (6) as

$$\sum_{\langle\alpha,2\gamma-\beta\rangle\neq 0}\omega^{\langle\gamma,\alpha\rangle}C_{\alpha\beta\gamma}(\eta,\tau)=\sum_{\gamma\in Z_n^2}\omega^{\langle\gamma,\alpha\rangle}C_{\alpha\beta\gamma}(\eta,\tau)=0. \tag{11}$$

However, we should prove the validity for the latter without the summation-restriction equations. Denote

$$f(\eta)=\sum_{\gamma\in Z_n^2}\omega^{\langle\gamma,\alpha\rangle}C_{\alpha\beta\gamma}(\eta,\tau),$$

with $C_{\alpha\beta\gamma}$ as in (10). It is obvious that $f(\eta)/\sigma_\beta(2\eta)$ is holomorphic for $\eta\neq(\delta_1\tau+\delta_2)/n$, $\delta\in Z_n^2$. Using (10) we evaluate $f(\eta)$ at $\eta=\epsilon+(\delta_1\tau+\delta_2)/n$. For $\epsilon\to 0$ there are four singular terms, $\gamma=\alpha-\delta,\alpha+\beta+\delta,-\delta,$ and $\beta+\delta$, However, the singular parts cancel exactly:

$$f(\epsilon+(\delta_1+\delta_2)/n)=\sigma_\beta(2\eta)\cdot(\text{regular terms at } \varepsilon=0).$$

Hence $f(\eta)/\sigma_\beta(2\eta)$ is an entire function wrt η. As an entire doubly periodic function, $f(\eta)$ is independent of η. Taking $2\eta_0=-(\beta_1\tau+\beta_2)/n$ we get $f(\eta_0)=0$. This proves $f(\eta)=0$. This also gives another explicit proof of Belavin's ansatz[3] as a solution of YBE.

We shall see in the next section that $C_{\alpha\beta\gamma}$'s are quantum algebra structure constants (QSC's). In order to be more convenient for expressing the symmetries of the QSC's and the relations between the QSC's and the classical algebra structure constants (CSC's) we introduce a modified QSC,

$$F_{\alpha\beta\gamma}(\eta,\tau)=\sigma_0'(0)\sigma_\alpha(0)C_{\alpha\beta\gamma}(\eta,\tau),\quad \alpha\neq 0,$$
$$F_{0\beta\gamma}(\eta,\tau)=\sigma_0'^2(0)C_{0\beta\gamma}(\eta,\tau). \tag{12}$$

Equation (10) gives an explicit expression for $F_{\alpha\beta\gamma}$ or $F_{0\beta\gamma}$ Two other expressions come from (9) by setting $v=0$, $u\to 0$ or $u=0$, $v\to 0$:

$$\begin{aligned}F_{\alpha\beta\gamma}(\eta,\tau)&=\frac{\partial}{\partial\eta}\ln\frac{\sigma_{\gamma-\alpha}(\eta)\sigma_{\alpha+\beta-\gamma}(\eta)}{\sigma_{\beta-\gamma}(\eta)\sigma_\gamma(\eta)}\\&=\zeta\left(\eta+\frac{\gamma_1-\alpha_1}{n}\tau+\frac{\gamma_2-\alpha_2}{n}\right)+\zeta\left(\eta+\frac{\alpha_1+\beta_1-\gamma_1}{n}\tau+\frac{\alpha_2+\beta_2-\gamma_2}{n}\right)-\zeta\left(\eta+\frac{\beta_1-\gamma_1}{n}\tau+\frac{\beta_2-\gamma_2}{n}\right)\\&\quad-\zeta\left(\eta+\frac{\gamma_1}{n}\tau+\frac{\gamma_2}{n}\right),\quad\alpha\neq 0,\\F_{0\beta\gamma}(\eta,\tau)&=\frac{\partial^2}{\partial\eta^2}\ln\frac{\sigma_{\beta-\gamma}(\eta)}{\sigma_\gamma(\eta)}\\&=\mathscr{P}\left(\eta+\frac{\gamma_1\tau+\gamma_2}{n}\right)-\mathscr{P}\left(\eta+\frac{\beta_1-\gamma_1}{n}\tau+\frac{\beta_2-\gamma_2}{n}\right),\end{aligned} \tag{13}$$

where $\zeta(x)$ and $\mathscr{P}(x)$ are the Weierstrass zeta function and elliptic function, respectively

Using (10), (13), and

$$\sigma_{\alpha_1+np,\,\alpha_2+nq}(u)=e^{2i\pi(1/2+\alpha_1/n)q}\sigma_\alpha(u),\quad p,q\in Z, \tag{14}$$

we obtain the following relations:

$$F_{\alpha\beta\gamma}=-F_{-\alpha,\beta,\gamma-\alpha}=F_{2\gamma-\alpha-\beta,\beta,\gamma},\quad\alpha\neq 0, \tag{15}$$

$$=-F_{\alpha,\beta,\alpha+\beta-\gamma}, \tag{16}$$

$$F_{0,\alpha+\beta,\beta}+F_{0,\beta+\gamma,\gamma}+\cdots+F_{0,\delta+\rho,\rho}+F_{0,\rho+\alpha,\alpha}=0, \tag{17}$$

$$\sum_{\gamma\in Z_n^2}F_{0\beta\gamma}=0,$$

$$\sum_{\gamma\in Z_n^2}F_{\alpha\beta\gamma}=\sum_{\beta\in Z_n^2}F_{\alpha\beta\gamma}=0,\quad\alpha\neq 0. \tag{18}$$

Now (11) is rewritten as

$$\sum_{\langle\alpha,2\gamma-\beta\rangle\neq 0}\omega^{\langle\gamma,\alpha\rangle}F_{\alpha\beta\gamma}(\eta,\tau)=\sum_{\gamma\in Z_n^2}\omega^{\langle\gamma,\alpha\rangle}F_{\alpha\beta\gamma}(\eta\tau)=0. \tag{19}$$

III. THE QUATNUM ALGEBRA

The operator representation of the YBE,

$$L_j(u)=\sum_{\alpha\in Z_n^2}W_\alpha(u)I_\alpha^{(j)}S_\alpha, \tag{20}$$

satisfies

$$R_{12}(u-v)L_1(u)L_2(v)=L_2(v)L_1(u)R_{12}(u-v). \tag{21}$$

Using (4), (21) is equivalent to

$$\sum_{\gamma\in Z_n^2}W_{\alpha\beta\gamma}(u,v)\,\omega^{(\beta_1-\gamma_1)(\gamma_2-\alpha_2)}S_{\alpha+\beta-\gamma}S_\gamma=0, \tag{22}$$

where $S_\alpha = S_{\alpha(\mathrm{mod}\, n)}$. For the Z_n-SES (8) using (9), (22) is reduced to a spectroparameter independent form

$$\sum_{\gamma\in Z_n^2} F_{\alpha\beta\gamma}(\eta,\tau)\, \omega^{(\beta_1-\gamma_1)(\gamma_2-\alpha_2)} S_{\alpha+\beta-\gamma} S_\gamma = 0, \quad (23)$$

This is an algebra for the quantum operators S's, and (11) is its compatibility conditions. Define a normalized QSC,

$$F^1_{\alpha\beta\gamma} = F_{\alpha\beta\gamma}/F_{\alpha\beta\alpha}. \quad (24)$$

From (10) we have

$$F^1_{\alpha\beta\gamma}(\eta,\tau) = \omega^{\gamma_1-\alpha_1} \times \frac{\sigma_{\alpha+\beta-2\gamma}(0)\sigma_0(\eta)\sigma_\beta(\eta)\sigma_\alpha(\eta)\sigma_{\beta-\alpha}(\eta)}{\sigma_{\beta-\alpha}(0)\sigma_{\gamma-\alpha}(\eta)\sigma_{\alpha+\beta-\gamma}(\eta)\sigma_\gamma(\eta)\sigma_{\beta-\gamma}(\eta)}. \quad (25)$$

On account of $F_{\alpha\beta\beta} = -F_{\alpha\beta\alpha}$, we rewrite (23) as

$$[S_\alpha, S_\beta] = \sum_{\gamma\neq\alpha,\beta} F^1_{\alpha\beta\gamma}(\eta,\tau) \times \omega^{(\beta_1-\gamma_1)(\gamma_2-\alpha_2)} S_{\alpha+\beta-\gamma} S_\gamma, \quad \alpha\neq\beta, \quad (26a)$$

$$\sum_{\gamma\in Z_n^2} F_{\alpha,\alpha,\alpha+\gamma}(\eta,\tau)\omega^{-\gamma_1\gamma_2}[S_{\alpha-\gamma}, S_{\alpha+\gamma}] = 0 \quad (n\geqslant 3). \quad (26b)$$

For $n=2$, we have

$$\begin{aligned}
&F^1_{10,11,01} = F^1_{11,01,10} = F^1_{01,10,11} = 1,\\
&F^1_{10,11,00} = F^1_{11,01,00} = F^1_{01,10,00} = -1,\\
&F^1_{0,01,10} = -F^1_{0,01,11} = -\frac{\vartheta_1^2(\eta)\vartheta_2^2(\eta)}{\vartheta_3^2(\eta)\vartheta_4^2(\eta)} \equiv J_{12},\\
&F^1_{0,10,11} = -F^1_{0,10,01} = -\frac{\vartheta_1^2(\eta)\vartheta_4^2(\eta)}{\vartheta_3^2(\eta)\vartheta_2^2(\eta)} \equiv J_{23},\\
&F^1_{0,11,01} = -F^1_{0,11,10} = \frac{\vartheta_1^2(\eta)\vartheta_3^2(\eta)}{\vartheta_2^2(\eta)\vartheta_4^2(\eta)} \equiv J_{31}.
\end{aligned} \quad (27)$$

Note that $I_{10}=\sigma_1$, $I_{11}=-i\sigma_2$, $I_{01}=\sigma_3$, $I_0=1$, and write $S_1=S_{10}$, $iS_2=S_{11}$, $S_3=S_{01}$, $S_0=S_{00}$, we obtain the result of Ref. 8,

$$[S_0, S_a] = iJ_{bc}[S_b, S_c]_+, \quad (28)$$

$$[S_a, S_b] = i[S_c, S_0]_+, \quad \text{abc: cycle of 123}. \quad (29)$$

Now let us see the simplest case $\eta\to 0$ for (26). We get $F^1_{\alpha\beta\gamma}(0,\tau) = 1$ for $\gamma=\alpha+\beta$, -1 for $\gamma=0$; 0 otherwise. Then (26a) turns to

$$[S_0, S_\alpha] = 0,$$

$$[S_\alpha, S_\beta] = (\omega^{-\alpha_1\beta_2} - \omega^{-\alpha_2\beta_1}) S_0 S_{\alpha+\beta}. \quad (30)$$

By means of (30), (16), and (19), the left-hand side of (26b),

$$\begin{aligned}
&\sum_{\gamma\in Z_n^2} F_{\alpha,\alpha,\alpha+\gamma}(0,\tau)\omega^{-\gamma_1\gamma_2}[S_{\alpha-\gamma}, S_{\alpha+\gamma}]\\
&= \omega^{-\alpha_1\alpha_2} \sum_{\gamma\in Z_n^2} F_{\alpha,\alpha,\alpha+\gamma}(0,\tau)(\omega^{\langle\gamma,\alpha\rangle} - \omega^{\langle\alpha,\gamma\rangle}) S_0 S_{2\alpha}\\
&= 2\omega^{-\alpha_1\alpha_2} \sum_{\gamma\in Z_n^2} F_{\alpha,\alpha,\gamma}(0,\tau)\omega^{\langle\gamma,\alpha\rangle} S_0 S_{2\alpha},
\end{aligned}$$

vanishes, hence (26b) does not give more relations except (26a). We conclude that the algebra with $\eta=0$ is equivalent tp (30) and in another basis it appears as the algebra $u(n)$. Moreover, in this case we have a matrix realization $S_\alpha \cong I^\dagger_\alpha$ for it.

IV. THE CLASSICAL YBE AND THE CLASSICAL ALGEBRA

The classical YBE corresponding to the YBE (3) is

$$[r_{12}(u-v), r_{13}(u) + r_{23}(v)] + [r_{13}(u), r_{23}(v)] = 0, \quad (31)$$

with

$$r_{jk}(u) = \sum_{\alpha\in Z_n^2} w_\alpha(u) I^{(j)}_\alpha I^{\dagger(k)}_\alpha, \quad (32)$$

$$w_\alpha(u) = \frac{\partial}{\partial\eta} W_\alpha(u)|_{\eta=0}, \quad w_0 = 0. \quad (33)$$

By means of (4), (31) is reduced to

$$(\omega^{\langle\alpha,\beta\rangle} - 1)[w_{-\alpha}(u-v)w_{\alpha+\beta}(u) - w_\beta(u-v) \times w_{\alpha+\beta}(v) + w_\beta(u)w_\alpha(v)] = 0, \quad \alpha,\beta\in Z_n^2. \quad (34)$$

The classical representation

$$\mathscr{L}_j(u) = s_0 + i\sum_{\alpha\neq 0} w_\alpha(u) I^{(j)}_\alpha s_\alpha \quad (35)$$

satisfies the fundamental Poisson bracket relations (FPR)

$$\{\mathscr{L}_1(u), \mathscr{L}_2(v)\} = [r_{12}(u-v), \mathscr{L}_1(u)\mathscr{L}_2(v)]. \quad (36)$$

Using (4), the FPR is reduced to

$$\{s_\alpha, s_\beta\} = \sum_{\gamma\in Z_n^2} \frac{w_\gamma(u-v)\bar{w}_{\alpha-\gamma}(u)\bar{w}_{\beta+\gamma}(v)}{\bar{w}_\alpha(u)\bar{w}_\beta(v)} \times [\omega^{(\alpha_1-\beta_1-\gamma_1)\gamma_2} - \omega^{(\alpha_2-\beta_2-\gamma_2)\gamma_1}] \times s_{\alpha-\gamma} s_{\beta+\gamma}, \quad (37)$$

with $\bar{w}_0 = 1$, $\bar{w}_\alpha = iw_\alpha$ $(\alpha\neq 0)$. Letting $\alpha=\beta'$, $\beta=\alpha'$; $\gamma=\gamma'-\alpha'$ for the former term and $\gamma=\beta'-\gamma'$ for the latter term on the right-hand side of (37) we get

$$\{s_\alpha, s_\beta\} = -\sum_{\langle\beta-\gamma,\gamma-\alpha\rangle\neq 0} f_{\alpha\beta\gamma}(u,v)\omega^{(\beta_1-\gamma_1)(\gamma_2-\alpha_2)} \times s_{\alpha+\beta-\gamma} s_\gamma, \quad \alpha\neq\beta, \quad (38)$$

where

$$f_{\alpha\beta\gamma}(u,v) = [\bar{w}_\beta(u)\bar{w}_\alpha(v)]^{-1} \times [w_{\gamma-\alpha}(u-v)\bar{w}_{\alpha+\beta-\gamma}(u)\bar{w}_\gamma(v) - w_{\beta-\gamma}(u-v)\bar{w}_\gamma(u)\bar{w}_{\alpha+\beta-\gamma}(v)]. \quad (39)$$

For the Z_n-SES (8) we have

$$w_\alpha(u) = \frac{\sigma_0'(0)}{\sigma_0(u)} \frac{\sigma_\alpha(u)}{\sigma_\alpha(0)}, \quad \alpha\neq 0, \quad w_0 = 0. \quad (40)$$

Taking a similar analysis as in Sec. II, we conclude that $f_{\alpha\beta\gamma}(u,v)$ is independent of both the spectroparameters u and v. Setting $v=0$, $u=-(\gamma_1\tau+\gamma_2)/n$ for $f_{\alpha\beta\gamma}$ and $f_{\alpha 0\gamma}$; $u=0$, $v=(\gamma_1\tau+\gamma_2)/n$ for $f_{0\beta\gamma}$; $v=0$ for $f_{\alpha\beta 0}$, we obtain the CSC's expressed in terms of the QSC's:

$$f_{\alpha\beta\gamma}(\tau)=F_{\alpha\beta\gamma}(0,\tau),\quad \alpha,\beta,\gamma,\alpha-\gamma,\beta-\gamma,\alpha+\beta-\gamma\neq 0,$$
$$f_{0\beta\gamma}(\tau)=f_{\beta 0\gamma}(\tau)=iF_{0\beta\gamma}(0,\tau),\quad \beta,\gamma,\beta-\gamma\neq 0,$$
$$f_{\alpha\beta 0}=-f_{\alpha,\beta,\alpha+\beta}=i,\quad \alpha,\beta,\alpha+\beta\neq 0. \tag{41}$$

By means of (10), (13), and (41) we get, besides (15)–(18), the following relations for the CSC's:

$$f_{\alpha\beta\gamma}=f_{\beta\alpha\gamma},\quad \gamma\neq\alpha,\beta,\quad \alpha+\beta,0, \tag{42}$$
$$=f_{-\alpha,-\beta,-\gamma},\quad \alpha,\beta,\gamma,\alpha-\gamma,\beta-\gamma,\alpha+\beta-\gamma\neq 0, \tag{43}$$
$$f_{0\beta\gamma}=f_{0,-\beta,-\gamma},\quad \beta,\gamma,\beta-\gamma\neq 0. \tag{44}$$

Moreover, the summation restriction in (38) can be weakened to $\gamma\neq\alpha,\beta$. To see this, using (16) we combine every pair terms of $\gamma=\gamma'$ and $\gamma=\alpha+\beta-\gamma'$ in (38) and denote the summation of all the pairs as Σ'',

$$\{s_\alpha,s_\beta\}=-\sum{}'' f_{\alpha\beta\gamma}(\tau)\,[\omega^{(\beta_1-\gamma_1)(\gamma_2-\alpha_2)}-\omega^{(\beta_2-\gamma_2)(\gamma_1-\alpha_1)}]\,s_{\alpha+\beta-\gamma}s_\gamma;$$

we see that the terms satisfying $\langle\beta-\gamma,\gamma-\alpha\rangle=0$ cancel. Therefore we write the classical algebra as

$$\begin{aligned}\{s_\alpha,s_\beta\}&=-\sum_{\langle\alpha-\gamma,\beta-\gamma\rangle\neq 0} f_{\alpha\beta\gamma}(\tau)\,\omega^{(\beta_1-\gamma_1)(\gamma_2-\alpha_2)}\times s_{\alpha+\beta-\gamma}s_\gamma,\\&=-\sum_{\gamma\neq\alpha,\beta} f_{\alpha\beta\gamma}(\tau)\,\omega^{(\beta_1-\gamma_1)(\gamma_2-\alpha_2)}\times s_{\alpha+\beta-\gamma}s_\gamma,\end{aligned} \tag{45}$$

For the $n=2$ case we have

$$f_{0,01,10}=-f_{0,01,11}=-i\pi^2\vartheta_2^4(0,\tau)\equiv ij_{12},$$
$$f_{0,10,11}=-f_{0,10,01}=-i\pi^2\vartheta_4^4(0,\tau)\equiv ij_{23},$$
$$f_{0,11,01}=-f_{0,11,10}=i\pi^2\vartheta_3^4(0,\tau)\equiv ij_{31},$$
$$f_{\alpha\beta 0}=-f_{\alpha,\beta,\alpha+\beta}=i,\quad \alpha,\beta\neq 0. \tag{46}$$

The summation in (45) is only for two terms. Denoting $s_{11}=is_2$, $s_{10}=s_1$, $s_{01}=s_3$ we get

$$\{s_0,s_a\}=-2j_{bc}s_bs_c, \tag{47}$$
$$\{s_a,s_b\}=-2s_cs_0,\quad abc\text{:cycle of }123. \tag{48}$$

V. RELATIONS BETWEEN THE QUANTUM AND CLASSICAL CASES

Based on the quantization correspondences

$$[A,B]\sim -i\hbar\{A,B\},\quad S_0\sim\hbar s_0,\quad S_\alpha\sim s_\alpha\quad(\alpha\neq 0),$$

where the Poisson bracket $\{A,B\}=\Sigma_i(\partial A/\partial p_i\partial B/\partial q_i-\partial A/\partial q_i\partial B/\partial p_i)$ and (33), $w_\alpha(u)=(\partial/\partial\eta)W(u)|_{\eta=0}$, all the classical quantities and equations may be obtained by taking first $\hbar$ terms of $\hbar$-expansions for the corresponding quantum ones. Suppressing the higher-$\hbar$ terms, the relations are

$$R=1+i\hbar r,\quad W_\alpha=\delta_\alpha^0+i\hbar\,w_\alpha,$$
$$F^1_{\alpha\beta\gamma}(i\hbar,\tau)=i\hbar f_{\alpha\beta\gamma}(\tau),\quad \alpha,\beta\neq 0,$$
$$F^1_{0\beta\gamma}(i\hbar,\tau)=i\hbar^2 f_{0\beta\gamma}(\tau),\quad \beta,\beta-\gamma\neq 0,$$
$$\text{YBE }(3)=-\hbar^2\text{ CYBE }(31),\quad (5)=-\hbar^2\,(34),$$
$$L\sim\hbar\mathscr{L},\quad (21)\sim -i\hbar^3(36),$$
$$(26\text{a})\sim -i\hbar\,(45),\quad \alpha\neq 0;\quad -i\hbar^2\,(45),\quad \alpha\neq 0;$$
$$(28)\sim -i\hbar^2(47),$$
$$(29)\sim -i\hbar^2\,(48),\quad J_{bc}=\hbar^2 j_{bc}.$$

VI. LIMITING TO THE TRIGONOMETRIC CASE

When the elliptic parameter $\tau\to i\infty$, the Z_n-SES turns to a Z_n-symmetric trigonometric solution. Correspondingly the algebras turn to trigonometric ones. The equations for the trigonometric algebras are the same as (23), (26), and (45). We exhibit the structure constants below. For compactness we use the notations

$$(u\quad p)=\sin\pi\Big(u+\frac{p}{n}\Big),\quad \begin{pmatrix}u & l\\ v & m\end{pmatrix}=\frac{(u\quad l)}{(v\quad m)},$$
$$\epsilon=\epsilon(\beta_1-\alpha_1)=\begin{cases}1,\beta_1>\alpha_1,\\ -1,\beta_1<\alpha_1,\end{cases}$$
$$\theta_+=\theta(\beta_1-\alpha_1)=\begin{cases}1,\beta_1>\alpha_1,\\ 0,\beta_1<\alpha_1,\end{cases}$$
$$F_{\alpha\beta\gamma}=F_{\alpha\beta\gamma}(\eta,i\infty),\quad \omega=e^{2\pi i/n},\quad q(n)=q(\operatorname{mod} n).$$

On account of $S_\alpha=S_{\alpha(\operatorname{mod} n)}$ in (23) we restrict below $0\leqslant\alpha_i,\beta_i,\gamma_i\leqslant n-1$; $F^1_{\alpha\beta\gamma}$ for $\alpha\neq\beta$, $\gamma\neq\alpha,\beta$ only; $\alpha_1\neq\beta_1$ when ϵ or θ_+ appears; $\alpha_1=\beta_1$ when $F^{1'}_{\alpha\beta\gamma}$ is used; $\alpha\neq(0,0)$ in 1- and 2-items. The nonvanishing QSC's are

1-0, $0<\gamma_1<\alpha_1,\beta_1$: $F_{\alpha\beta\gamma}=2\pi i$, $F^1_{\alpha\beta\gamma}=2ie^{-i\pi\eta\epsilon}(\eta\quad 0)$, $F^{1'}_{\alpha\beta\gamma}=2i(\eta\quad 0)\begin{pmatrix}\eta & \beta_2-\alpha_2\\ 0 & \beta_2-\alpha_2\end{pmatrix}$;

1-1, $0<\alpha_1,\beta_1<\gamma_1<\alpha_1+\beta_1$: $F_{\alpha\beta\gamma}=-2\pi i$, $F^1_{\alpha\beta\gamma}=-2ie^{-i\pi\eta\epsilon}(\eta\quad 0)$, $F^{1'}_{\alpha\beta\gamma}=-2i(\eta\quad 0)\begin{pmatrix}\eta & \beta_2-\alpha_2\\ 0 & \beta_2-\alpha_2\end{pmatrix}$;

1-2, $\begin{cases}\gamma_1=0,\\ \alpha_1,\beta_1,\alpha_1+\beta_1\neq 0(n):\end{cases}$ $F_{\alpha\beta\gamma}=-\omega^{-\gamma_2/2}e^{-i\pi\eta}\dfrac{\pi}{(\eta\quad\gamma_2)}$,

$$F^1_{\alpha\beta\gamma}=-\omega^{-\gamma_2/2}e^{-2i\pi\eta\theta_+}\begin{pmatrix}\eta & 0\\ \eta & \gamma_2\end{pmatrix},$$
$$F^{1'}_{\alpha\beta\gamma}=-\omega^{-\gamma_2/2}e^{-i\pi\eta}\begin{pmatrix}\eta & 0\\ \eta & \gamma_2\end{pmatrix}\begin{pmatrix}\eta & \beta_2-\alpha_2\\ 0 & \beta_2-\alpha_2\end{pmatrix};$$

1-3 $\begin{cases}\gamma_1=\alpha_1\neq 0,\\ \beta_1\neq 0,\alpha_1:\end{cases}$ $F_{\alpha\beta\gamma}=\omega^{(\gamma_2-\alpha_2)\epsilon/2}e^{i\pi\eta\epsilon}\dfrac{\pi}{(\eta\quad\gamma_2-\alpha_2)}$, $F^1_{\alpha\beta\gamma}=\omega^{(\gamma_2-\alpha_2)\epsilon/2}\begin{pmatrix}\eta & 0\\ \eta & \gamma_2-\alpha_2\end{pmatrix}$;

$$1\text{-}4\ \begin{cases}\gamma_1=\alpha_1+\beta_1\neq 0(n),\\ \alpha_1,\beta_1\neq 0:\end{cases}\quad F_{\alpha\beta\gamma}=\omega^{-(\alpha_2+\beta_2-\gamma_2)/2}\frac{e^{-i\pi\eta}\pi}{(\eta\quad \alpha_2+\beta_2-\gamma_2)},$$

$$F^1_{\alpha\beta\gamma}=\omega^{-(\alpha_2+\beta_2-\gamma_2)/2}e^{-2i\pi\eta\theta_+}\begin{pmatrix}\eta & 0\\ \eta & \alpha_2+\beta_2-\gamma_2\end{pmatrix},$$

$$F^{1'}_{\alpha\beta\gamma}=\omega^{-(\alpha_2+\beta_2-\gamma_2)/2}e^{-i\pi\eta}\begin{pmatrix}\eta & 0\\ 0 & \beta_2-\alpha_2\end{pmatrix}\begin{pmatrix}\eta & \beta_2-\alpha_2\\ \eta & \alpha_2+\beta_2-\gamma_2\end{pmatrix};$$

$$1\text{-}5\ \begin{cases}\gamma_1=\beta_1\neq 0,\\ \alpha_1\neq 0,\beta_1:\end{cases}\quad F_{\alpha\beta\gamma}=-\omega^{(\beta_2-\gamma_2)\epsilon/2}\frac{\pi e^{i\pi\eta\epsilon}}{(\eta\quad \beta_2-\gamma_2)},\quad F^1_{\alpha\beta\gamma}=-\omega^{(\beta_2-\gamma_2)\epsilon/2}\begin{pmatrix}\eta & 0\\ \eta & \beta_2-\gamma_2\end{pmatrix};$$

$$1\text{-}6\ \begin{cases}\gamma_1=\beta_1=0,\\ \alpha_1>0:\end{cases}\quad F_{\alpha\beta\gamma}=-\frac{\pi}{(\eta\quad \gamma_2)}\begin{pmatrix}2\eta & \beta_2\\ \eta & \beta_2-\gamma_2\end{pmatrix},\quad F^1_{\alpha\beta\gamma}=-\begin{pmatrix}\eta & 0\\ \eta & \gamma_2\end{pmatrix}\begin{pmatrix}\eta & \beta_2\\ \eta & \beta_2-\gamma_2\end{pmatrix};$$

$$1\text{-}7\ \begin{cases}\gamma_1=\alpha_1=0,\\ \beta_1>0:\end{cases}\quad F_{\alpha\beta\gamma}=\frac{\pi}{(\eta\quad \gamma_2)}\begin{pmatrix}0 & \alpha_2\\ \eta & \gamma_2-\alpha_2\end{pmatrix},\quad F^1_{\alpha\beta\gamma}=\begin{pmatrix}\eta & 0\\ \eta & \gamma_2\end{pmatrix}\begin{pmatrix}\eta & \alpha_2\\ \eta & \gamma_2-\alpha_2\end{pmatrix};$$

$$1\text{-}8\ \left[\begin{array}{l}\gamma_1=0,\\ \alpha_1+\beta_1=n,\\ \alpha_1,\beta_1\neq 0:\end{array}\right.\quad F_{\alpha\beta\gamma}=-\frac{\pi}{(\eta\quad \gamma_2)}\begin{pmatrix}0 & \alpha_2+\beta_2-2\gamma_2\\ \eta & \alpha_2+\beta_2-\gamma_2\end{pmatrix}$$

$$F^1_{\alpha\beta\gamma}=-e^{-i\pi\eta\epsilon}\begin{pmatrix}\eta & 0\\ \eta & \gamma_2\end{pmatrix}\begin{pmatrix}0 & \alpha_2+\beta_2-2\gamma_2\\ \eta & \alpha_2+\beta_2-\gamma_2\end{pmatrix},$$

$$F^{1'}_{\alpha\beta\gamma}=-\begin{pmatrix}\eta & 0\\ \eta & \gamma_2\end{pmatrix}\begin{pmatrix}\eta & \beta_2-\alpha_2\\ 0 & \beta_2-\alpha_2\end{pmatrix}\begin{pmatrix}0 & \alpha_2+\beta_2-2\gamma_2\\ \eta & \alpha_2+\beta_2-\gamma_2\end{pmatrix};$$

$$1\text{-}9\ \begin{cases}\beta_1=0,\\ \gamma_1=\alpha_1\neq 0:\end{cases}\quad F_{\alpha\beta\gamma}=\frac{\pi}{(\eta\quad \gamma_2-\alpha_2)}\begin{pmatrix}2\eta & \beta_2\\ \eta & \alpha_2+\beta_2-\gamma_2\end{pmatrix},\quad F^1_{\alpha\beta\gamma}=\begin{pmatrix}\eta & 0\\ \eta & \gamma_2-\alpha_2\end{pmatrix}\begin{pmatrix}\eta & \beta_2\\ \eta & \alpha_2+\beta_2-\gamma_2\end{pmatrix};$$

$$1\text{-}10\ \begin{cases}\gamma_1=\beta_1\neq 0,\\ \alpha_1=0:\end{cases}\quad F_{\alpha\beta\gamma}=-\frac{\pi}{(\eta\quad \beta_2-\gamma_2)}\begin{pmatrix}0 & \alpha_2\\ \eta & \alpha_2+\beta_2-\gamma_2\end{pmatrix},\quad F^1_{\alpha\beta\gamma}=-\begin{pmatrix}\eta & 0\\ \eta & \beta_2-\gamma_2\end{pmatrix}\begin{pmatrix}\eta & \alpha_2\\ \eta & \alpha_2+\beta_2-\gamma_2\end{pmatrix};$$

$$1\text{-}11\ \alpha_1=\beta_1=\gamma_1\neq 0:\quad F_{\alpha\beta\gamma}=\frac{\pi}{(\eta\quad \gamma_2-\alpha_2)}\begin{pmatrix}0 & \alpha_2+\beta_2-2\gamma_2\\ \eta & \beta_2-\gamma_2\end{pmatrix},$$

$$F^1_{\alpha\beta\gamma}=\begin{pmatrix}\eta & 0\\ \eta & \gamma_2-\alpha_2\end{pmatrix}\begin{pmatrix}\eta & \beta_2-\alpha_2\\ \eta & \beta_2-\gamma_2\end{pmatrix}\begin{pmatrix}0 & \alpha_2+\beta_2-2\gamma_2\\ 0 & \beta_2-\alpha_2\end{pmatrix};$$

$$1\text{-}12\ \alpha_1=\beta_1=\gamma_1=0:\quad F_{\alpha\beta\gamma}=\frac{\pi}{(\eta\quad \gamma_2)}\begin{pmatrix}0 & \alpha_2\\ \eta & \gamma_2-\alpha_2\end{pmatrix}\begin{pmatrix}0 & \alpha_2+\beta_2-2\gamma_2\\ \eta & \alpha_2+\beta_2-\gamma_2\end{pmatrix}\begin{pmatrix}2\eta & \beta_2\\ \eta & \beta_2-\gamma_2\end{pmatrix},$$

$$F^1_{\alpha\beta\gamma}=\begin{pmatrix}\eta & 0\\ \eta & \gamma_2\end{pmatrix}\begin{pmatrix}\eta & \alpha_2\\ \eta & \gamma_2-\alpha_2\end{pmatrix}\begin{pmatrix}\eta & \beta_2\\ \eta & \alpha_2+\beta_2-\gamma_2\end{pmatrix}\begin{pmatrix}\eta & \beta_2-\alpha_2\\ \eta & \beta_2-\gamma_2\end{pmatrix}\begin{pmatrix}0 & \alpha_2+\beta_2-2\gamma_2\\ 0 & \beta_2-\alpha_2\end{pmatrix};$$

$$2\text{-}1\ \begin{cases}\alpha_1,\beta_1\neq 0,\\ \alpha_1\neq\beta_1:\end{cases}\quad F_{\alpha\beta\alpha}=e^{i\pi\eta\epsilon}\frac{\pi}{(\eta\quad 0)};$$

$$2\text{-}2\quad \alpha_1=\beta_1\neq 0: F_{\alpha\beta\alpha}=\frac{\pi}{(\eta\quad 0)}\begin{pmatrix}0 & \beta_2-\alpha_2\\ \eta & \beta_2-\alpha_2\end{pmatrix};$$

$$2\text{-}3\ \begin{cases}\beta_1=0,\\ \alpha_1\neq 0:\end{cases}\quad F_{\alpha\beta\alpha}=\frac{\pi}{(\eta\quad 0)}\begin{pmatrix}2\eta & \beta_2\\ \eta & \beta_2\end{pmatrix};$$

$$2\text{-}4\ \begin{cases}\alpha_1=0,\\ \beta_1\neq 0:\end{cases}\quad F_{\alpha\beta\alpha}=\frac{\pi}{(\eta\quad 0)}\begin{pmatrix}0 & \alpha_2\\ \eta & \alpha_2\end{pmatrix};$$

$$2\text{-}5\quad \alpha_1=\beta_1=0: F_{\alpha\beta\alpha}=\frac{\pi}{(\eta\quad 0)}\begin{pmatrix}0 & \alpha_2\\ \eta & \alpha_2\end{pmatrix}\begin{pmatrix}2\eta & \beta_2\\ \eta & \beta_2\end{pmatrix}\begin{pmatrix}0 & \beta_2-\alpha_2\\ \eta & \beta_2-\alpha_2\end{pmatrix};$$

$$3\text{-}1\quad \gamma_1\neq 0,\beta_1:\quad F_{0\beta\gamma}=0;$$

$$3\text{-}2\ \begin{cases}\gamma_1=0,\\ \beta_1\neq 0:\end{cases}\quad F_{0\beta\gamma}=\frac{\pi^2}{(\eta\quad \gamma_2)^2},\quad F^1_{0\beta\gamma}=\begin{pmatrix}\eta & 0\\ \eta & \gamma_2\end{pmatrix}^2;$$

$$3\text{-}3\quad \gamma_1=\beta_1\neq 0:\quad F_{0\beta\gamma}=-\frac{\pi^2}{(\eta\quad \beta_2-\gamma_2)^2},\quad F^1_{0\beta\gamma}=-\begin{pmatrix}\eta & 0\\ \eta & \beta_2-\gamma_2\end{pmatrix}^2;$$

$$3\text{-}4\quad \gamma_1=\beta_1=0:\quad F_{0\beta\gamma}=\frac{\pi^2}{(\eta\quad \gamma_2)^2}\frac{(0\quad \beta_2-2\gamma_2)(2\eta\quad \beta_2)}{(\eta\quad \beta_2-\gamma_2)^2},\quad F^1_{0\beta\gamma}\begin{pmatrix}\eta & 0\\ \eta & \gamma_2\end{pmatrix}^2\begin{pmatrix}\eta & \beta_2\\ \eta & \beta_2-\gamma_2\end{pmatrix}^2\begin{pmatrix}0 & \beta_2-2\gamma_2\\ 0 & \beta_2\end{pmatrix};$$

$$4\text{-}1\quad \beta_1\neq 0:\quad F_{0\beta 0}=\frac{\pi^2}{(\eta\quad 0)^2};$$

4-2 $\beta_1 = 0$: $F_{0\beta 0} = \frac{\pi^2}{(\eta \quad 0)^2} \frac{(0 \quad \beta_2)(2\eta \quad \beta_2)}{(\eta \quad \beta_2)^2}$.

For $n = 2$ we have

$F^1_{0,10,11}(\eta, i\infty) = J_{23}(i\infty) = -\tan^2 \pi\eta,$

$F^1_{0,11,01}(\eta, i\infty) = J_{31}(i\infty) = \tan^2 \pi\eta,$

$F^1_{0,01,10}(\eta, i\infty) = J_{12}(i\infty) = 0.$

The CSC's $f_{\alpha\beta\gamma}(i\infty)$'s may be written down from above $F_{\alpha\beta\gamma}(\eta, i\infty)$ by means of (41). We see that nonvanishing $f_{\alpha\beta\gamma}(i\infty)$'s exhibit a nontrivial classical trigonometric algebra.

VII. CONCLUDING REMARKS

Recently quantum algebra (quantum group) has become an interesting subject and many approaches to the trigonometric quantum group are appearing. However, the algebra in the elliptic case is more complicated; on that subject, fewer results have been published, and its relation to the trigonometric case is known only for $n = 2$. In this article we have, for $n \geqslant 2$, derived the classical and quantum algebras as well as their compatibility conditions, obtained some explicit expressions for the structure constants and their symmetric relations, given exact relations between the classical algebra and the quantum one, and worked out the corresponding trigonometric algebras.

ACKNOWLEDGMENTS

The authors are indebted to Professor C. N, Yang for his stimulating lectures on statistical mechanics and braid group at Nanki University.

This work was supported in part by the Natural Science Fund of China.

[1] C. N. Yang, Phys. Rev. Lett. **19**, 1312 (1967).
[2] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic, London, 1982).
[3] A. A. Belavin, Nucl. Phys. B **180**, 189 (1980).
[4] B. Y. Hou, M. L. Yan, and Y. K. Zhou, Nucl. Phys. B **324**, 715 (1989).
[5] P. P. Kulish, N. Yu. Reshetikhin, and E. K. Sklyanin, Lett. Math. Phys. **5**, 393 (1981).
[6] E. Date, M. Jimbo, A. Kuniba, T. Miwa, and M. Okado, Nucl. Phys. B **290**, 231 (1987).
[7] V. G. Drinfeld, Proc. of the Intl. Congress of Mathematicians, Berkeley, 1986, Vol. 1, p. 798.
[8] E. K. Sklyanin, Funct. Anal. Appl. **16**, 263 (1982); **17**, 273 (1983).
[9] I. V. Cherednik, Func. Anal. Appl. **19**,77 (1985).
[10] V. Jones, Ann. Math. **126**, 335 (1987).
[11] T. Kohno, Ann. Inst. Fourier **37**, 139 (1987).
[12] Y. Akutsu and M. Wadati, J. Phys. Soc. Jpn. **56**, 839 (1987).
[13] A. Tsuchiya and Y. Kanie, Lett. Math. Phys. **13**, 303 (1987).
[14] E. Witten, Proc. of the IAMP Congress, Swansea, July, 1988.
[15] H. Wei and B. Y. Hou, NWU-IMP-89-59.

The four sets of additive quantum numbers of SU(3)

J. Patera
Centre de Recherches Mathématiques, Université de Montréal, Montréal, Québec H3C 3J7, Canada

(Received 19 July 1988; accepted for publication 2 August 1989)

The four maximal sets of additive quantum numbers and related fine gradings of the Lie algebra sl(3,$\mathbb{C}$) are described in detail. The quantum numbers are determined by a grading of the Lie algebra, fine gradings providing maximal sets of them. Two sets of additive quantum numbers are equivalent precisely if the corresponding gradings are equivalent under a transformation from the automorphism group of the Lie algebra.

I. INTRODUCTION

In quantum mechanics it is often convenient to distinguish three types of quantum numbers: additive, multiplicative, and the others. The additive ones are eigenvalues of operators of infinitesimal transformations, the multiplicative ones are eigenvalues of operators performing finite transformations, and the remaining ones are exemplified by the Casimir operators.

Quantum numbers are in principle measurable quantities, hence real numbers, which are used to label states of a physical system. Their physical interpretation ranges from obvious symmetries of a crystal to, for instance, the rather obscure eight additive quantum numbers of E_8 (assuming that this group reflects some properties of the physical world). Whatever may be the situation with interpretation of some quantum numbers at a given time, the corresponding symmetry will remain at best a speculative frontier of physics for as long as those quantum numbers do not become tangible, at least in principle, for experimentalists.

Conversely, for the same reason it is important to know all really different choices of quantum numbers for a given, say established, symmetry: What are the really different quantum numbers one may use with that symmetry? Here by "really different" we obviously mean more than just eigenvalues of different linear combinations of the same set of diagonal operators. Surprisingly the answer cannot be given except for the case described in this paper because the corresponding questions in mathematics are only now being asked.[1,2] Moreover, it is quite likely that states of a physical system can be built in a much simpler way in one basis of the corresponding space than in another, because the states of the system are more simply characterized by one set of quantum numbers than by another. An unsuitable basis may obscure the role of such a symmetry, for instance, by "mixing pure states."

The simple Lie group SU(3) is ever present in elementary particle physics. The corresponding quantum numbers may refer to colors, flavors, properties of harmonic oscillators, and many other quantities related to SU(3) symmetry. In spite of that the following question has not been answered before: *What are the really different choices of additive SU(3) quantum numbers?*

The answer is given in this paper: there are precisely four different maximal sets of additive quantum numbers related to SU(3).

In order to begin, we need to explain what is meant by the additivity of a set of quantum numbers and how to decide whether two sets of additive quantum numbers are equivalent or not. In the paper it is first noticed that the difference commonly made between additive and multiplicative quantum numbers if not essential for the purpose here (logarithms of multiplicative quantum numbers add up). It is explained that the existence of a set of additive quantum numbers hinges on the existence of a grading of the corresponding linear space V. In this paper V is the Lie algebra sl(3,$\mathbb{C}$). Then a natural equivalence relation for gradings of Lie algebras[1] is the equivalence of the corresponding grading decompositions of the Lie algebra under the action of the group of automorphisms of the Lie algebra.

Each set of SL(3,$\mathbb{C}$) quantum numbers is described in a general way and exemplified by a realization of the generators of the Lie algebra by matrices. Their nonequivalence is then easy to verify. The fact that there are no other but the four possibilities described below is a result of Refs. 1 and 2; it is not proven here. A representation of the gradings is found in the appendix of Ref. 1.

It turns out that only one of the four sets is the one familiar in particle physics [or rather all common choices of additive SU(3) quantum numbers are equivalent to it]. Existence of the other three possibilities raises amusing questions in SU(3) representation theory. Namely, how to use those quantum numbers in describing representations of SU(3). No answer has yet been given to that but, in general, it is clear that equivalent gradings of the Lie algebra produce nonequivalent gradings of representations. These problems are studied elsewhere.[3,4] Here we encounter only the lowest representations that pose no problems.

It is a common practice in physics to speak about a compact simple Lie group such as SU(2), SU(3),..., while really using the group (or Lie algebra) with complex parameters. For instance, the generators L_1 and L_{-1} of the angular momentum theory generate the Lie algebra of SL(2,$\mathbb{C}$) rather than SU(2). Such an inconsistency is then remedied at the end of a calculation typically by taking appropriate linear combinations and restricting parameters to real values when needed. Similarly in this paper the relevant Lie algebra is sl(3,$\mathbb{C}$) of the group SL(3,$\mathbb{C}$).

The question that is asked in this paper is by no means limited to SU(3) or rather to SL(3,$\mathbb{C}$), however, all maximal additive sets of quantum numbers are not known for simple Lie groups others than SL(2,$\mathbb{C}$) and SL(3,$\mathbb{C}$) [there are apparently 9 of them for SL(4,ℂ)]. A method to construct them is the main content of Ref. 2. Some uncommon sets of quantum numbers for the Lie algebras of the groups SL(n,ℂ) are described in Ref. 5.

II. NONEQUIVALENT ADDITIVE QUANTUM NUMBERS

The physical states that are labeled (often not completely) by a set of quantum numbers are eigenvectors of corresponding "labeling" operators Q (there may be several of them at a time and they necessarily commute). Practically this means that in the corresponding space V of physical states one can choose a basis $\{|\lambda\rangle, \lambda\in S\}$ of eigenvectors of Q's, where S is the set of eigenvalues of Q's in the space V.

If λ_1 and λ_2 are two different quantum numbers corresponding to the same labeling operator Q, one has the additivity of the quantum numbers of the physical states

$$|\lambda_1\rangle|\lambda_2\rangle = |\lambda_1+\lambda_2\rangle, \qquad |\lambda_1\rangle\in V_1,\ |\lambda_2\rangle\in V_2,\ |\lambda_1\rangle|\lambda_2\rangle\in V_1\otimes V_2 \tag{1}$$

as a consequence of the infinitesimal action of Q: on a product state Q acts as a derivation,

$$\begin{aligned} Q(|\lambda_1\rangle|\lambda_2\rangle) &= Q(|\lambda_1\rangle)|\lambda_2\rangle + |\lambda_1\rangle Q(|\lambda_2\rangle) \\ &= \lambda_1|\lambda_1\rangle|\lambda_2\rangle + \lambda_2|\lambda_1\rangle|\lambda_2\rangle \\ &= (\lambda_1+\lambda_2)|\lambda_1\rangle|\lambda_2\rangle. \end{aligned}$$

Suppose now that λ_1 and λ_2 are multiplicative quantum numbers corresponding to the operator denoted again by Q. Then one has the multiplicativity of the quantum numbers of the states

$$|\lambda_1\rangle|\lambda_2\rangle = |\lambda_1\lambda_2\rangle \tag{2}$$

following from the finite transformation action of Q:

$$Q(|\lambda_1\rangle|\lambda_2\rangle) = Q(|\lambda_1\rangle)Q(|\lambda_2\rangle) = \lambda_1\lambda_2|\lambda_1\rangle|\lambda_2\rangle.$$

In the case of the nonadditive and nonmultiplicative quantum numbers, the quantum number of the state $|\lambda_1\rangle|\lambda_2\rangle$ is *not* determined by λ_1 and λ_2 because Q is not diagonal in the basis $|\lambda_1\rangle|\lambda_2\rangle$.

For our purposes the distinction between the multiplicative and additive quantum numbers is not a major one. Indeed, nothing is lost by making a different labeling convention in (2), replacing λ_1 and λ_2 by $\exp(\ln\lambda_1)$ and $\exp(\ln\lambda_2)$, and using the logarithms of λ_1 and λ_2 as the labels—quantum numbers. Then (2) is rewritten in the additive form (1):

$$|\ln\lambda_1\rangle|\ln\lambda_2\rangle = |\ln\lambda_1+\ln\lambda_2\rangle \tag{3}$$

and we speak of the additive form of the multiplicative quantum numbers.

Consequently we need to distinguish only two classes: the *additive* and *nonadditive* quantum numbers. In this paper we are concerned with the former ones, assuming always that multiplicative quantum numbers are used in their additive form.

The *crucial* property is the fact that on the left side of (1) there is a state labeled by $\lambda_1+\lambda_2$ and not a linear combination of states with different values of the quantum number. Only then do λ_1 and λ_2 determine the quantum number of the product state $|\lambda_1\rangle|\lambda_2\rangle$; that is the additivity of the quantum numbers. Mathematically the general property that reflects that fact and thus underlies the notion of the additive quantum numbers is called *grading*.

In general, the space V of the physical states can be decomposed into a direct sum of subspaces

$$V = \bigoplus_{\lambda\in S} V_\lambda \tag{4}$$

labeled by the eigenvalues λ of an operator Q. It is called the *grading decomposition* of V provided there is a label set S such that (1) holds for any

$$|\lambda_1\rangle\in V_{\lambda_1},\quad |\lambda_2\rangle\in V_{\lambda_2}, \qquad |\lambda_1+\lambda_2\rangle\in V_{\lambda_1+\lambda_2}\quad (\lambda_1,\ \lambda_2,\ \lambda_1+\lambda_2\in S). \tag{5}$$

Note that it is not required that the subspaces V_λ are one dimensional or that S is unique. In general, one has several (commuting) operators Q, each providing different quantum numbers. In that case the subspaces V_λ of (4) may be further decomposed. We say that the grading decomposition (4) is refined. It is convenient then to use multicomponent labels for subspaces and for states so that each component is one quantum number. If no further refinement is possible the grading is called *fine*.

Practically, a grading of V is accomplished provided one knows a labeling set S of the quantum numbers and a basis for each subspace V_λ of (4), such that (1) and (5) are fulfilled. Two sets S and S' of additive quantum numbers need to be distinguished only if the corresponding decompositions (4) are "really" different.

The question about all different maximal sets of commuting, simultaneously diagonalizable operators Q acting in the space V giving maximal sets of additive quantum numbers is thus reduced to the question about all really different fine gradings of the space V. In order to be able to give an answer to such a question we have to say more about the space V.

In this paper we consider the case $V=\mathrm{sl}(3,\mathbb{C})$. Therefore our task is to present the Lie algebra in the form of a direct sum of subspaces

$$\mathrm{sl}(3,\mathbb{C}) = \bigoplus_{\lambda\in S} V_\lambda, \tag{6}$$

called a *grading decomposition*, such that the grading property

$$[V_\lambda, V_\mu] \subseteq V_{\lambda+\mu}, \quad \lambda,\ \mu,\ \lambda+\mu\in S \tag{7}$$

of the commutation relations is satisfied. Here S is a set of distinct labels—quantum numbers. The grading subspaces are pairwise orthogonal:

$$\mathrm{tr}\, v_\lambda v_\mu = 0, \quad \text{for any } \lambda+\mu\neq 0, \qquad v_\lambda\in V_\lambda, v_\mu\in V_\mu. \tag{8}$$

Practically, we need to know generators of each subspace V_λ and the set of quantum numbers S for which the subspaces are nontrivial (dimension $\dim V_\lambda>0$). Moreover, since we are interested in fine gradings of sl(3,ℂ) (which cannot be made finer) we require as well that the decomposition (6) contains as many terms as possible. Note that for $[V_\lambda, V_{\lambda'}] = 0$ Eq. (7) implies no restriction. For instance the subspace $V_{\lambda+\lambda'}$ may be trivial, dim $V_{\lambda+\lambda'} = 0$.

There is a natural *equivalence relation* for gradings of a Lie algebra[1] L: Two gradings are equivalent if the corresponding grading decompositions of L are transformed into each other by an action of the group Aut(L) of automorphisms of L. In particular, two gradings whose grading subspaces have different dimensions are obviously nonequivalent.

The maximal subgroup of Aut(L), which preserves a grading decomposition Γ of L, is the *grading* group of Γ. Two gradings of L are equivalent if their grading groups are conjugate under Aut(L). It was established in Ref. 1 that there is a one-to-one correspondence between maximal abelian diagonable subgroups (MAD groups) of Aut(L) and fine gradings of L.

For a fine grading a simple Lie algebra L it is sufficient to use only one or two suitably chosen elements of a MAD group with sufficiently many distinct eigenvalues when acting on L.

III. ADDITIVE QUANTUM NUMBERS OF sl(2,ℂ)

There are exactly two fine gradings[5] of sl(2,ℂ):

I. $\mathrm{sl}(2,\mathbb{C}) = V_1 \oplus V_0 \oplus V_{-1}$

$$[V_j, V_k] \subseteq V_{j+k},$$
$$(j,\ k,\ j+k \in S = \{1,0,-1\}).$$

II. $\mathrm{sl}(2,\mathbb{C}) = V_{11} \oplus V_{01} \oplus V_{10}$

$$[V_{ab}, V_{cd}] \subseteq V_{a+c,b+d},$$
$$(a,b,c,d,a,+c,b+d \text{ are integers mod } 2).$$

In the first case the grading subspaces are generated by the familiar infinitesimal operators L_1, L_0, L_{-1} of angular momentum theory in arbitrary normalization. The generators in the second case are the Pauli matrices σ_x, σ_y, σ_z also arbitrarily normalized and denoted by a pair (a,b) of integers mod 2 such that not both integers are simultaneously zero.

Nonequivalence of the two gradings becomes clear once we notice that there are two subspaces, V_1 and V_{-1}, in the first case consisting entirely of nilpotent elements of the Lie algebra, while in the second case there is no such grading subspace.

In the first case the grading group is an SL(2,ℂ) torus [cf. (34) below] while in the second one it is the adjoint action of the finite group[5] $\wp_2$ of order 8 generated by the Pauli matrices that are normalized to be antihermitian and of determinant one.

It is not an accident that among the infinity of possible choices of sl(2,ℂ) generators the two bases above are practically the only ones used. Other choices are quite awkward and are used only for special purposes [cf. Ref. 6]. It is a direct consequence of the fact that bases dictated by fine gradings reflect structural properties of the algebra and are therefore far more convenient to use.

IV. FOUR SETS OF ADDITIVE QUANTUM NUMBERS OF SL(3,ℂ)

From now on we assume that V is the linear space of the Lie algebra of complex 3×3 traceless matrices, the simple Lie algebra denoted either by A_2 or sl(3,ℂ). The operators Q are either elements of the Lie algebra that can be diagonalized (semisimple elements), or they are the finite invertible transformations that act on any element X of A_2 as QXQ^{-1} and leave the whole Lie algebra invariant. In the latter case Q belong to the group of automorphisms of A_2 that properly contains the Lie group SL(3,ℂ).

Let us consider the four gradings one by one.

(1) Cartan decomposition, also called the root decomposition, or the toroidal decomposition, is well known:

$$\mathrm{sl}(3,\mathbb{C}) = V_0 \oplus V_\alpha \oplus V_{-\alpha} \oplus V_\beta \oplus V_{-\beta} \oplus V_{\alpha+\beta} \oplus V_{-\alpha-\beta}. \tag{9}$$

The grading implies the commutation relations in the form (7) where $S = \{\pm\alpha, \pm\beta, \pm(\alpha+\beta), 0\}$ is the set of roots of the algebra (including two zero roots). If $\lambda + \mu \notin S$ one must have $[V_\mu, V_\lambda] = 0$. The subspace V_0 is two dimensional, other grading subspaces are of dimension 1. Consequently six generators are prescribed up to a normalization (multiplication by a nonzero complex valued constant) by their quantum numbers. The two generators Q_1 and Q_2 of V_0 are a *matter of choice*. In particular, the orthogonality (8) of Q_1 and Q_2 is not automatic, they must be chosen accordingly for that. Once those generators are chosen, eigenvalues λ_1 and λ_2 of their commutation action

$$[Q_j, X] = \lambda_j X, \quad j = 1,2, \tag{10}$$

on the elements X of sl(3,ℂ), are the additive quantum numbers (coordinates of the roots). The commuting generators Q_1 and Q_2 are therefore labeled by the same quantum numbers $\lambda_1 = \lambda_2 = 0$. Choosing instead of Q_1 and Q_2 their linear combinations, one does not change the grading of the Lie algebra; the roots will of course be given relative to a different basis of V_0. Therefore such choices are equivalent. A standard choice in physics makes one of the quantum numbers into the projection of the isospin and the other into the hypercharge, another common choice makes them the projections of u spin and v spin. A standard choice in mathematics is that of (13) below.

The generators of the decomposition (9) are often represented by the 3×3 matrices $E_{jk} = (\delta_{jr}\delta_{sk})$ satisfying the commutation relations

$$[E_{jk}, E_{mn}] = \delta_{km} E_{jn} - \delta_{nj} E_{mk}. \tag{11}$$

The grading decomposition is then the sum of six one-dimensional subspaces of off-diagonal matrices labeled in (9) by the positive and negative roots $\pm\alpha$, $\pm\beta$, $\pm(\alpha+\beta)$, and a subspace of dimension 2 of traceless diagonal matrices:

$$\mathrm{sl}(3,\mathbb{C}) = (\mathbb{C}(E_{11} - E_{22}) + \mathbb{C}(E_{22} - E_{33})) \oplus \bigoplus_{\substack{j,k=1\\ j\neq k}}^{3} (\mathbb{C}E_{jk}). \tag{12}$$

Here ℂ denotes an arbitrary complex valued coefficient. Therefore

$$V_0 = \mathbb{C}(E_{11} - E_{22}) + \mathbb{C}(E_{22} - E_{33}). \tag{13}$$

The rest of the correspondence between (9) and (12) can be set up in many ways. For example,

$$E_{12}\in V_\alpha, \quad E_{23}\in V_\beta, \quad E_{13}\in V_{\alpha+\beta},$$
$$E_{21}\in V_{-\alpha}, \quad E_{32}\in V_{-\beta}, \quad E_{31}\in V_{-\alpha-\beta}.$$

The seven distinct pairs of quantum numbers (coordinates of roots) is then drawn in the familiar hexagonal form with the origin containing the two weights (0,0).

The grading group in this case is the maximal torus of SL(3,$\mathbb{C}$) generated by a basis of V_0.

(2). The grading of sl(3,$\mathbb{C}$) by the finite group $\wp_3$ of generalized Pauli matrices is described in Ref. 5. On a number of occasions this basis of sl(3,$\mathbb{C}$) appeared[7] in physics literature in a different context. Let us recall it briefly;

$$\mathrm{gl}(3,\mathbb{C}) = \bigoplus_{a,d=-1}^{1} \{\mathbb{C}(a,d)\} = \mathbb{C}(0,0) \oplus \mathrm{sl}(3,\mathbb{C})$$
$$= \mathbb{C}(0,0) \oplus \bigoplus_{\substack{a,d=-1\\ a=d\neq 0}}^{1} \{\mathbb{C}(a,d)\}. \tag{14}$$

The generators (a,d) satisfy the commutation relations

$$[(a,d),(a',d')] = (\omega^{ad'} - \omega^{a'd})(a+a',d+d'),$$
$$\omega = \exp(2\pi i/3). \tag{15}$$

Here a, a', d, d' and the sums $a+a'$ and $d+d'$ are taken mod 3. The grading property is evident in (15).

A faithful matrix representation of the generators is, for instance, given by[5]

$$(0,-1) = \begin{pmatrix} 0 & 1 & 0\\ 0 & 0 & 1\\ 1 & 0 & 0\end{pmatrix}, \quad (0,1) = \begin{pmatrix} 0 & 0 & 1\\ 1 & 0 & 0\\ 0 & 1 & 0\end{pmatrix},$$

$$(1,0) = \begin{pmatrix} 1 & 0 & 0\\ 0 & \omega & 0\\ 0 & 0 & \omega^2\end{pmatrix}, \quad (-1,0) = \begin{pmatrix} 1 & 0 & 0\\ 0 & \omega^2 & 0\\ 0 & 0 & \omega\end{pmatrix},$$

$$(1,-1) = \begin{pmatrix} 0 & \omega & 0\\ 0 & 0 & \omega^2\\ 1 & 0 & 0\end{pmatrix}, \quad (-1,1) = \begin{pmatrix} 0 & 0 & \omega\\ 1 & 0 & 0\\ 0 & \omega^2 & 0\end{pmatrix},$$

$$(-1,-1) = \begin{pmatrix} 0 & \omega^2 & 0\\ 0 & 0 & \omega\\ 1 & 0 & 0\end{pmatrix}, \quad (1,1) = \begin{pmatrix} 0 & 0 & \omega^2\\ 1 & 0 & 0\\ 0 & \omega & 0\end{pmatrix},$$

$$(0,0) = \begin{pmatrix} 1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 1\end{pmatrix}. \tag{16}$$

In this case the eight-dimensional Lie algebra sl(3,$\mathbb{C}$) is graded into eight one-dimensional subspaces that are pairwise orthogonal with respect to the Killing form (8). Hence *all* the generators in this case are *prescribed* by the grading up to a normalization. We say that the grading is not only fine but also finest. The appearance of the matrices (16) representing the generators is, of course, basis dependent. Note that the matrices (16) under matrix multiplication generate a finite group of order 27 denoted by $\wp_3$ in Ref. 5.

In the previous case relations between generators were conveniently visualized as six points of a regular hexagon and the origin at the center. In the present case it would be necessary to draw the similar picture on the surface of a two-dimensional torus with the origin at the point (0,0) on the torus.

The quantum numbers now are the exponents a and d of the eigenvalues of the group action

$$Q_a X Q_a^{-1} = \omega^a X, \quad Q_d X Q_d^{-1} = \omega^d X,$$
$$\omega = \exp(2\pi i/3), \tag{17}$$

of the matrices $Q_a = (0,-1)$ and $Q_d = (1,0)$ of (16), respectively. Equivalently, we could have chosen any other pair of noncommuting matrices (16) as the labeling operators Q_a and Q_d. There is no other freedom in choosing the quantum numbers for this grading.

(3). An essential part of the next case is an o(3) grading of sl(3,$\mathbb{C}$). However, the grading of sl(3,$\mathbb{C}$) by the torus of O(3) is not fine. Therefore the torus has to be extended by the outer automorphism of sl(3,$\mathbb{C}$) in order to produce a maximal grading group in Aut(sl(3,$\mathbb{C}$)), which then yields a fine grading. As a result, one has the grading decomposition

$$\mathrm{sl}(3,\mathbb{C}) = \bigoplus_{k=-2}^{2} (\mathbb{C}(1,k)) \bigoplus_{k=-1}^{1} (\mathbb{C}(0,k)). \tag{18}$$

The decomposition (18) is a grading, provided one has

$$[(b,c),(b'c')] \subseteq \mathbb{C}(b+b',c+c'), \tag{19}$$

where the first components of generators are taken mod 2 and the second ones as ordinary integers. As in the previous case the grading subspaces are one dimensional. The generators are labeled by distinct quantum numbers and their orthogonality (8) is automatic. They are again determined by the grading up to a normalization.

The conditions (18) and (19) are satisfied, for example, by the following matrix representation of the generators:

$$(0,1) = \begin{pmatrix} 0 & 1 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\end{pmatrix}, \quad (0,0) = \begin{pmatrix} 1 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & -1\end{pmatrix},$$

$$(0,-1) = \begin{pmatrix} 0 & 0 & 0\\ 1 & 0 & 0\\ 0 & 1 & 0\end{pmatrix},$$

$$(1,2) = \begin{pmatrix} 0 & 0 & 1\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix}, \quad (1,1) = \begin{pmatrix} 0 & 1 & 0\\ 0 & 0 & -1\\ 0 & 0 & 0\end{pmatrix},$$

$$(1,0) = \begin{pmatrix} 1 & 0 & 0\\ 0 & -2 & 0\\ 0 & 0 & 1\end{pmatrix},$$

$$(1,-1) = \begin{pmatrix} 0 & 0 & 0\\ 1 & 0 & 0\\ 0 & -1 & 0\end{pmatrix}, \quad (1,-2) = \begin{pmatrix} 0 & 0 & 0\\ 0 & 0 & 0\\ 1 & 0 & 0\end{pmatrix}. \tag{20}$$

Note that (0,1), (0,0), and (0,−1) generate the Lie algebra o(3) of the O(3) subgroup of SL(3,$\mathbb{C}$). With this choice of matrix representation of the generators of (18) and (19), it is obvious that the structure constants are integers.

Consider now the labeling operators Q_b and Q_c, whose action on the generators (b,c) provides the quantum numbers b and c. As one of them we have $Q_c = (0,0)$, the generator of an O(3) torus. Its action

$$[Q_c,(b,c)] = [(0,0),(b,c)] = c(b,c) \tag{21}$$

can be verified directly using (20). We cannot choose Q_b as the other generator that commutes with (0,0), namely (1,0), because then we would have a different grading than (18), namely, a grading equivalent to the Cartan decomposition (9). Therefore one has to look for Q_b outside of the group SL(3,$\mathbb{C}$) and consequently the operator Q_b cannot be realized as a 3×3 matrix.

A faithful representation of the outer automorphisms of sl(3,$\mathbb{C}$) is possible by matrices 6×6. For every generator (b,c) we consider the block diagonal 6×6 matrix

$$\begin{pmatrix} (b,c) & \\ & -(b,c)^T \end{pmatrix}. \tag{22}$$

Then we require that

$$Q_b\begin{pmatrix} (b,c) & \\ & -(b,c)^T \end{pmatrix}Q_b^{-1} = (-1)^b\begin{pmatrix} (b,c) & \\ & -(b,c)^T \end{pmatrix}. \tag{23}$$

Given the specific choice (20) of the generator matrices, the labeling operator Q_b is determined. Namely,

$$Q_b = \begin{pmatrix} & M \\ M & \end{pmatrix}, \quad M = \begin{pmatrix} & & 1 \\ & -1 & \\ 1 & & \end{pmatrix}. \tag{24}$$

The grading (18) cannot be equivalent to (14) because the latter does not contain among its grading subspaces a subalgebra sl(2,$\mathbb{C}$). It is also nonequivalent to (9) because of the absence of a two-dimensional grading subspace in it. The grading group consists of the SL(2,$\mathbb{C}$) torus and by the outer automorphism of sl(2,$\mathbb{C}$).

An analogous picture to the hexagon in the case of the root decomposition (9) now would be a set of eight distinct points on the surface of a cylinder with the quantum number c varying along the axis of the cylinder.

The fact that the outer automorphism of sl(n,$\mathbb{C}$) $n\geqslant 3$, leaves invariant the o(n,$\mathbb{C}$) subalgebra is well known. Its particular matrix representation (24) is a consequence of our choice of the o(3) generators in (20). Instead, if one is guided by the simplicity of Q_b then a favorite choice might be

$$Q_b = \begin{pmatrix} 0 & I_n \\ I_n & 0 \end{pmatrix}, \tag{24$'$}$$

where I_n is the $n\times n$ identity matrix. Then one has

$$Q_b\begin{pmatrix} X & \\ & -X^T \end{pmatrix}Q_b^{-1} = \begin{pmatrix} -X^T & \\ & X \end{pmatrix} = (-1)^b\begin{pmatrix} X & \\ & -X^T \end{pmatrix}.$$

In this case the o(n,$\mathbb{C}$) subalgebra corresponding to the $+1$ eigenvalue consists of all antisymmetric matrices $X^T = -X$, while the remaining sl(n,$\mathbb{C}$) generators (eigenvalue -1) are symmetric matrices. These properties have been used many times in physics, for example, in connection of the Wigner–Inönü contractions, in defining the rotational bands[8] of SU(3), and elsewhere.

(4). In the last case the subalgebra sl(2,$\mathbb{C}$) plays a prominent role but not gl(2,$\mathbb{C}$). Let us first describe the grading:

$$\mathrm{sl}(3,\mathbb{C}) = \bigotimes_{k=0,\pm 2}(\mathbb{C}(0,k)) \bigoplus_{k,j=\pm 1}(\mathbb{C}(j,k)) \oplus \mathbb{C}(2,0), \tag{25}$$

$$[(j,k),(j',k')] \subseteq \mathbb{C}(j+j',k+k'),$$
$$j,j',j+j' \text{ are integers mod 4;}$$
$$k,k',k+k' \text{ are integers.} \tag{26}$$

The generators of the decomposition (25) are denoted by (j,k). Since the decomposition contains eight terms, each generator is determined by it up to a normalization. Relations (26) imply that $Q_1 = (0,0)$ must be the diagonal generator of sl(2,$\mathbb{C}$) and also that we have

$$[(0,0),(j,k)] = k(j,k). \tag{27}$$

Furthermore, $(0,\pm 2)$ must be the remaining generators of the sl(2,$\mathbb{C}$) subalgebra, and together with (2,0) they generate gl(2,$\mathbb{C}$) in sl(3,$\mathbb{C}$). If we would now adopt the generator (2,0) as Q_2, the grading property (26) would not hold because (0,0) and (2,0) would have to be labeled by the same quantum numbers; such a grading would be equivalent to (9). Therefore again, as in the previous case, we have to look for Q_2 outside of the SL(3,$\mathbb{C}$) group.

A matrix representation of the generators of the grading decomposition (25) can be chosen with integer structure constants, for example, as follows:

$$(0,0) = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (1,1) = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & -i & 0 \end{pmatrix},$$

$$(0,2) = \begin{pmatrix} 0 & 1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (1,-1) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ i & 0 & 0 \end{pmatrix},$$

$$(0,-2) = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad (-1,1) = \begin{pmatrix} 0 & 0 & i \\ 0 & 0 & 0 \\ 0 & -1 & 0 \end{pmatrix},$$

$$(2,0) = \begin{pmatrix} -i & 0 & 0 \\ 0 & -i & 0 \\ 0 & 0 & 2i \end{pmatrix},$$

$$(-1,-1) = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & -i \\ -1 & 0 & 0 \end{pmatrix}. \tag{28}$$

The first label of each generator (j,k) is defined by the action

$$Q_2\{(j,k) \oplus -(j,\kappa)^T\}Q_2^{-1} = i^j\{(j,k) \oplus -(j,\kappa)^T\} \tag{29}$$

of Q_2 on $\{(j,k)\oplus -(j,\kappa)^T\}$ as in (22) and (23). Adopting the representation (28) of the generators, we have determined Q_2 as the matrix

$$Q_2=\begin{pmatrix} & & & 0 & 1 & 0\\ & & & -1 & 0 & 0\\ & & & 0 & 0 & 1\\ 0 & 1 & 0 & & & \\ -1 & 0 & 0 & & & \\ 0 & 0 & 1 & & & \end{pmatrix}. \tag{30}$$

The grading (25) is not equivalent to (9) and (14) for the same reason that (18) is not equivalent to (9) and (14); it is also not equivalent to (18) because the grading subspaces of (18) and (25) decompose differently with respect to their respective subalgebras sl(2,$\mathbb{C}$).

The grading group now consists of the O(3) torus, generated by (0,0), and by the outer automorphism of sl(3,$\mathbb{C}$).

IV. MULTIPLICATIVE FORMS OF ADDITIVE QUANTUM NUMBERS AND EQUIVALENT GRADINGS

The modular quantum numbers appear in gradings (14), (18), and (25) because the corresponding labeling operators Q act as finite transformations of finite order on the Lie algebra [cf. (17)] and because we replaced their multiplicative form by their additive form in the sense of (3).

The nonmodular additive quantum numbers can also be replaced by the modular ones; in fact, in many ways. Such a replacement is equivalent for a limited number of low representations, and it may sometimes prove advantageous. Therefore in this section we describe how it can be done.

Consider the first grading of Sec. III explicitly in the lowest representation:

$$L_-=\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix},\ L_0=\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix},\ L_+=\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}. \tag{31}$$

Here the subscripts of the generators refer to the traditional choice of quantum numbers according to the commutation action of L_0 on L_+ and L_-:

$$[L_0,L_\pm]=\pm 2L_\pm . \tag{32}$$

The role of Q is played by L_0. The grading is preserved by any diagonal matrix

$$Q=\exp(iL_0\theta)=\operatorname{diag}\{\exp(i\theta),\exp(-i\theta)\}$$
$$\in \mathrm{SU}(2),\theta\in\mathbb{R}, \tag{33}$$

representing an element of the torus of SU(2). Indeed one has

$$QL_0Q^{-1}=L_0,\ QL_+Q^{-1}=\exp(i\theta)L_+,$$
$$QL_-Q^{-1}=\exp(-i\theta)L_-. \tag{34}$$

Consequently, whole torus preserves that grading of the Lie algebra. It suffices to use just one element of the torus, i.e., to fix a value of θ that is not an integer multiple of π, to provide the additive quantum number labeling the generators (31). With the simple choice $\theta=2\pi/3$, the eigenvalues of (34) are third roots of 1. Thus the generators (31) are labeled by the powers of $\omega=\exp(2\pi i/3)$. Thus a grading of sl(3,$\mathbb{C}$) by a subgroup of the torus, namely the cyclic group of order 3 generated by $Q=\exp(2\pi iL_0/3)$, produces the same grading. However, the action of that cyclic group on representations of sl(3,$\mathbb{C}$) of dimensions >3 produces a much coarser grading than that of the torus.

In the example above we have taken a particular element of order 3 in SU(2) and used it to grade the Lie algebra sl(2,$\mathbb{C}$). The element represents the SU(2) conjagacy class denoted[9] by Refs. 1 and 2. A general element $[s_0,s_1]$ of finite order,[9] where s_0,s_1 are relatively prime non-negative integers, in its diagonal form is then represented by the matrix

$$Q=\operatorname{diag}\{\exp(\pi is_1/(s_0+s_1)),\exp(-\pi is_1/(s_0+s_1))\}. \tag{35}$$

Similarly, in the case of sl(3,$\mathbb{C}$) one may replace a quantum number defined as an eigenvalue of the commutation action of a diagonal generator of a torus of the Lie group by the quantum number which the eigenvalue of the action (34) of the "exponential form" of the generator with a fixed value of the parameter. In other words, one chooses a minimal number of particular elements of the grading group to provide a desired grading. In particular, combining a suitable chosen element of SL(3,$\mathbb{C}$) torus with the outer automorphism, it is possible to replace the labeling operator (30) for instance, by

$$Q=\begin{pmatrix} & & & 0 & \xi^7 & 0\\ & & & \xi & 0 & 0\\ & & & 0 & 0 & i\\ 0 & \xi^5 & 0 & & & \\ \xi^3 & 0 & 0 & & & \\ 0 & 0 & i & & & \end{pmatrix},\quad \xi=e^{2\pi i/8}. \tag{36}$$

Unlike Q_2 in (30), here Q this is an operator of order 8. Indeed, $Q^8=1$ while $Q^4\neq 1$. It has eight distinct eigenvalues, eight roots of 1, which alone can be used to relabel the generators (28) of sl(3,$\mathbb{C}$). Denoting a generator by (k), where k is the exponent of an eigenvalue ξ^k, we have

$$(0)=\begin{pmatrix}1 & 0 & 0\\ 0 & -1 & 0\\ 0 & 0 & 0\end{pmatrix},\ (7)=\begin{pmatrix}0 & 0 & 1\\ 0 & 0 & 0\\ 0 & -i & 0\end{pmatrix},$$
$$(2)=\begin{pmatrix}0 & 1 & 0\\ 0 & 0 & 0\\ 0 & 0 & 0\end{pmatrix},\ (5)=\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & 1\\ i & 0 & 0\end{pmatrix},$$
$$(6)=\begin{pmatrix}0 & 0 & 0\\ 1 & 0 & 0\\ 0 & 0 & 0\end{pmatrix},\ (3)=\begin{pmatrix}0 & 0 & i\\ 0 & 0 & 0\\ 0 & -1 & 0\end{pmatrix},$$
$$(4)=\begin{pmatrix}-i & 0 & 0\\ 0 & -i & 0\\ 0 & 0 & 2i\end{pmatrix},\ (1)=\begin{pmatrix}0 & 0 & 0\\ 0 & 0 & -i\\ -1 & 0 & 0\end{pmatrix}. \tag{37}$$

It is the same set of generators graded as before although the commutation relations now look particularly simple:

$$[(e),(f)]\subseteq\mathbb{C}(e+f), \tag{38}$$

where $e, f, e+f$ are integers mod 8. In this case the generators are easily visualized as the vertices of a regular octagon.

ACKNOWLEDGMENTS

Dr. R. T. Sharp and Dr. J. Van der Jeugt are acknowledged for comments and helpful suggestions.

This work was supported in part by the National Science and Engineering Research Council of Canada and by the Ministére de l'Education du Québec.

[1] J. Patera and H. Zassenhaus, "On Lie gradings I," Linear Algebra Appl. **112**, 87 (1989).

[2] J. Patera and H. Zassenhaus, "On Lie gradings II," in preparation.

[3] J. Patera, R. T. Sharp, and J. Van der Jeugt, "New gradings of sl(3,ℂ) representations," J. Math. Phys. **30**, 2763 (1989).

[4] C. Cummins, J. Patera, and R. T. Sharp, "SL(3,ℂ) to $\wp_3$ basis states and generator matrix elements," in preparation.

[5] J. Patera and H. Zassenhaus, "The Pauli matrices in n-dimensions and finest gradings of Lie algebras of type A_{n-1}," J. Math. Phys. **29**, 665 (1988).

[6] J. Patera and P. Winternitz, "A new basis for the representations of the rotation group, Lamé and Heun polynomials," J. Math. Phys. **14**, 1130 (1974).

[7] A. A. Belavin, "Discrete groups and integrability of quantum systems," Funct. Analysis Appl. **14**, 18 (1980); G. v Gehlen, V. Rittenberg, and H. Ruegg, "Conformal invariance and finite-dimensional quantum chains," J. Phys. A: Math. Gen. **19**, 107 (1985); D. S. Freed and C. Vafa, "Global anomalies on orbifolds," Commun. Math. Phys. **110**, 349 (1987).

[8] L. C. Biedenharn, "Are the rotational bands assigned correctly in the nuclear SU(3) model?," Phys. Lett. B **28**, 537 (1969).

[9] R. V. Moody, J. Patera, and R. T. Sharp, "Character generators for elements of finite order in simple Lie groups A_1, A_2, A_3, B_2, G_2," J. Math. Phys. **24**, 2387 (1983).

New gradings of sl(3,C) representations

J. Patera
Centre de Recherches Mathématiques, Université de Montréal, Montréal, Québec H3C 3J7, Canada
R. T. Sharp
Department of Physics, McGill University, Montréal, Québec H3A 2T8, Canada
J. Van Der Jeugt
Faculty of Mathematical Studies, University of Southampton, Southampton, England

(Received 25 August 1988; accepted for publication 9 August 1989)

Two new gradings of the Lie algebra sl(3,C) are the refined sl(2,C) and o(3) gradings. The grading in each case utilizes the sl(2,C), or o(3), weight together with a new additive modular label. A complete set of sl(3,C) representation basis states labeled by each of the two sets of additive quantum numbers is found. The new labeling operator in both cases fails to commute with the cubic sl(3,C) Casimir operator, and hence mixes states of the contragredient representations (p,q) and (q,p).

I. INTRODUCTION

The complex Lie algebra of traceless 3×3 matrices appears in many applications in physics either in its own right, as a stepping stone to the SU(3) representations, or as a subalgebra of a larger symmetry algebra. The history of those applications is extensive and long. Superficially one may be inclined to share the opinion that "whatever is needed about sl(3,C) and its finite-dimensional representations is already known." But recently new questions have been raised[1] about sl(3,C) (and any other simple Lie algebra), which should be of obvious interest to physicists. One of them concerns nonequivalent choices of additive quantum numbers. It turns out that for sl(3,C) there are four different possibilities and that only one is generally known and used (in a number of disguises); only part of the others is used.

An additive quantum number λ labeling physical states $|\lambda\rangle$ in the most general form[1] can be defined as a number satisfying one of the equalities

$$|\lambda_1\rangle|\lambda_2\rangle = |\lambda_1+\lambda_2\rangle \text{ or } |\lambda_1\rangle|\lambda_2\rangle = |\lambda_1\cdot\lambda_2\rangle. \qquad (1.1)$$

In the latter case one often speaks of a multiplicative quantum numbers; however, without loss of generality one may use $\ln\lambda_1$ and $\ln\lambda_2$ as labels instead of λ_1 and λ_2 and rewrite the relation in the additive form.

Each set of additive quantum numbers of a given Lie algebra opens the possibility of using them in describing the representations of the algebra. Practically, that means that we label simultaneously the generators of the algebra and basis states of representation spaces by the same set of additive quantum numbers.

Two sets of additive sl(3,C) quantum numbers are equivalent for a space V if they decompose V into the same subspaces:

$$V=\bigoplus_{\lambda\in S} V_\lambda, \qquad (1.2)$$

such that

$$[V_{\lambda_1},V_{\lambda_2}]\subseteq V_{\lambda_1+\lambda_2};\quad \lambda_1,\lambda_2,\lambda_1+\lambda_2\in S. \qquad (1.3)$$

In all our cases V is a finite-dimensional representation space of sl(3,C). In the case of the adjoint representation V could be taken to coincide with sl(3,C). In particular, if two sets of quantum numbers S_1 and S_2 provide the same decomposition of V, they are equivalent even if the actual values of the labels are different. Similarly different linear combinations of labeling operators always produce only equivalent sets of additive quantum numbers. Thus, for instance, the projection of isospin and hypercharge are not essentially different additive quantum numbers from, say, the projections of the u and v spins.

It may be astonishing but the question about nonequivalent sets S was apparently not investigated before Ref. 2. Therefore the description of representations using the uncommon sets S is undertaken here for a simple Lie algebra of rank >1 for the first time. The traditional representation theory of semisimple Lie algebras has an elegant and fairly uniform form for each type of algebra; perhaps that is why the alternatives were not pursued with any vigor until now. It is conceivable that in some problems the new bases and new quantum numbers will prove advantageous.

Nonequivalent sets of additive quantum numbers, which label a basis of the Lie algebra are in one-to-one correspondence with nonequivalent fine gradings of the Lie algebra.[1,2] For two equivalent gradings the related quantum numbers are equivalent in principle but practically can be defined to look very different. Moreover, equivalent gradings of a Lie algebra may not be equivalent gradings of a set of representations. Then what is the optimal choice? The answer undoubtedly depends on the set of representations under consideration.

This paper contains the first exploratory study of the alternative ways to describe sl(3,C) representations. We consider two of the four cases.[2] The third and fourth cases are the traditional theory (with its additive quantum numbers) and the $\wp_3$ grading of Refs. 3 and 4. The application of $\wp_3$ gradings to representation theory is somewhat different because it is based on a finite group rather than a Lie one as here. The two cases considered here have in common a prominent role played by a maximal simple subalgebra of rank 1 in sl(3,C), namely, sl(2,C) or o(3). These subalgebras have been used many times before but only here is the additive quantum number the subalgebra provides supplemented by a new additive quantum number, which is *not* an eigenvalue of the second diagonal generator of sl(3,C).

For convenience we identify the two cases considered here by the prominent simple subalgebra, either $sl(2,C)$ or $o(3)$. In both cases *part* of our construction is thus the reduction of $sl(3,C)$ to the subalgebra and among the generators of $sl(3,C)$ we find three generating the subalgebra. Among them must be at least one that is diagonalizable. Thus one of the two additive quantum numbers in each case is the eigenvalue of the diagonal generator of the subalgebra (projection of angular momentum, spin, isospin, etc.).

The traditional way to proceed in the $sl(2,C)$ case is to use the second diagonal generator of $sl(3,C)$, chosen in such a way that it commutes with the subalgebra and use its eigenvalue as the second additive quantum number. Such a choice is equivalent to the exploitation of the Cartan decomposition as the relevant grading of the algebra. Indeed, both additive quantum numbers are provided by two simultaneously diagonalizable generators of $sl(3,C)$. Also let us point out that the two additive quantum numbers in this case do not define a complete basis of $sl(3,C)$ nor of its representations in general. Indeed, the additive quantum numbers do not distinguish the diagonal generators, being zero for both of them. The supplementary labels used in the traditional way are then *nonadditive* quantum numbers related to the subgroup Casimir operators.

In the $o(3)$ case the first additive quantum number is the $o(3)$ weight. In fact the corresponding diagonal generator could be chosen to coincide with that of the $sl(2,C)$ case. The two cases would then differ in how the remaining two generators are chosen. For $o(3)$ that is done so that there is *no* other generator of $sl(3,C)$ commuting with $o(3)$. Then one faces what is often called the "missing label problem." Over the years a number of ways were invented to cope with it.[5–7]

The second labeling operator in the $O(3)$ contains the outer automorphism $\mathbf{Q}_3$ of $sl(3,\mathbf{C})$ often combined with an inner one. Its existence is well known and it was used in mathematics, for instance, in classifying the real forms of $sl(3,\mathbf{C})$, more precisely $sl(3,\mathbf{R})$ for $\mathbf{Q}_3$. In physics it is used for the Wigner–Inönü contractions of $sl(3,\mathbf{C})$ and for classification of rotational bands in $SU(3)$. Note, however, that even in those situations irreducible self-contragedient representations labeled by two equal integers (p,p) are always involved.

Abandoning the traditional scenarios above we define[1] a new quantum number of the additive type that can be used simultaneously with that defined by the corresponding subalgebra. Such a number is an eigenvalue of a finite transformation of $sl(3,C)$, which is not in the group $SL(3,C)$. The purpose of this paper is to investigate how the finite-dimensional representations of $sl(3,C)$ can be described in a basis labeled by the new pairs of additive quantum numbers. The main features of the description are the following: (1) The $sl(3,C)$ generators are labeled by two distinct additive quantum numbers; (2) one additive quantum number is of the "subalgebra type", the other is modular (mod N); (3) every representation space V is simultaneously graded with the Lie algebra and then decomposes as in (1.2) into N subspaces V_λ, $\lambda = 1,2, \dots ,N$ or $0,1, \dots ,N-1$; (4) a basis for each V_i has to be built resolving the degeneracies of the N-modular eigenvalues; (5) the choice of the modular eigenvalue is not unique. There are in general many ways to choose it, which are equivalent from the point of view of labeling the generators but nonequivalent from the point of view of representations.

II. REFINED sl(2,*C*) GRADING

Rather than the physicists' notation $SU(3) \supset SU(2)$, we use in this paper the nomenclature of the complexified Lie algebras, $sl(3,C) \supset sl(2,C)$. We hope this does not complexify the reading.

The two grading labels are described, for the $sl(3,C)$ algebra, in Refs. 1 and 2. The first, m, is the eigenvalue of the diagonal generator T_0 of the $sl(2,C)$ subalgebra; it takes integer values. The second, α, is the exponent of the imaginary unit i in the eigenvalue of an outer automorphism operator $\mathbf{Q}_2$; the eigenvalue takes the values ± 1, $\pm i$, so that α, an integer determined only modulo 4, may be limited to the values 0, 2, and ± 1. The labeling operators T_0 and $\mathbf{Q}_2$ commute with each other; they also commute with the $sl(2,C)$ Casimir operator T^2, the square Y^2 of the $sl(3,C)$ hypercharge generator, the second degree $sl(3,C)$ Casimir operator $C^{(2)}$ and the square $C^{(3)2}$ of the third degree $sl(3,C)$ Casimir operator. However, $\mathbf{Q}_2$ does not commute with $C^{(3)}$ nor with the hypercharge operator Y. Consequently, a graded basis state cannot belong in general to a single $sl(3,C)$ irreducible representation, nor have a definite hypercharge label. However, we can choose it to lie in the representation space of the pair of contragradient representations $(p,q) \oplus (q,p)$ [or in the single representation space (p,p) only if the representation is self-contragredient]; similarly, the graded state may involve only the equal and opposite hypercharges $\pm y$ (or a single value only if $y = 0$).

We review briefly the $sl(3,C)$ generators in our graded basis and the action of the grading operator $\mathbf{Q}_2$ on them. The matrix of a generator takes the form

$$\begin{pmatrix} \Omega & 0 \\ 0 & -\Omega^T \end{pmatrix},$$

suitable for acting on a state lying in the composite $sl(3,C)$ representation $(1,0) \oplus (0,1)$ in which the basis states are the usual η_1, η_2, η_3, η_1^*, η_2^*, η_3^*. Then, the eight generators are [we show only Ω, suitable for acting on $(1,0)$ states referred to η_1, η_2, and η_3]:

$$\mathbf{(2,0,0)} = T_0 = \begin{pmatrix} 1 & 0 & 0 \\ 0 & -1 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{(2,2,0)} = -T_+ = \begin{pmatrix} 0 & -1 & 0 \\ 0 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$
$$\mathbf{(2,-2,0)} = T_- = \begin{pmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix}, \quad \mathbf{(0,0,2)} = -3iY = \begin{pmatrix} -i & 0 & 0 \\ 0 & -i & 0 \\ 0 & 0 & 2i \end{pmatrix}, \tag{2.1}$$

$$(\mathbf{1,1,1}) = E_{13} - iE_{32} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & -i & 0 \end{pmatrix}, \quad (\mathbf{1,-1,1}) = E_{23} + iE_{31} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ i & 0 & 0 \end{pmatrix},$$

$$(\mathbf{1,1,-1}) = E_{13} + iE_{32} = \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ 0 & i & 0 \end{pmatrix}, \quad (\mathbf{1,-1,-1}) = E_{23} - iE_{31} = \begin{pmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ -i & 0 & 0 \end{pmatrix}.$$

The notation is $(\mathbf{t,m,\alpha})$. The label t is the sl$(2,C)$ representation label; it is not an additive grading label, but since the sl$(2,C)$ Casimir commutes with the grading operators T_0 and $\mathbf{Q}_2$, it is useful to retain it as an additional label; the generators are all distinguished without it, but it will be useful in labeling basis states of larger representations. In our basis, $\mathbf{Q}_2$ takes the form

$$\mathbf{Q}_2 = \begin{pmatrix} & 0 & & 0 & i & 0 \\ & & & -i & 0 & 0 \\ & & & 0 & 0 & i \\ 0 & i & 0 & & & \\ -i & 0 & 0 & & 0 & \\ 0 & 0 & i & & & \end{pmatrix}. \tag{2.2}$$

Then, the α labels in (2.1) are found from

$$\mathbf{Q}_2 \begin{pmatrix} \Omega & 0 \\ 0 & -\Omega^T \end{pmatrix} \mathbf{Q}_2^{-1} = i^\alpha \begin{pmatrix} \Omega & 0 \\ 0 & -\Omega^T \end{pmatrix}. \tag{2.3}$$

The label m arises from $[(\mathbf{2,0,0}),(\mathbf{t,m,\alpha})] = m(\mathbf{t,m,\alpha})$. The action of $\mathbf{Q}_2$ on the basis states of $(1,0) \oplus (0,1)$ is found from (2.2) to be

$$\mathbf{Q}_2\eta_1 = -i\eta_2^*, \quad \mathbf{Q}_2\eta_2 = i\eta_1^*, \quad \mathbf{Q}_2\eta_3 = i\eta_3^*,$$
$$\mathbf{Q}_2\eta_1^* = -i\eta_2, \quad \mathbf{Q}_2\eta_2^* = i\eta_1, \quad \mathbf{Q}_2\eta_3^* = i\eta_3. \tag{2.4}$$

The common eigenstates of T_0 and $\mathbf{Q}_2$ in this basis are found to be

$$|0,0,1\rangle = \eta_3 + \eta_3^*, \quad |0,0,-1\rangle = \eta_3 - \eta_3^*, \tag{2.5a}$$
$$|1,1,0\rangle = \eta_1 - i\eta_2^*, \quad |1,1,2\rangle = \eta_1 + i\eta_2^*, \tag{2.5b}$$
$$|1,1,0\rangle = \eta_2 + i\eta_1^*, \quad |1,-1,2\rangle = \eta_2 - i\eta_1^*. \tag{2.5c}$$

Again the notation is $|t,m,\alpha\rangle$. Note that (2.5a) and (2.5b) are the highest states of sl$(2,C)$ singlets and doublets, respectively.

Products of powers of the four states (2.5a) and (2.5b) of total degree n lie in the space V_n of representations (p,q) with $p+q=n$ (it is understood from now on that η_i, η_i^* stand for the traceless variables defined in the Appendix). It can be verified that they span the subspace of V_n with $m=t$. Thus a generating function for a basis of this subspace is given by

$$1/(1-U)^2(1-UT)^2. \tag{2.6}$$

The coefficient of U^nT^t in the power expansion of (2.6) is the multiplicity of the sl$(2,C)$ representation t in sl$(3,C)$ representations with $p+q=n$. The elements of the integrity basis corresponding to the respective denominator factors of (2.6) are $|0,0,1\rangle$, $|0,0,-1\rangle$, $|1,1,0\rangle$, $|1,1,2\rangle$. States for which the exponents of these elementary multiplets are, respectively, a,b,c,d, with $c+d=t$, and $a+b+c+d=n$ span that part of the sl$(3,C)$ representation space for which $p+q=n$, and $m=t$. States with $m<t$ are obtained by application of $T_- = (\mathbf{2,-2,0})$. The α value of an sl$(2,C)$ multiplet with exponents a, b, c, d is $a-b+2d$ mod 4. Consequently the decomposition (1.2) of a representation space V is of the form

$$V = \bigoplus_{m,\alpha} V_{m,\alpha}, \tag{2.7}$$

where the range of m depends on the particular case and the range of α is integers mod 4.

If eigenstates of $C^{(2)}$ and Y^2, as well as of T_0 and $\mathbf{Q}_2$, are desired, they can be projected from the states implied by (2.6); but it is complicated to specify from which states to make the projections in order to obtain a complete, independent set. Instead, we construct the states directly, starting with Gel'fand states.

A Gel'fand state $\left|{p,q \atop y,t,m}\right\rangle$ (unnormalized) with $m=t$ may be written[5]

$$\left|\begin{matrix} p, q \\ y, t, t \end{matrix}\right\rangle = \eta_1^{(1/2)(t+y)+(1/3)(p-q)} \times \eta_3^{(1/3)(2p+q)-(1/2)(t+y)} \times \eta_2^{*(1/2)(t-y)+(1/3)(q-p)} \times \eta_3^{*(1/3)(p+2p)+(1/2)(y-t)}. \tag{2.8}$$

The Gel'fand states with $-t \leqslant m < t$ are obtained from (2.8) by using T_-.

For the graded states belonging to $(p,q) \oplus (q,p)$, we may then write, in an obvious notation,

$$\left|\begin{matrix} p_>, p_< \\ y,t,m,\alpha \end{matrix}\right\rangle = \left|\begin{matrix} p_>, p_< \\ y,t,m \end{matrix}\right\rangle \pm i^{(2/3)(p_< - p_>) - y} \left|\begin{matrix} p_<, p_> \\ -y,t,m \end{matrix}\right\rangle \tag{2.9}$$

with $\alpha = p+q-t$ or $p+q-t+2$ mod 4 according to whether the positive or negative sign is taken in (2.9). For $p=q$, $y=0$, the states are

$$\left|\begin{matrix} p,p \\ 0,t,t,\alpha \end{matrix}\right\rangle = (\eta_1\eta_2^*)^{(1/2)t}(\eta_3\eta_3^*)^{p-(1/2)t} \tag{2.10}$$

with $\alpha = 2p - t$.

Finally the grading structure of the Lie algebra and its representation spaces can be expressed in terms of the subspaces $V_{m\alpha}$ of (2.7) as follows:

$$V_{m_1\alpha_1} \otimes V_{m_2\alpha_2} \subseteq V_{m_1+m_2,\alpha_1+\alpha_2}. \tag{2.11}$$

In particular, if a generator $(\mathbf{t,m_1,\alpha_1})$ replaces V_{m_1,α_1}, (2.11) becomes

$$(\mathbf{t,m_1,\alpha_1})V_{m_2\alpha_2} \subseteq V_{m_1+m_2,\alpha_1+\alpha_2}. \tag{2.12}$$

III. REFINED o(3) GRADING

The algebra chain sl$(3,C) \supset$ o(3) has been of importance in many physical models. It appears, for example, in the description of the simple three-dimensional harmonic oscillator,[6] in the Elliott model of the nucleus,[7] and in the rotational limit of the Interacting Boson Model.[8] It is known that the sl(3,C) generators split into the components L_μ ($\mu=0,\pm 1$) of a vector [i.e., the o(3) or angular momentum generators] and the components Q_μ ($\mu=0, \pm 1, \pm 2$) of a five-dimensional irreducible o(3) representation, the so-called quintet. Since the early days of sl(3,C)$\supset$o(3), it was known that the commutation relations satisfy

$$\begin{aligned}
&[L_\mu,L_\nu]\sim L_{\mu+\nu},\\
&[L_\mu,Q_\nu]\sim Q_{\mu+\nu},\\
&[Q_\mu,Q_\nu]\sim L_{\mu+\nu},
\end{aligned}\tag{3.1}$$

which displays the grading decomposition of sl(3,C) into the o(3) subspace V_0 generated by L's and the complementary subspace V_1 generated by Q's. Hence one has

$$(V_i,V_0)\subseteq V_{i+j},\quad i,j,i+j \bmod 2. \tag{3.1'}$$

Clearly, the index μ refers to the projection of the angular momentum and provides an additive quantum number and hence an additive grading of the Lie algebra sl(3,C). What has not been stressed in the literature is the fact that this grading can be refined as a result of the splitting of the sl(3,C) Lie algebra to L and Q components and the relations (3.1). In fact, in Refs. 1 and 2, it is shown that this splitting is exactly one of the four fine gradings of the Lie algebra of sl(3,C). The sl(3,C) generators can then actually be labeled by two numbers (m,β). The first, m, is the eigenvalue of the diagonal generator L_0 of the o(3) subalgebra and takes integer values. It is equal to the index μ in (3.1). The second, β, is the exponent of i in the eigenvalue of an outer automorphism operator $\mathbf{Q}_3$ (not to be confused with one of the components $\mathbf{Q}_\mu$ of the quintet); the eigenvalues of $\mathbf{Q}_3$ on the adjoint representation take the value ± 1, so that β is 0 or 2. Clearly, although β is a $\mathbf{Z}_4$ number, it defines only a $\mathbf{Z}_2$ grading for the generators in the adjoint representation. Here, the L_μ's have $\beta=0$, and the Q_μ's have $\beta=2$. The labeling operators L_0 and $\mathbf{Q}_3$ commute with each other and with the total angular momentum operator L^2. Therefore l is still a good quantum number when we want to label basis vectors of sl(3,C) representations by means of their L_0 and $\mathbf{Q}_3$ eigenvalues. However, just as in the sl(2,C) case, $\mathbf{Q}_3$ commutes with $C^{(2)}$ but not with $C^{(3)}$. As a consequence, irreducible sl(3,C) representations do not in general allow a consistent refined o(3) grading. Only the self-contragredient representations $(p,q)\oplus(q,p)(p\neq q)$ or (p,p) have graded refined o(3) basis states.

Let us first recall the explicit form of the sl(3,C) generators in the graded basis. As in Sec. II, they can be written in matrix notation

$$\begin{pmatrix}\Omega & 0\\ 0 & -\Omega^T\end{pmatrix}$$

and have a natural action on the composite sl(3,C) representation $(1,0)\oplus(0,1)$ in which the basis states are now denoted by η_1, η_0, η_{-1}, η_1^*, η_0^*, and η_{-1}^*. The Ω matrices for the eight generators are

$$\begin{gathered}
\mathbf{(1,0,0)}=\begin{pmatrix}1&0&0\\0&-1&0\\0&0&0\end{pmatrix},\quad \mathbf{(1,-1,0)}=\begin{pmatrix}0&0&0\\1&0&0\\0&1&0\end{pmatrix},\\
\mathbf{(1,1,0)}=\begin{pmatrix}0&1&0\\0&0&1\\0&0&0\end{pmatrix},\\
\mathbf{(2,-2,2)}=\begin{pmatrix}0&0&0\\0&0&0\\1&0&0\end{pmatrix},\\
\mathbf{(2,-1,2)}=\begin{pmatrix}0&0&0\\1&0&0\\0&-1&0\end{pmatrix},\\
\mathbf{(2,0,2)}=\begin{pmatrix}1&0&0\\0&-2&0\\0&0&1\end{pmatrix},\\
\mathbf{(2,1,2)}=\begin{pmatrix}0&1&0\\0&0&-1\\0&0&0\end{pmatrix},\quad \mathbf{(2,2,2)}=\begin{pmatrix}0&0&1\\0&0&0\\0&0&0\end{pmatrix}.
\end{gathered}\tag{3.2}$$

The notation is $(\mathbf{l,m,\beta})$. The label l is the o(3) representation label: it is unnecessary for the labeling of the generators and is retained in (3.2) only for convenience. The grading of the sl(3,C) Lie algebra follows from

$$[(\mathbf{l,m,\beta}),(\mathbf{l',m',\beta'})]=\text{coeff}(\mathbf{l'',m+m',\beta+\beta'}). \tag{3.3}$$

The additive labels m and β arise from

$$[\mathbf{(1,0,0)},(\mathbf{l,m,\beta})]=m(\mathbf{l,m,\beta}) \tag{3.4}$$

and

$$\mathbf{Q}_3\begin{pmatrix}\Omega&0\\0&-\Omega^T\end{pmatrix}\mathbf{Q}_3^{-1}=i^\beta\begin{pmatrix}\Omega&0\\0&-\Omega^T\end{pmatrix}, \tag{3.5}$$

where

$$\mathbf{Q}_3=\begin{pmatrix}&&&0&0&i\\&&&0&-i&0\\&&&i&0&0\\0&0&i&&&\\0&-i&0&&&\\i&0&0&&&\end{pmatrix}. \tag{3.6}$$

Hence the action of $\mathbf{Q}_3$ on the basis states of $(1,0)\oplus(0,1)$ is

$$\begin{aligned}
&\mathbf{Q}_3\eta_1=i\eta_{-1}^*,\quad \mathbf{Q}_3\eta_0=-i\eta_0^*,\quad \mathbf{Q}_3\eta_{-1}=i\eta_1^*,\\
&\mathbf{Q}_3\eta_1^*=i\eta_{-1},\quad \mathbf{Q}_3\eta_0^*=-i\eta_0,\quad \mathbf{Q}_3\eta_{-1}^*=i\eta_1.
\end{aligned}\tag{3.7}$$

The common eigenstates of L_0 and $\mathbf{Q}_3$ are now given by

$$\begin{aligned}
&|1,-1,1\rangle=\eta_{-1}+\eta_1^*,\quad &&|1,-1,-1\rangle=\eta_{-1}-\eta_1^*,\\
&|1,0,1\rangle=\eta_0-\eta_0^*,\quad &&|1,0,-1\rangle=\eta_0+\eta_0^*,\\
&|1,1,1\rangle=\eta_1+\eta_{-1}^*,\quad &&|1,1,-1\rangle=\eta_1-\eta_{-1}^*;
\end{aligned}\tag{3.8}$$

the notation is $|\mathbf{l,m,\beta}\rangle$, where $L_0|\mathbf{l,m,\beta}\rangle=m|\mathbf{l,m,\beta}\rangle$ and $\mathbf{Q}_3|\mathbf{l,m,\beta}\rangle=i^\beta|\mathbf{l,m,\beta}\rangle$. Note that in the representation

(3.8) only $\pm i$ appear as $\mathbf{Q}_3$ eigenvalues. This implies that (3.8) is $\mathbf{Z}_2$ graded by the label β, and that

$$(l,\mathbf{m},\beta)\,|l',\mathbf{m}',\beta'\rangle = \mathrm{coeff}|l'',\mathbf{m}+\mathbf{m}',\beta+\beta'\rangle. \quad (3.9)$$

A generating function for basis states of higher sl(3,C) representations with proper grading labels is given by

$$(1+U^2L/(1-U^2)(1-UL)^2. \qquad (3.10)$$

The coefficient of $U^n L_l$ in the expansion of (3.10) is the multiplicity of the o(3) representation l in the space of representations (p,q) with $p+q=n$ and $m=l$.

The elements of the integrity basis implied by (3.10) are

$$\begin{aligned}
&U^2L \to \eta_0\eta^*_{-1} + \eta_1\eta^*_0,\\
&(U^2)_1 \to \eta_0^2 - 2\eta_1\eta_{-1} + \eta_0^{*2} - 2\eta_1^*\eta^*_{-1},\\
&(U^2)_2 \to \eta_0^2 - 2\eta_1\eta_{-1} - \eta_0^{*2} + 2\eta_1^*\eta^*_{-1},\\
&(UL)_1 \to \eta_1 + \eta^*_{-1}, \quad (UL)_2 \to \eta_1 - \eta^*_{-1}. \qquad (3.11)
\end{aligned}$$

The eigenvalues of $\mathbf{Q}_3$ on the five elementary states (3.11) are $+1, -1, +1, +i, -i$, respectively. Contrary to what one might expect this does not lead to a $\mathbf{Z}_4$ but to a $\mathbf{Z}_2$ grading, since on a single representation $(p,q) \oplus (q,p)$, $(p \neq q)$, or (p,p), it still has only two different eigenvalues: either $+1$ and -1 or else $+i$ and $-i$. Products of powers of the elements (3.11) with respective exponents z, a, b, c, d $(z=0,1)$, with $2z+2a+2b+c+d=n$ and $z+c+d=l$, span that part of the sl(3,C) representation space for which $p+q=n$ and $m=l$. Basis states with $m<l$ are obtained by applying $L_- = \mathbf{(1,-1,0)}$. The product states formed from (3.11) have a good grading label $\beta = 2a+c-d$ mod 4. It follows that the $\mathbf{Q}_3$ eigenvalues are ± 1 when $p+q$ is even and $\pm i$ when $p+q$ is odd.

Just as in the previous case, one can proceed in two ways. This first is to use the states in the explicit expansion of (3.10), which are eigenstates of L^2, L_0, and $\mathbf{Q}_3$. The graded states of $(p,q) \oplus (q,p)$, $(p \neq q)$, or (p,p) are then projections of the product states. However, it is not easy to specify which product states in (3.10) will give an independent set of basis states after projection. An alternative way is to construct the basis states explicitly, starting for example with the so-called stretched states for sl(3,C) $\supset$ o(3). In terms of the η and η^* variables, a stretched state with $m=l$ may be written as[5]

$$\left|\begin{matrix}p,q\\ l_1,l,l\end{matrix}\right\rangle = \eta_1^{l_1}\eta^{*\,l-l_1}_{-1}(2\eta_1\eta_{-1}-\eta_0^2)^{(1/2)(p-l_1)} \times (2\eta_1^*\eta^*_{-1}-\eta_0^{*2})^{(1/2)(q-l+l_1)} \qquad (3.12)$$

for $p+q-l$ even, and as

$$\left|\begin{matrix}p,q\\ l_1,l,l\end{matrix}\right\rangle = \eta_1^{l_1-1}\eta^{*\,l-l_1}_{-1}(2\eta_1\eta_{-1}-\eta_0^2)^{(1/2)(p-l_1)} \times (2\eta_1^*\eta^*_{-1}-\eta_0^{*2})^{(1/2)(q-l+l_1-1)} \times (\eta_1\eta_0^* + \eta_0\eta^*_{-1}), \qquad (3.13)$$

for $p+q-l$ odd. Using the notation $p_> = \max\{p,q\}$, $p_< = \min\{p,q\}$, we can now give an expression for the graded basis states of the representation $(p,q) \oplus (q,p)$, $(q \neq p)$, or (p,p):

$$\left|\begin{matrix}p_>,p_<\\ l_1,l,m,\beta_\pm\end{matrix}\right\rangle = \left|\begin{matrix}p_>,p_<\\ l_1,l,m\end{matrix}\right\rangle \pm \left|\begin{matrix}p_>,p_<\\ l-l_1,l,m,\end{matrix}\right\rangle, \qquad (3.14)$$

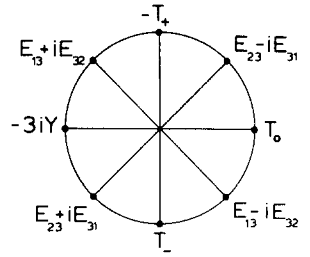


FIG. 1. The generators (2.1) in a $\mathbf{Z}_8$ weight space.

where $\beta_+ = 2l-p-q$ mod 4 and $\beta_- = \beta_+ + 2$ mod 4. The $\mathbf{Q}_3$ eigenvalues are then given by i^β. Note that when $p=q$, (3.14) will give a basis for the irreducible sl(3,C) representation (p,p); then when $l=2l_1$, only the state with the plus sign survives in (3.14). As in the previous case, these are the only irreducible sl(3,C) representations for which the grading is possible.

The grading properties are expressed by Eqs. (2.11) and (2.12), with α replaced by β, when the additive quantum numbers m and β are defined by (3.4) and (3.5), respectively.

IV. CONCLUDING REMARKS

It has been observed in Ref. 2 that the grading presented in Sec. II is equivalent to a $\mathbf{Z}_8$ grading of the Lie algebra sl(3,C). Indeed, the basis elements $(\mathbf{t},\mathbf{m},\alpha)$ given in (2.1) can be renamed as follows:

$$\begin{aligned}
&(0) = \mathbf{(2,0,0)}, (1) = \mathbf{(1,-1,-1)}, (2) = \mathbf{(2,2,0)},\\
&(3) = \mathbf{(1,1,-1)}, (4) = \mathbf{(0,0,2)}, (5) = \mathbf{(1,-1,1)}, \qquad (4.1)\\
&(6) = \mathbf{(2,-2,0)}, (7) = \mathbf{(1,1,1)}.
\end{aligned}$$

The fact that this provides a $\mathbf{Z}_8$ grading follows from

$$[(a),(b)] = \mathrm{coeff}(a+b), \quad a,b,a+b \in \mathbf{Z}_8. \qquad (4.2)$$

The $\mathbf{Z}_8$ label a is again related to an outer automorphism operator $\mathbf{Q}_8$, explicitly given by

$$\mathbf{Q}_8 = \begin{bmatrix} & & & 0 & \theta^7 & 0\\ & & & \theta & 0 & 0\\ & & & 0 & 0 & i\\ 0 & \theta^5 & 0 & & & \\ \theta^3 & 0 & 0 & & & \\ 0 & 0 & i & & & \end{bmatrix}, \qquad (4.3)$$

where

$$\theta = e^{2\pi i/8} = (1+i)/\sqrt{2}. \qquad (4.4)$$

It grades the basis elements (4.1) by

$$\mathbf{Q}_8\begin{pmatrix}(a) & 0\\ 0 & -(a)^T\end{pmatrix}\mathbf{Q}_8^{-1} = \theta^a\begin{pmatrix}(a) & 0\\ 0 & -(a)^T\end{pmatrix}. \qquad (4.5)$$

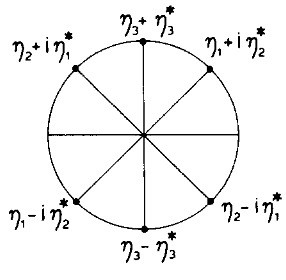


FIG. 2. The quark (η_1,η_2,η_3) and antiquark $(\eta_1^*,\eta_2^*,\eta_3^*)$ states in the $\mathbf{Z}_8$ weight space.

TABLE I. Upper part of the commutation table of the sl(3,C) generators (2.1) modified as in (4.9).

	(0,0)	**(2,0)**	**(− 2,0)**	**(0,2)**	**(− 1,1)**	**(1,1)**	**(− 1, − 1)**	**(1, − 1)**
(0,0)	0	2	− 2	0	− 1	1	− 1	1
(2,0)		0	1	0	− 1	0	1	0
(− 2,0)			0	0	0	− 1	0	1
(0,2)				0	3	− 3	− 3	3
(− 1,1)					0	− 1	− 2	− 1
(1,1)						0	1	− 2
(− 1, − 1)							0	1
(1, − 1)								0

It is interesting to see now what effect this $\mathbf{Z}_8$ grading has on the representations. For the $(1,0) \oplus (0,1)$ (i.e., quark and antiquark states) representation, the following basis vectors:

$$|1\rangle = \eta_1 + i\eta_2^*, \quad |2\rangle = \eta_3 + \eta_3^*, \quad |3\rangle = \eta_2 + i\eta_1^*,$$
$$|5\rangle = \eta_1 - i\eta_2^*, \quad |6\rangle = \eta_3 - \eta_3^*, \quad |7\rangle = \eta_2 - i\eta_1^*, \tag{4.6}$$

satisfy

$$\mathbf{Q}_8|b\rangle = \theta^b|b\rangle, \quad b = 1,2,3,5,6,7. \tag{4.7}$$

This implies the grading for the representation (4.6):

$$(a)|b\rangle \sim |a+b\rangle, \quad (a,b,a+b\in\mathbf{Z}_8). \tag{4.8}$$

Just as for the Cartan decomposition, the present graded basis of generators and graded states can be described in weight space. This gives rise to unusual pictures, manifesting a different sort of symmetry and possessing very elegant properties. In Fig. 1, we give the "weights" for the generators, following from (2.1) and (4.1); the "weights" of the quark and antiquark states (4.6) are depicted in Fig. 2. From (4.8), it follows, for example, that acting with T_+ on the states of Fig. 2 rotates the states by an angle $\pi/2$; acting with Y rotates the states by an angle π, etc. Also note that the commutator of two elements in Fig. 1 is proportional to the element obtained by adding the two angles of the original elements. This case illustrates very well that, although the gradings of the Lie algebra may be equal as in (2.1) and (4.1), their effect on graded representations is quite different. For example, the $(2,0) \oplus (0,2)$ has two multiplicity 2 subspaces when graded by the refined sl(3,C) labels (m,α) of Sec. II and has four multiplicity 2 subspaces when graded only by the $\mathbf{Z}_8$ quantum number (b).

The Cartan decomposition has the property of leading to integer structure constants, provided a proper normalization is chosen. The two different gradings discussed in this paper share this property. Choosing the following notation $(\mathbf{m},\alpha)$ and normalization for the basis elements (2.1),

$$\mathbf{(0,0) = (2,0,0), \quad (2,0) = -(2,2,0), \quad (-2,0) = (2,-2,0),}$$
$$\mathbf{(0,2) = (0,0,2),}$$
$$\mathbf{(-1,1) = (1,-1,1), \quad (1,1) = -(1,1,1),}$$
$$\mathbf{(-1,-1)} = -i\mathbf{(1,-1,-1)}, \quad \mathbf{(1,-1)} = -i\mathbf{(1,1,-1)} \tag{4.9}$$

the constants in the commutation relations

$$[(\mathbf{m},\alpha),(\mathbf{m}',\alpha')] = \mathrm{const}(\mathbf{m}+\mathbf{m}',\alpha+\alpha') \tag{4.10}$$

have the integer values given in Table I.

Similarly, for the basis (3.2), the structure constants in the relation (3.3) have the integer values shown in Table II [we suppress the label l for the generators and use only the additive grading labels $(\mathbf{m},\beta)$].

When a generator acts on a graded state, whether of refined sl(2,C) or refined so(3) type, the result is always a linear combination of graded states satisfying the usual additivity rules for the grading labels. The generator matrix elements are closely related to those for the conventional Gel'fand basis, or stretched basis in the respective cases. We give one example for each case, with the states lying in the composite representation $(1,0) \oplus (0,1)$.

For refined sl(3,C), we have

$$\mathbf{(1,1,1)}|0,0,-1\rangle$$
$$= (\eta_1\partial_3 - i\eta_3\partial_2 - \eta_3^*\partial_1^* + i\eta_2^*\partial_3^*)(\eta 3 - \eta_3^*)$$
$$= \eta_1 - i\eta_2^* = |1,1,0\rangle, \tag{4.11}$$

and for refined o(3)

$$\mathbf{(2,-1,2)}|1,1,1\rangle$$
$$= (\eta_0\partial_1 - \eta_{-1}\partial_0 - \eta_1^*\partial_0^* + \eta_0^*\partial_{-1}^*)(\eta_1 + \eta_{-1}^*)$$
$$= \eta_0 + \eta_0^* = |0,0,-1\rangle. \tag{4.12}$$

TABLE II. Upper part of the commutation table of the sl(3,C) generators (3.2).

	(0,0)	**(− 1,0)**	**(1,0)**	**(− 2,2)**	**(− 1,2)**	**(0,2)**	**(1,2)**	**(2,2)**
(0,0)	0	− 1	1	− 2	− 1	0	1	2
(− 1,0)		0	− 1	0	2	3	− 1	− 1
(1,0)			0	1	1	− 3	− 2	0
(− 2,2)				0	0	0	1	− 1
(− 1,2)					0	3	− 1	1
(0,2)						0	3	0
(1,2)							0	0
(2,2)								0

ACKNOWLEDGMENTS

We are grateful to H. Zassenhaus for helpful discussions and comments.

One of us (J.V.d.J.) thanks NSERC for a visiting scientist grant, and the British SERC for a Research Fellowship GR/D 49909.

APPENDIX

In order to extract the traceless (i.e., orthogonal to any state containing the scalar $B=\eta_1\eta_1^*+\eta_2\eta_2^*+\eta_3\eta_3^*$ as a factor, and therefore transforming by a lower representation) part of any state, we adapt a prescription given by Lohe[9] for orthogonal groups, and make the replacements

$$\eta_i\to\eta_i'=\eta_i-B(N+3)^{-1}\partial_i^*,$$
$$\eta_i^*\to\eta_i'^*=\eta_i^*-B(N+3)^{-1}\partial_i \tag{A1}$$

in any polynomial in the starred and unstarred variables; $N=\Sigma_i(\eta_i\partial_i+\eta_i^*\partial_i^*)$ is the total degree of its operand.

The criterion for a traceless state $|\psi\rangle$ is

$$\langle B\varphi|\psi\rangle=0, \tag{A2}$$

where φ is an arbitrary state. Equation (A2) can be written

$$\langle\varphi|B^\dagger\psi\rangle=0, \tag{A3}$$

or, since φ is arbitrary,

$$B^\dagger\psi=(\partial_1\partial_1^*+\partial_2\partial_2^*+\partial_3\partial_3^*)\psi=0. \tag{A4}$$

We show by induction that a product of primed variables is traceless by calculating

$$B^\dagger\eta_i'=(\eta_i-B(N+5)^{-1}\partial_i^*)B^\dagger. \tag{A5}$$

Thus, if $B^\dagger$ annihilates a state, it also annihilates the state multiplied by η_i' (or $\eta_i'^*$).

It is also seen by actual calculation that the η_i' and $\eta_j'^*$ mutually commute, so the order of applying them is immaterial.

The primed and unprimed variables transform in the same way under $\mathrm{sl}(3,C)$ transformations.

[1]J. Patera and H. Zassenhaus, "On Lie gradings I," Linear Algebra Appl. **112**, 87 (1989); "On Lie gradings II," in preparation.

[2]J. Patera, "The four sets of additive quantum numbers of SU(3)," J. Math. Phys. **30**, 2756 (1989).

[3]J. Patera and H. Zassenhaus, "The Pauli matrices in n dimensions and finest gradings of simple Lie algebras of type A_{n-1}," J. Math. Phys. **29**, 665 (1988).

[4]C. J. Cummins, J. Patera, and R. T. Sharp, "SL(3,C) $\supset\wp_3$ basis states and generator matrix elements," in preparation.

[5]M. Moshinsky, J. Patera, R. T. Sharp, and P. Winternitz, "Everything you always wanted to know about SU(3)$\supset$O(3)," Ann. Phys. **95**, 139 (1975); R. T. Sharp, H. C. von Baeyer, and S. C. Pieper, "Polynomial bases and Wigner coefficients for SU(3)$\supset R_3$," Nucl. Phys. A **127**, 513 (1969).

[6]M. Moshinsky, *The Harmonic Oscillator in Modern Physics* (Gordon and Breach, New York, 1969).

[7]J. P. Elliott, "Collective motion in the nuclear shell model. I. Clasification scheme for states mixed configurations; II. The introduction of intrinsic wave functions," Proc. Roy. Soc. London A **245**, 128, 562 (1959).

[8]A. Arima and F. Iachello, "Interacting boson model of collective nuclear states. II. The rotational limit," Ann. Phys. **111**, 201 (1978).

[9]M. A. Lohe, Ph.D. thesis, University of Adelaide, 1974; see also M. A. Lohe and C. A. Hurst, "The boson calculus for the orthogonal and symplectic groups," J. Math. Phys. **12**, 1882 (1971).

Lie algebras associated with scalar second-order ordinary differential equations

F. M. Mahomed and P. G. L. Leach
Centre for Nonlinear Studies and Department of Computational and Applied Mathematics, University of the Witwatersrand, Johannesburg, P. O. Wits, 2050 South Africa

(Received 18 November 1988; accepted for publication 28 June 1989)

Second-order ordinary differential equations are classified according to their Lie algebra of point symmetries. The existence of these symmetries provides a way to solve the equations or to transform them to simpler forms. Canonical forms of generators for equations with three-point symmetries are established. It is further shown that an equation cannot have exactly $r \in \{4,5,6,7\}$ point symmetries. Representative(s) of equivalence class(es) of equations possessing $s \in \{1,2,3,8\}$ point symmetry generator(s) are then obtained.

I. INTRODUCTION

It is by now well established that when one realizes a real low-dimensional Lie algebra in terms of vector fields in two coordinates more than one canonical form may occur. A familiar example of this is found in the two canonical forms of generators obtained for each of the two real two-dimensional Lie algebras $2A_1$ and A_2 given in Lie.[1]

Lie showed that if a second-order ordinary differential equation $\ddot{q} = E(t,q,\dot{q})$ admits a two-dimensional algebra (Abelian algebra $2A_1$ or the solvable algebra A_2) of point symmetries, then its point symmetries G_1 and G_2 can be either connected [i.e., there exists a function $\rho(t,q)$ such that $G_2 = \rho(t,q)G_1$] or unconnected [i.e., for any function $\psi(t,q)G_2 \neq \psi(t,q)G_1$]. Thus he was led to distinguish between connected and unconnected operators for each of the algebras $2A_1$ and A_2, respectively. Consequently he had four cases (see Ref. 2) to consider (Types I–IV, see Ref. 3). Lie deduced that, if a second-order equation admits a two-dimensional algebra with operators G_1 and G_2 satisfying $G_2 = \rho(t,q)G_1$ for a suitable function ρ (i.e., Type II or IV of Ref. 3, then it is linearizable via a point transformation, i.e., it has SL(3,$\mathbb{R}$) symmetry (see Ref. 2).

In Ref. 2 second-order equations having two commuting unconnected point symmetries (Type I) were investigated for linearizability. It was shown that such an equation is linearizable provided the regular point transformation that brings these symmetries into their canonical form reduces the equation to one which is at most cubic in the first derivative. Accordingly, for equations of this type, the complete symmetry group SL(3, $\mathbb{R}$) and the corresponding linearizing point transformations were obtained.

The study of second-order equations admitting two noncommuting unconnected point symmetries (Type III), with a view to linearization, was undertaken in Ref. 3. It was found that such an equation has the SL(3,$\mathbb{R}$) symmetry group provided the point transformation that casts these symmetries to their canonical form reduces the equation to the form

$$t\ddot{q} = a\dot{q}^3 + b\dot{q}^2 + (1 + b^2/3a)\dot{q} + b/3a + b^3/27a^2, \tag{i}$$

where a ($\neq 0$) and b are arbitrary constants.

A necessary condition for a second-order equation to admit the sl(3,$\mathbb{R}$) algebra (see Refs. 2 and 3) is that it be of the form

$$\ddot{q} = \mathscr{A}(t,q)\dot{q}^3 + \mathscr{B}(t,q)\dot{q}^2 + \mathscr{C}(t,q)\dot{q} + \mathscr{D}(t,q), \tag{ii}$$

where the functions $\mathscr{A}$, $\mathscr{B}$, $\mathscr{C}$, and $\mathscr{D}$ are analytic. Sufficient conditions for a second-order equation to admit the sl(3,$\mathbb{R}$) algebra are given by (see Ref. 3)

$$3\mathscr{A}_{tt} + 3\mathscr{A}_t\mathscr{C} - 3\mathscr{A}_q\mathscr{D} + 3\mathscr{A}\mathscr{C}_t + \mathscr{C}_{qq} - 6\mathscr{A}\mathscr{D}_q + \mathscr{B}\mathscr{C}_q - 2\mathscr{B}\mathscr{B}_t - 2\mathscr{B}_{tq} = 0, \tag{iii}$$

$$6\mathscr{A}_t\mathscr{D} - 3\mathscr{B}_q\mathscr{D} + 3\mathscr{A}\mathscr{D}_t + \mathscr{B}_{tt} - 2\mathscr{C}_{tq} - 3\mathscr{B}\mathscr{D}_q + 3\mathscr{D}_{qq} + 2\mathscr{C}\mathscr{C}_q - \mathscr{C}\mathscr{B}_t = 0, \tag{iv}$$

where the suffices refer to partial derivatives.

In Ref. 2 and 3 we have treated all four cases (Types I–IV) of second-order equations admitting two-dimensional algebras of point symmetries which are linearizable via a point transformation. Thus, if an equation of the form (ii) passes the linearization test, i.e., conditions (iii) and (iv) hold, then one requires only two-point symmetries of the equation to obtain a linearizing point transformation for the equation.

In this paper we thoroughly investigate second-order equations admitting real Lie algebras of dimension two and higher (and at most eight which is the maximum dimension for such equations). We give a complete and rigorous treatment of three-dimensional Lie algebra realizations in terms of vector fields defined on the plane (see Sec. II) since in this respect there have been omissions in the works of Lie. Lie did not take into account the realizations of the family of algebras $A^b_{3,7}$ (b is a real parameter) and the algebras $A_{3,6}$ and $A_{3,9}$ (see Table I for their commutation relations). The canonical realizations in two coordinates obtained for the low-dimensional Lie algebras (see Tables I–III) are utilized systematically to derive all equivalence classes of second-order equations (see Sec. III) which admit point symmetry algebras. More precisely, we obtain five representatives [see (45)] of equivalence classes of second-order equations having exactly three-point symmetries. It is also deduced that second-order equations possessing exactly two-point symmetries can belong to one of two equivalence classes [see (46)]. Moreover, we prove, in Sec. II, that a second-order equation cannot admit exactly $r \in \{4,5,6,7\}$ point symme-

TABLE I. Lie algebras of dimension 3.

Algebra	Nonzero commutation relations
$3A_1$	
$A_1 \oplus A_2$	$[G_1,G_3] = G_1$
$A_{3,1}$ (Weyl)	$[G_2,G_3] = G_1$
$A_{3,2}$	$[G_1,G_3] = G_1, \quad [G_2,G_3] = G_1 + G_2$
$A_{3,3}$ $(D \otimes_s T_2)$	$[G_1,G_3] = G_1, \quad [G_2,G_3] = G_2$
$A_{3,4}$(E(1,1))	$[G_1,G_3] = G_1, \quad [G_2,G_3] = -G_2$
$A^a_{3,5}$ $(0<\lvert a\rvert<1)$	$[G_1,G_3] = G_1, \quad [G_2,G_3] = aG_2$
$A_{3,6}$(E(2))	$[G_1,G_3] = -G_2, \quad [G_2,G_3] = G_1$
$A^b_{3,7}$ $(b>0)$	$[G_1,G_3] = bG_1 - G_2, \quad [G_2,G_3] = G_1 + bG_2$
$A_{3,8}$(SL(2,ℝ))	$[G_1,G_2] = G_1, [G_2,G_3] = G_3, \quad [G_3,G_1] = -2G_2$
$A_{3,9}$(SO(3))	$[G_1,G_2] = G_3, [G_2,G_3] = G_1, \quad [G_3,G_1] = G_2$

tries. This in effect means that a second-order equation can admit exactly one of 0,1,2,3, or 8 point symmetries.

II. LIE ALGEBRA REALIZATIONS IN TWO COORDINATES

The Lie algebra classification used is the Mubarakzyanov classification given in Patera *et al.*[4] For the three- and four-dimensional algebras our list also includes the decomposable algebras. The notation is the same as given in Ref. 4. Thus when referring to $A^a_{r,j}$ we simply mean the jth algebra of dimension r. The superscript(s), if any, indicate(s) the parameter(s) on which the algebra depends. The range of parameters is restricted to avoid double counting and algebraic sums of lower algebras. Assignment of specific values to a parameter singles out special algebras, within a family, which are well-known and which in some cases result in the linearization of the associated second-order equation. There are 11 Lie algebras of dimension three[4] (decomposable and indecomposable) two of which depend on parameters. For convenience and easy reference we present their commutation relations in a convenient basis $\{G_i : i = 1,3\}$ in Table I.

The decomposable Lie algebras are the Abelian algebra $3A_1$ and the non-Abelian one $A_1 \oplus A_2$. In many cases we have indicated the corresponding Lie groups in parentheses. These are the Weyl group, the semidirect product of dilations and translations $D \otimes_s T_2$, the Euclidean group E(2), the pseudo-Euclidean group E(1,1), the special linear group SL(2,ℝ), and the special orthogonal group SO(3).

The following easily verified identities will be used in the proof of theorems. Let G_1,G_2,G_3 be operators of the form

$$G = \xi(t,q)\frac{\partial}{\partial t} + \eta(t,q)\frac{\partial}{\partial q}.$$

Then (a) $[G_1,G_2] = (-G_2\rho)G_2$, if $G_1 = \rho(t,q)G_2$ for a suitable function ρ, (b) $[G_1,G_3] = (G_1\psi)G_2 + \psi[G_1,G_2]$, if $G_3 = \psi(t,q)G_2$ for a suitable function ψ.

Theorem 1: A second-order ordinary differential equation does not admit the Abelian Lie algebra, $3A_1$.

Proof: Suppose a general second-order equation of the form $\ddot q = H(\dot q,q,t)$ has the Abelian algebra. The generators of symmetry G_i $(i = 1,3)$ then satisfy the Abelian commutation relations $3A_1$. This being the case, we firstly show that one cannot have $G_1 = \rho(t,q)G_2$ or $G_3 = \psi(t,q)G_2$ (for any nonconstant functions ρ and ψ). Assume the contrary is true. Then for the case $G_1 = \rho G_2$, the first and last commutators of $3A_1$ imply

$$G_2\rho = \xi_2\frac{\partial\rho}{\partial t} + \eta_2\frac{\partial\rho}{\partial q} = 0,$$

$$G_3\rho = \xi_3\frac{\partial\rho}{\partial t} + \eta_3\frac{\partial\rho}{\partial q} = 0.$$

As ρ cannot be a constant, at least one of $\partial\rho/\partial t$ or $\partial\rho/\partial q$ is nonzero. This in turn leads to $\xi_2\eta_3 - \xi_3\eta_2 = 0$ and consequently

$$G_3 = (\xi_3/\xi_2)G_2.$$

Invoking the second commutator, $[G_2,G_3] = 0$, we obtain

$$G_2(\xi_3/\xi_2) = 0.$$

Therefore we must have $\xi_3/\xi_2 = f(\rho)$ for some function f (since also $G_2\rho = 0$). Thus eventually we have

$$G_1 = \rho G_2, \quad G_3 = f(\rho)G_2 \text{ with } G_2\rho = 0.$$

Transforming G_2 to $\bar G_2 = \partial/\partial Q$ and solving $\bar G_2\bar\rho = 0$ $[\bar\rho = \bar\rho(Q,T)$ where $Q = Q(t,q)$ and $T = T(t,q)$ are the transformed coordinates] we obtain $\bar\rho = g(T)$ for some function g. Without loss of generality we may set $g(T) = T$ making the generators appear as (cf. Lie[1])

$$\bar G_1 = T\frac{\partial}{\partial Q}, \quad \bar G_2 = \frac{\partial}{\partial Q}, \quad \bar G_3 = f(T)\frac{\partial}{\partial Q}. \tag{1}$$

Expressing the invariance of the differential equation for Q with respect to (1) results in f being linear in T contradicting the linear independence of the G_i's. The preceding argument applies equally to the case $G_3 = \psi G_2$ since G_1 and G_3 play interchangeable roles. It follows therefore that $G_1 \neq \rho G_2$ and $G_3 \neq \psi G_2$ for any functions ρ and ψ. This, however, leads to (upon reducing G_1, G_2 to canonical form)

$$\bar G_1 = \frac{\partial}{\partial T}, \quad \bar G_2 = \frac{\partial}{\partial Q}, \quad \bar G_3 = \alpha\frac{\partial}{\partial T} + \beta\frac{\partial}{\partial Q}$$

(α, β constants and $\alpha \neq 0$) which obviously are linearly dependent. ◇

An immediate consequence of the above theorem is that a differential equation does not admit a Lie algebra which contains the Abelian three-dimensional algebra $3A_1$, as a subalgebra. As a result we need not consider all the real Lie algebras of dimension four (24 if we include, as we have, the decomposable ones). Indeed we have tabulated only nine (see Table II below). Of the five-dimensional algebras, Theorem 1 eliminates all except three of the indecomposable algebras $A_{5,36}$, $A_{5,37}$, and $A_{5,40}$ (see Table III) and two of the decomposable algebras $A_{3,8} \oplus A_2$ and $A_{3,9} \oplus A_2$. We list the four-dimensional algebras (in convenient basis $\{G_i : i = 1,4\}$) of relevance together with their three-dimensional subalgebras (Patera and Winternitz[5]) in Table II. The subalgebras listed in Table II are enclosed in parentheses (rather than in braces), e.g., $(G_1,G_3;G_2)$. This is to indicate that they are maximal subalgebras of the Lie algebras considered. The generators of the derived algebra are written to the right of the semicolon.

Concerning the five-dimensional real Lie algebras (indecomposable), all but three of the algebras do not contain

TABLE II. Relevant algebras of dimension 4 and their three-dimensional subalgebras.

Algebra	Nonzero commutators	Dimension 3
$2A_2$	$[G_1,G_2]=G_2,$	$A_1\oplus A_2$: $(G_1,G_3;G_2)(G_1,G_4;G_2)$
	$[G_3,G_4]=G_4$	$(G_1,G_3;G_4)(G_2,G_3;G_4)$
		$A_{3,3}$: $(G_1+G_3;G_2,G_4)$
		$A_{3,4}$: $(G_1-G_3;G_2,G_4)$
		$A_{3,5}^a$: $(G_1+xG_3;G_2,G_4)$, $a=\begin{cases}x, & 0<\lvert x\rvert<1,\\ 1/x, & 1<\lvert x\rvert<\infty,\end{cases}$
$A_{3,8}\oplus A_1$	$[G_1,G_3]=2G_2,$	$A_2\oplus A_1$: $(G_2,G_4;G_1)$
	$[G_1,G_2]=1,$	
	$[G_2,G_3]=G_3$	$A_{3,8}$: $(;G_1,G_2,G_3)$
$A_{3,9}\oplus A_1$	$[G_1,G_2]=G_3,$	$A_{3,9}$: $(;G_1,G_2,G_3)$
	$[G_2,G_3]=G_1,$	
	$[G_3,G_1]=G_2$	
$A_{4,7}$	$[G_1,G_4]=2G_1,$	$A_{3,1}$: $(G_2,G_3;G_1)$
	$[G_2,G_4]=G_2,$	
	$[G_3,G_4]=G_2+G_3,$	$A_{3,5}^{1/2}$: $(G_4;G_1,G_2)$
	$[G_2,G_3]=G_1$	
$A_{4,8}$	$[G_2,G_3]=G_1,$	$A_{3,1}$: $(G_2,G_3;G_1)$
	$[G_2,G_4]=G_2,$	
	$[G_3,G_4]=-G_3$	$A_2\oplus A_1$: $(G_4,G_1;G_2)$, $(G_4,G_1;G_3)$
$A_{4,9}^b$	$(G_2,G_3]=G_1,$	$A_{3,1}$: $(G_2,G_3;G_1)$
$(0<\lvert b\rvert<1)$	$[G_1,G_4]=(1+b)G_1,$	$A_{3,5}^v$: $(G_4;G_1,G_2)$,
	$[G_2,G_4]=G_2,$	$v=\begin{cases}1+b, & \lvert 1+b\rvert<1,\\ 1/(1+b), & \lvert 1+b\rvert>1,\end{cases}$
	$[G_3,G_4]=bG_3$	$A_{3,4}$: $(G_4;G_1,G_3)$, $b=-\frac{1}{2}$
		$A_{3,5}^w$: $(G_4;G_1,G_3)$, $w=\begin{cases}b/(1+b), & \lvert b/(1+b)\rvert<1,\\ (1+b)/b, & \lvert b/(1+b)\rvert>1,\end{cases}$
$A_{4,9}^1$	$[G_2,G_3]=G_1,$	$A_{3,1}$: $(G_2,G_3;G_1)$
	$[G_1,G_4]=2G_1,$	
	$[G_2,G_4]=G_2,$	$A_{3,5}^{1/2}$: $(G_4;G_1,\cos\phi G_2+\sin\phi G_3)$
	$[G_3,G_4]=G_3$	
		$0\leqslant\phi<\pi$
$A_{4,9}^0$	$[G_2,G_3]=G_1,$	$A_{3,1}$: $(G_2,G_3;G_1)$
	$[G_1,G_4]=G_1,$	
	$[G_2,G_4]=G_2$	$A_2\oplus A_1$: $(G_3,G_4;G_1)$
		$A_{3,3}$: $(G_4;G_1,G_2)$
		$A_{3,2}$: $(G_4+xG_3;G_1,G_2)$ $(x\neq 0)$
$A_{4,10}$	$[G_2,G_3]=G_1,$	$A_{3,1}$: $(G_2,G_3;G_1)$
	$[G_2,G_4]=-G_3,$	
	$[G_3,G_4]=G_2$	
$A_{4,11}^a$	$[G_2,G_3]=G_1,$	$A_{3,1}$: $(G_2,G_3;G_1)$
	$[G_1,G_4]=2aG_1,$	
$(a>0)$	$[G_2,G_4]=aG_2-G_3,$	
	$[G_3,G_4]=G_2+aG_3$	
$A_{4,12}$	$[G_1,G_3]=G_1,$	$A_{3,3}$: $(G_3;G_1,G_2)$
	$[G_2,G_3]=G_2,$	
	$[G_1,G_4]=-G_2,$	$A_{3,6}$: $(G_4;G_1,G_2)$
	$[G_2,G_4]=G_1$	$A_{3,7}^{\lvert x\rvert}$: $(G_4+xG_3;G_1,G_2)$ $(x\neq 0)$

TABLE III. Relevant algebras of dimension 5.

Algebra	Nonzero commutation relations
$A_{5,36}$	$[G_2,G_3]=G_1,\quad [G_1,G_4]=G_1,\quad [G_2,G_4]=G_2$
	$[G_2,G_5]=-G_2,\quad [G_3,G_5]=G_3$
$A_{5,37}$	$[G_2,G_3]=G_1,\quad [G_1,G_4]=2G_1,\quad [G_2,G_4]=G_2$
	$[G_3,G_4]=G_3,\quad [G_2,G_5]=-G_3,\quad [G_3,G_5]=G_2$
$A_{5,40}$	$[G_1,G_2]=2G_1,\quad [G_1,G_3]=-G_2,\quad [G_2,G_3]=2G_3$
	$[G_1,G_4]=G_5,\quad [G_2,G_4]=G_4,\quad [G_2,G_5]=-G_5,$
	$[G_3,G_5]=G_4$

$3A_1$ as a subalgebra. Their commutation relations in a convenient bases $\{G_i: i=1,5\}$ are given in Table III.

We shall discuss the Lie algebras listed in Tables II and III after investigating the three-dimensional Lie algebra realizations of Table I.

Proposition 2: If a second-order equation admits the Lie algebra $\mathrm{sl}(2,\mathbb{R})$ ($A_{3,8}$), then it has either three or eight generators of symmetry.

Proof: Suppose an equation admits the Lie algebra $\mathrm{sl}(2,\mathbb{R})$, i.e., the generators of symmetry satisfy the $\mathrm{sl}(2,\mathbb{R})$ commutation relations. Proceeding as in Theorem 1 we cannot have $G_1=\rho(t,q)G_2$ or $G_3=\psi(t,q)G_2$ (for any nonconstant functions ρ and ψ) as, in the case $G_1=\rho G_2$, the first and last commutators yield

$$G_2\rho+\rho=0,\quad G_3=(1/\rho)G_2, \tag{2}$$

with the remaining one giving rise to

$$G_2(1/\rho)=1/\rho, \tag{3}$$

these in turn leading to the generators [on transforming G_2 to $\bar{G}_2=Q(\partial/\partial Q)$ and choosing the simplest solution of (2a)]

$$\bar{G}_1=\frac{\partial}{\partial Q},\quad \bar{G}_2=Q\frac{\partial}{\partial Q},\quad \bar{G}_3=Q^2\frac{\partial}{\partial Q}, \tag{4}$$

which as is seen by elementary calculation cannot be symmetries of a second-order equation! It follows therefore that $G_1\neq\rho G_2$ for any function ρ. Similarly it can be shown that no function ψ exists such that $G_3=\psi G_2$. Hence $G_1\neq\rho G_2$ and $G_3\neq\psi G_2$ for any functions ρ and ψ. Consequently, there exists a regular point transformation $Q=Q(t,q)$, $T=T(t,q)$, transforming G_1, G_2 to the canonical form

$$\bar{G}_1=\frac{\partial}{\partial Q},\quad \bar{G}_2=T\frac{\partial}{\partial T}+Q\frac{\partial}{\partial Q}. \tag{5a}$$

The commutators involving $\bar{G}_3$ then give us

$$\bar{G}_3=(2TQ+\alpha T^2)\frac{\partial}{\partial T}+(Q^2+\beta T^2)\frac{\partial}{\partial Q}, \tag{5b}$$

where α,β are constants.

It is now a simple matter to obtain the differential equations for Q having at least the three symmetries (5). They are

$$TQ''=\frac{-1}{(\alpha^2+4\beta)}\{(2Q'+\alpha)(2Q'^2+2\alpha Q'-2\beta)+A(2Q'^2+2\alpha Q'-2\beta)^{3/2}\}, \tag{6}$$

with $\alpha^2+4\beta\neq 0$, A constant and

$$TQ''=-\frac{1}{2}(Q'+\alpha/2)+B(Q'+\alpha/2)^3, \tag{7}$$

with $\alpha^2+4\beta=0$, B constant. The emergence of two differential equations (6) and (7) should imply the existence of two canonical forms for the generators. This in fact is the case. We find that the linear transformations

$$\tilde{T}=(i/2)(\alpha^2+4\beta)^{1/2}T,\quad i=\sqrt{-1},$$
$$\tilde{Q}=Q+(\alpha/2)T,\quad \alpha^2+4\beta\neq 0,$$

and

$$\tilde{T}=B^{-1/2}T,\quad B\neq 0,$$
$$\tilde{Q}=Q+(\alpha/2)T,\quad \alpha^2+4\beta=0,$$

do provide reductions to the two canonical forms

$$\tilde{G}_1=\frac{\partial}{\partial\tilde{Q}},\quad \tilde{G}_2=\tilde{T}\frac{\partial}{\partial\tilde{T}}+\tilde{Q}\frac{\partial}{\partial\tilde{Q}},$$
$$\tilde{G}_3=2\tilde{T}\tilde{Q}\frac{\partial}{\partial\tilde{T}}+(\tilde{Q}^2-\tilde{T}^2)\frac{\partial}{\partial\tilde{Q}} \tag{8}$$

and

$$\tilde{G}_1=\frac{\partial}{\partial\tilde{Q}},\quad \tilde{G}_2=\tilde{T}\frac{\partial}{\partial\tilde{T}}+\tilde{Q}\frac{\partial}{\partial\tilde{Q}},$$
$$\tilde{G}_3=2\tilde{T}\tilde{Q}\frac{\partial}{\partial\tilde{T}}+\tilde{Q}^2\frac{\partial}{\partial\tilde{Q}}. \tag{9}$$

The associated differential equations are

$$\tilde{T}\tilde{Q}''=\tilde{Q}'^3+\tilde{Q}'+\mathscr{A}(1+\tilde{Q}'^2)^{3/2}, \tag{10}$$
$$\tilde{T}\tilde{Q}''=\tilde{Q}'^3-\frac{1}{2}\tilde{Q}', \tag{11}$$

where $\mathscr{A}$ is a constant. It now becomes clear that Eq. (10) has more than the three given symmetries (8) whenever $\mathscr{A}=0$. This follows straightforwardly from the results of Ref. 3 [see Sec. I, (i)]. It is also evident that Eq. (7) is linear when $B=0$. For $B\neq 0$ we have the form (11). Equation (10) has exactly the three symmetries (8) whenever $\mathscr{A}$ is nonzero.

It should also be mentioned that the transformations

$$\bar{T}=Q+aT,\quad \bar{Q}=Q+bT,\quad a,(\neq b)\text{ constants},$$

with $\alpha=a+b, \beta=-ab$ and

$$\bar{T}=Q+aT,\quad \bar{Q}=T^{1/2},$$

with $\alpha=2a, \beta=-a^2$ do provide reductions to the two Lie canonical forms

$$\bar{G}'_1=\frac{\partial}{\partial\bar{T}}+\frac{\partial}{\partial\bar{Q}},\quad G'_2=\bar{T}\frac{\partial}{\partial\bar{T}}+\bar{Q}\frac{\partial}{\partial\bar{Q}},$$
$$\bar{G}'_3=\bar{T}^2\frac{\partial}{\partial\bar{T}}+\bar{Q}^2\frac{\partial}{\partial\bar{Q}}, \tag{12a}$$

and

$$\bar{G}'_1=\frac{\partial}{\partial\bar{T}},\quad \bar{G}'_2=\bar{T}\frac{\partial}{\partial\bar{T}}+\frac{1}{2}\bar{Q}\frac{\partial}{\partial\bar{Q}},$$
$$\bar{G}'_3=\bar{T}^2\frac{\partial}{\partial\bar{T}}+\bar{T}\bar{Q}\frac{\partial}{\partial\bar{Q}}. \tag{12b}$$

The corresponding differential equations are

$$(\bar{T}-\bar{Q})\bar{Q}''+2(\bar{Q}'+\bar{A}\bar{Q}'^{3/2}+\bar{Q}'^2)=0, \tag{13a}$$
$$\bar{Q}^3\bar{Q}''+\bar{B}=0. \tag{13b}$$

where $\bar{A}$ and $\bar{B}$ are constants.

Proposition 3: If a second-order equation admits the Lie algebra $A^b_{3,7}$ $(b>0)$ $A_{3,6}$, then it has either three or eight generators of symmetry.

Proof: Suppose an equation admits the algebra $A^b_{3,7}$ $(b>0)$, $A_{3,6}$. Then we cannot have $G_2=\rho(t,q)G_1$ and $G_3=\psi(t,q)G_1$ for any nonconstant functions ρ and ψ as this would lead to $\rho^2+1=0$. We thus assume $G_2\neq\rho G_1$ for any ρ. This produces the following generators:

$$\bar{G}_1=\frac{\partial}{\partial T},\quad \bar{G}_2=\frac{\partial}{\partial Q},$$
$$\bar{G}_3=(bT+Q)\frac{\partial}{\partial T}+(bQ-T)\frac{\partial}{\partial Q}. \tag{14}$$

The differential equation for Q is

$$Q''=A(1+Q'^2)^{3/2}\exp(b\arctan Q'), \tag{15}$$

where A is a constant. It is fairly simple to observe that the rhs of Eq. (15) cannot be a polynomial which is at most cubic in Q' unless $A=0$. Hence it follows from the conclusion of Ref. 2 (see Sec. I) that (15) is not linearizable via a point transformation if $A\neq 0$. It has three symmetries. For $A=0$ Eq. (15) clearly has five more symmetries.

We next assume that $G_3\neq\psi G_1$ for any function ψ. Without loss of generality we may further assume that $G_1=\phi(t,q)G_2$ for some function ϕ. Since we have $[G_1,G_2]=0$, the reduction to canonical form of the generators G_1 and G_2 is given by

$$\bar{G}_1=T\frac{\partial}{\partial Q},\quad \bar{G}_2=\frac{\partial}{\partial Q}, \tag{16a}$$

and

$$\bar{G}_3=(1+T^2)\frac{\partial}{\partial T}+(TQ+bQ)\frac{\partial}{\partial Q}, \tag{16b}$$

following easily from the commutators involving G_3. The associated differential equation is

$$Q''=B(1+T^2)^{-3/2}\exp(b\arctan T), \tag{17}$$

where B is a constant. This equation clearly is linear. ◇

The canonical forms (14) and (16) are presented for the first time. The realizations of the family of Lie algebras $A^b_{3,7}$ $(b>0)$ and $A_{3,6}$ and the corresponding second-order equations were not considered by Lie.

Remark: Suppose we have an equation of the precise form (15) or (17) but with $b<0$. Then we can introduce the basis $\{V_1=\bar{G}_2, V_2=\bar{G}_1, V_3=-\bar{G}_3\}$ so that the resulting Lie algebra is of the desired form $A^{-b}_{3,7}(-b>0)$.

Proposition 4: If an equation admits the Lie algebra $A_{3,2}$, then it has either three or eight generators of symmetry.

Proof: Suppose an equation admits the algebra $A_{3,2}$. We find that the symmetry generators G_i cannot be proportional to each other. Therefore we resort to proof by cases. In the first case we assume that no function ψ exists such that $G_3=\psi(t,q)G_1$. This, proceeding as before, gives rise to the following generators:

$$\bar{G}_1=\frac{\partial}{\partial Q},\quad \bar{G}_2=\beta\frac{\partial}{\partial T}-\ln T\frac{\partial}{\partial Q},\quad \beta\text{ constant},$$
$$\bar{G}_3=T\frac{\partial}{\partial T}+Q\frac{\partial}{\partial Q} \tag{18}$$

(disregarding an additive constant multiple of G_1). Expressing the invariance of a general differential equation for Q with respect to (18) yields

$$TQ''=-1/\beta+A\exp(-\beta Q'),\quad \beta\neq 0 \tag{19}$$

and

$$TQ'' = -Q' + B, \quad \beta = 0, \tag{20}$$

where A and B are constants. Equation (20) is linear. It is immediate from the results of Ref. 3 [see (i)] that (19) is not linearizable (via a point transformation) if $A \neq 0$. In this case (19) has exactly three symmetries.

The transformations

$$\bar{Q} = Q + \frac{T}{\beta}\ln T - \frac{T}{\beta}, \quad \bar{T} = \frac{T}{\beta}, \quad \beta \neq 0,$$

and

$$\bar{Q} = -Q, \quad \bar{T} = -\ln T, \quad \beta = 0,$$

give the Lie canonical forms

$$\bar{G}_1' = \frac{\partial}{\partial \bar{Q}}, \quad \bar{G}_2' = \frac{\partial}{\partial \bar{T}}, \qquad \bar{G}_3' = \bar{T}\frac{\partial}{\partial \bar{T}} + (\bar{T} + \bar{Q})\frac{\partial}{\partial \bar{Q}}, \tag{21}$$

and

$$\bar{G}_1' = -\frac{\partial}{\partial \bar{Q}}, \quad \bar{G}_2' = \bar{T}\frac{\partial}{\partial \bar{Q}}, \qquad \bar{G}_3' = \frac{\partial}{\partial \bar{T}} + \bar{Q}\frac{\partial}{\partial \bar{Q}}. \tag{22}$$

The associated differential equations are

$$\bar{Q}'' = \bar{A}\exp(-\bar{Q}'), \tag{23}$$

$$\bar{Q}'' = \bar{B}\exp(\bar{T}), \tag{24}$$

where $\bar{A}$ and $\bar{B}$ are constants.

The case $G_2 \neq \rho G_1$ for ρ arbitrary gives the generators (21) above. ◊

Proposition 5: If an equation admits the Lie algebra $A_{3,5}^{a}$ $(0 < |a| < 1)$, $A_{3,4}$, $A_1 \oplus A_2$, then it has either three or eight generators of symmetry.

Proof: Suppose an equation admits the algebra $A_{3,5}^{a}$ $(0 < |a| < 1)$, $A_{3,4}$, $A_1 \oplus A_2$.
Then we certainly must have $G_2 \neq \rho(t,q)G_1$ or $G_3 \neq \psi(t,q)G_1$ for any functions ρ and ψ. Otherwise the commutators would imply $G_1\psi = a$ $(\neq 1)$ and $G_1\psi = 1$. Thus we first assume that no function ρ exists such that $G_2 = \rho G_1$. As a result the reduction to canonical form of the generators G_1 and G_2 is immediate and we finally obtain

$$\bar{G}_1 = \frac{\partial}{\partial T}, \quad \bar{G}_2 = \frac{\partial}{\partial Q}, \qquad \bar{G}_3 = T\frac{\partial}{\partial T} + aQ\frac{\partial}{\partial Q}. \tag{25}$$

The differential equation for Q invariant under at least the three symmetry generators (25) is

$$Q'' = AQ'^{(a-2)/(a-1)}, \tag{26}$$

where A is a constant. It follows from Ref. 2 (see Sec. I) that Eq. (26) is not linearizable (via a point transformation) if $A \neq 0$ and $a \neq 0, \frac{1}{2}, 2$. In this case (26) has three symmetries. Note that for $|a| > 1$ we can introduce the basis $\{V_1 = \bar{G}_2, V_2 = \bar{G}_1, V_3 = (1/a)\bar{G}_3\}$ so that the resulting Lie algebra is $A_{3,5}^{1/a}(0 < 1/|a| < 1)$.

For the second part of the proof we assume $G_3 \neq \psi G_1$ for any function ψ. This gives rise to the symmetry generators

$$\bar{G}_1 = \frac{\partial}{\partial Q}, \quad \bar{G}_2 = \alpha T^{1-a}\frac{\partial}{\partial T} + \beta T^{1-a}\frac{\partial}{\partial Q}, \qquad \bar{G}_3 = T\frac{\partial}{\partial T} + Q\frac{\partial}{\partial Q}, \tag{27}$$

where α and β are constants. The associated differential equations having at least these three symmetries are then given by

$$TQ'' = (1/\alpha)(a-1)(\alpha Q' - \beta) + B(\alpha Q' - \beta)^{(2a-1)/(a-1)}, \quad \alpha \neq 0, \tag{28}$$

$$TQ'' = -aQ' + C, \quad \alpha = 0, \tag{29}$$

where B and C are constants. Equation (29) being linear has eight symmetries, three of them given by (27) (with $\alpha = 0$). By means of a linear transformation we can reduce (28) to the form [may as well set $\alpha = 1, \beta = 0$ in (28)]

$$TQ'' = (a-1)Q' + \mathscr{B}Q'^{(2a-1)/(a-1)}, \tag{30}$$

where $\mathscr{B}$ is a constant. Evidently this equation is not linearizable if $\mathscr{B} \neq 0$ and $a \neq 0, \frac{1}{2}$. For $a = 2$ and $\mathscr{B} \neq 0$, (30) takes the form

$$TQ'' = \mathscr{B}Q'^3 + Q'. \tag{31}$$

By a rescaling of time we can make the constant $\mathscr{B}$ unity, whence we have the familiar equation of Ref. 3 [see (i)] for which we obtained eight-point symmetries. ◊

Under the transformation

$$Q = t, \quad T = a^{1/a}q^{1/a}, \quad a \neq 0, \tag{32}$$

(25) (in lower case variables and with $a \neq 0$) is equivalent to (27) $(\alpha = 1, \beta = 0)$. When $a = 0$ $(\alpha = 1, \beta = 0)$ the generators (27) are reducible to the simple form

$$\bar{G}_1 = \frac{\partial}{\partial Q}, \quad \bar{G}_2 = T\frac{\partial}{\partial T}, \quad G_3' = Q\frac{\partial}{\partial Q}, \tag{33}$$

where $G_3' = \bar{G}_3 - \bar{G}_2$. The generators (25) (in lower case variables and with $a = 0$) are now equivalent to (33) by means of the transformation

$$T = \exp q, \quad Q = t.$$

We also point out that the transformation

$$Q = q, \quad T = \beta t^{1-a}, \quad \beta \neq 0,$$

maps (27) (in lower case variables with $\alpha = 0$) to

$$\bar{G}_1 = \frac{\partial}{\partial Q}, \quad \bar{G}_2 = T\frac{\partial}{\partial Q}, \qquad \bar{G}_3 = (1-a)T\frac{\partial}{\partial T} + Q\frac{\partial}{\partial Q}. \tag{34}$$

The related equation is

$$Q'' = DT^{(1-2a)/(a-1)}, \tag{35}$$

where D is a constant.

In view of the above discussion we have two canonical forms for the generators. They are either the Lie canonical forms given by (25) and (34) or those given by (27) upon setting $\alpha = 0, \beta = 1$ and $\alpha = 1, \beta = 0$, respectively.

We now prove theorems relating to the linearizability (via a point transformation) of a second-order equation.

Theorem 6: In order that a differential equation $\ddot{q} = N(\dot{q},q,t)$ possesses sl(3,$\mathbb{R}$) algebra it is necessary and sufficient that it has the algebra (A) $A_1 \oplus A_2$, $A_{3,5}^{1/2}$, (B) $A_{3,3}$, (C) $A_{3,1}$, (D) $A_{3,9}$ [so(3)].

Proof (A): To prove sufficiency we need only refer to Proposition 5 and the discussion following it. We restrict our attention to (27) with $\alpha = 0$, $\beta = 1$ and $\alpha = 1$, $\beta = 0$, respectively, bearing in mind that a can take on the values 0, $\frac{1}{2}$, or 2. When $a = 2$ we always can introduce a basis so that the Lie algebra is $A_{3,5}^{1/2}$. This much is apparent from the previous theorem. The relevant equations (29) and (30) each have eight generators of symmetry whenever $a = 0$, $\frac{1}{2}$, 2. Thus they admit the sl(3,$\mathbb{R}$) algebra. The original equation $\ddot{q} = N(\dot{q},q,t)$ accordingly has the same algebra. Proof of the necessity is trivial because $A_1 \oplus A_2$ and $A_{3,5}^{1/2}$ are subalgebras of sl (3,$\mathbb{R}$). ◊

We proved part (B) of the above theorem in Ref. 6 which deals with the reverse procedure to find the point symmetry algebra from three given first integrals of a second-order equation that possesses sl(3,$\mathbb{R}$) algebra. The algebra $A_{3,3}$ arises in connection with each of the three triplets of symmetry generators of the second-order equation.

The proof of part (C) is given in Ref. 2. The algebra $A_{3,1}$ plays an important role in the determination of linearizing transformations for equations of Type I considered in Ref. 2. In Ref. 2 the Lie algebra $A_{3,1}$ was referred to as the $\aleph$ algebra. Let us also mention that some of the structure of the proofs in this paper are closely patterned along the lines of the proofs of the propositions of Refs. 2 and 6. We now prove part (D).

Proof (*D*): Suppose an equation has the Lie algebra so(3) ($A_{3,9}$). Then we cannot have $G_1 = \rho(t,q)G_2$ or $G_3 = \psi(t,q)G_2$ (for any nonconstant functions ρ and ψ) as, in the case $G_1 = \rho G_2$, we deduce

$$(G_2\rho)^2 + \rho^2 = -1, \quad G_3 = -(G_{2\rho})G_2, \tag{36}$$

which in turn lead to the generators [on transforming G_2 to $\bar{G}_2 = \partial/\partial Q$ and choosing the simplest solution of (36a)]

$$\bar{G}_1 = i \sin Q \frac{\partial}{\partial Q}, \quad \bar{G}_2 = \frac{\partial}{\partial Q},$$
$$\bar{G}_3 = i \cos Q \frac{\partial}{\partial Q}, \quad i = \sqrt{-1}, \tag{37}$$

which are not symmetries of a second-order equation. Therefore $G_1 \neq \rho G_2$ for any function ρ. Likewise we can show that no function ψ exists such that $G_3 = \psi G_2$. Hence $G_1 \neq \rho G_2$ and $G_3 \neq \psi G_2$. This being the case we always can choose coordinates in which one of the generators appears as a generator of time translation. Thus the only admissible transformation from now on will be of the form

$$Q = \alpha(q), \quad T = t + \beta(q), \tag{38}$$

where α and β are as yet arbitrary. We write the three generators as

$$G_1 = \frac{\partial}{\partial t}, \quad G_2 = \xi_2 \frac{\partial}{\partial t} + \eta_2 \frac{\partial}{\partial q},$$
$$G_3 = \xi_3 \frac{\partial}{\partial t} + \eta_3 \frac{\partial}{\partial q}, \tag{39}$$

where the ξ's and η's are functions of t and q. Invoking the commutators $[G_1,G_2] = G_3$ and $[G_1,G_3] = -G_2$ of the Lie algebra so(3), we obtain

$$\frac{\partial \xi_2}{\partial t} = \xi_3, \quad \frac{\partial \xi_3}{\partial t} = -\xi_2,$$
$$\frac{\partial \eta_2}{\partial t} = \eta_3, \quad \frac{\partial \eta_3}{\partial t} = -\eta_2, \tag{40}$$

which in turn imply

$$\frac{\partial^2 \xi_2}{\partial t^2} + \xi_2 = 0, \quad \frac{\partial^2 \eta_2}{\partial t^2} + \eta_2 = 0.$$

Consequently ξ_2 and η_2 are

$$\xi_2 = a \cos t + b \sin t, \quad \eta_2 = c \cos t + d \sin t,$$

where a,b,c, and d are functions of q. The coordinate functions ξ_3 and η_3 are immediately given by

$$\xi_3 = -a \sin t + b \cos t, \quad \eta_3 = -c \sin t + d \cos t.$$

The remaining commutator $[G_2,G_3] = G_1$ then yields the conditions

$$cb' - da' = 1 + a^2 + b^2, \quad cd' - dc' = bd + ac. \tag{41}$$

Since the generators G_1, G_2, and G_3 are unconnected we must have $c \neq 0$ or $d \neq 0$. We firstly assume that $c \neq 0$. Without loss of generality we may further take $d = 0$. The reason for this is straightforward. For under the transformation (38) G_2 transforms to

$$\bar{G}_2 = \bar{\xi}_2 \frac{\partial}{\partial T} + \bar{\eta}_2 \frac{\partial}{\partial Q},$$

where

$$\bar{\xi}_2 = \bar{a} \cos T + \bar{b} \sin T, \quad \bar{\eta}_2 = \bar{c} \cos T + \bar{d} \sin T,$$

with

$$\bar{d} = \alpha'(c \sin \beta + d \cos \beta).$$

Similar expressions hold for $\bar{a}$, $\bar{b}$, and $\bar{c}$. Hence, if $d \neq 0$ at the outset, we can make $\bar{d} = 0$ by the requirement that β be determined from the relation $\cot \beta = -c/d$. This relation is not violated when G_3 transforms to $\bar{G}_3$.

Henceforth we work in coordinates in which $d = 0$. In order to preserve this property we further restrict the transformations (38) to be of the form

$$Q = \alpha(q), \quad T = t. \tag{42}$$

The conditions (41) with $d = 0$ imply

$$a = 0, \quad cb' = 1 + b^2,$$

making the generators appear in simplified form. By means of the transformation [set $\alpha = b$ in (42)]

$$Q = b(q), \quad T = t,$$

the generators G_i acquire the form

$$\bar{G}_1=\frac{\partial}{\partial T},\quad \bar{G}_2=Q\sin T\frac{\partial}{\partial T}+(1+Q^2)\cos T\frac{\partial}{\partial Q},$$
$$\bar{G}_3=Q\cos T\frac{\partial}{\partial T}-(1+Q^2)\sin T\frac{\partial}{\partial Q}. \tag{43}$$

The generators (43) are reduced even further via the transformation

$$\tilde{T}=\tan T,\quad \tilde{Q}=Q/\cos T.$$

They transform to

$$\tilde{G}_1=(1+\tilde{T}^2)\frac{\partial}{\partial\tilde{T}}+\tilde{T}\tilde{Q}\frac{\partial}{\partial\tilde{Q}},$$
$$\tilde{G}_2=\tilde{T}\tilde{Q}\frac{\partial}{\partial\tilde{T}}+(1+\tilde{Q}^2)\frac{\partial}{\partial\tilde{Q}}, \tag{44}$$
$$\tilde{G}_3=\tilde{Q}\frac{\partial}{\partial\tilde{T}}-\tilde{T}\frac{\partial}{\partial\tilde{Q}}.$$

The differential equation for $\tilde{Q}$ is the free particle equation. A similar argument holds for the case $d\neq 0$, once again giving rise to the above generators (43). It therefore follows that the equation for q has the sl(3,$\mathbb{R}$) algebra, concluding sufficiency. The necessity is trivial since so(3) is a subalgebra of sl(3,$\mathbb{R}$). ◊

We point out that the generators (43) were previously obtained by Wulfman and Wybourne[7] in their treatment of the simple harmonic oscillator.

The so(3) algebra when realized in two coordinates has a unique canonical form, viz., (44) which are generators of symmetry of a second-order equation. The so(3) realizations derived above were omitted by Lie.

We now concentrate on four- and higher-dimensional Lie algebras. The four-dimensional algebras of interest are listed in Table II. Each one of them contains a three-dimensional subalgebra which implies linearization by Theorem 6. Hence, if a second-order equation admits a four-dimensional algebra, then it is linearizable. However, there are four-dimensional algebras which are not admitted by any second-order equation. In fact one can prove the following.

Proposition 7: A second-order equation does not admit the Lie algebra (A) $A_{3,9}\oplus A_1$, (B) $A_{4,7}$, (C) $A_{4,10}$, and (D) $A^a_{4,11}$ $(a>0)$, respectively.

Remark: We only prove the result for (A). The proofs for (B), (C), and (D) are similar and are therefore omitted.

Proof: Suppose a second-order equation admits the algebra $A_{3,9}\oplus A_1$. Then we can use the $A_{3,9}$ realization given in (43) (now using lower case) since (43) is the only canonical realization of $A_{3,9}$ that generates symmetries of a second-order equation.

Writing G_4 in the general form $G_4=\xi(\partial/\partial t)+\eta(\partial/\partial q)$ and applying $[G_1,G_4]=0$, we find that ξ and η are independent of t. The remaining commutators then require that both ξ and η be zero, contradicting the assumption. ◊

A second-order equation does not admit the five-dimensional algebra $A_{3,9}\oplus A_2$. This is a consequence of Proposition 7 by noting that $A_{3,9}\oplus A_1$ is a subalgebra of $A_{3,9}\oplus A_2$. Moreover, it is not difficult to verify $A_{3,8}\oplus A_2$ is not admitted by a second-order equation. Each of the Lie algebras $A_{5,36}$, $A_{5,37}$, and $A_{5,40}$ (see Table III) has a four-dimensional subalgebra and hence imply linearization for a second-order equation that admits it.

We now investigate the case of a second-order equation which admits a six- or seven- dimensional real Lie algebra. The only simple algebra of dimension less than eight (greater than five) is so(3,1) (see, e.g., Ref. 8). If a second-order equation admits so(3,1) then it is linearizable since so(3,1) contains so(3) as a subalgebra which implies linearization by Theorem 6. Every six- or seven-dimensional nonsemisimple real Lie algebra is Levi decomposable or can be written as a direct sum of lower dimensional Lie algebras. A six- or seven-dimensional nonsemisimple real Lie algebra has a four-dimensional subalgebra (see Ref. 8). This in turn implies linearization for a second-order equation that admits a six- or seven-dimensional nonsemisimple Lie algebra.

An immediate consequence of the preceding propositions, theorems, and related discussions is the following interesting result.

Theorem 8: A second-order equation does not admit exactly an $r\in\{4,5,6,7\}$ dimensional point symmetry algebra.

III. EQUIVALENCE CLASSES OF EQUATIONS

A second-order ordinary differential equation has either 0,1,2,3, or 8 point symmetries. We exclude the case of equations possessing no point symmetry as we cannot in general write representatives of equivalence classes for such equations. If an equation has one-point symmetry, it can be reduced to an autonomous form by means of a point transformation which brings the symmetry to a generator of time translation. Thus an equation with a single point symmetry belongs to the equivalence class of

$$\ddot{q}=f(q,\dot{q}),$$

where $\partial/\partial t$ (realizes A_1) is the standard form for the symmetry and f is a definite function of q and $\dot{q}$. We treat the case of equations possessing two-point symmetries after investigating the three-point symmetry case. For the equivalence classes of equations having three-point symmetries we need only recall the results of the previous section. It follows that there are five representatives of equivalence classes (in each case $A\neq 0$ and $a\in\mathbb{R}$). They are

$$t\ddot{q}=\dot{q}^3+\dot{q}+A(1+\dot{q}^2)^{3/2}, \tag{45a}$$
$$t\ddot{q}=A\dot{q}^3-\tfrac{1}{2}\dot{q}, \tag{45b}$$
$$t\ddot{q}=(a-1)\dot{q}+A\dot{q}^{(2a-1)/(a-1)}\ \text{or}\ \ddot{q}=A\dot{q}^{(a-2)/(a-1)},\quad a\neq 0,\tfrac{1}{2},1,2, \tag{45c}$$
$$t\ddot{q}=-1+A\exp(-\dot{q})\ \text{or}\ \ddot{q}=A\exp(-\dot{q}), \tag{45d}$$
$$\ddot{q}=A(1+\dot{q}^2)^{3/2}\exp(a\arctan\dot{q}). \tag{45e}$$

It is now simple to deduce that equations possessing two-point symmetries belong to either of the equivalence classes

$$\ddot{q}=f(\dot{q}),\quad t\ddot{q}=g(\dot{q}), \tag{46}$$

where f is not a polynomial which is at most cubic in $\dot{q}$ and it is not of the form given in (45) and g is not linear in $\dot{q}$ and neither is g of the form given in (i) nor of the form given in (45).

Equations admitting the richest number of point symmetries belong to the equivalence class of the free particle equation, $\ddot{q}=0$.

IV. CONCLUSION

In this work we have shown how second-order ordinary differential equations can be classified by investigating the realizations of real low-dimensional Lie algebras in terms of vector fields in two coordinates. In this way we were able to associate differential equations to those realizations that are generators of symmetry of a second-order equation. Our results show that an equation admitting exactly an r-dimensional ($r \in \{1,2,3,8\}$) Lie algebra of point symmetry generator(s) possess(es) a canonical representative for the corresponding differential equation.

For the case $r=1$ the differential equation can be transformed to an autonomous form. The case $r=2$ resulted in two classes of representative equations. For equations having three-point symmetries ($r=3$ case), we obtained five representatives of equivalence classes. The subject of linearization ($r=8$ case) was addressed and a number of linearizability results were proved.

We have in fact obtained a general (local) structure theory for second-order equations which admit point symmetry algebras.

Finally, let us note that many further questions remain open. We, however, content ourselves with just a few of them. It would certainly be of great interest to compare the Lie classification of equations given here with the Painlevé classification which produces 50 equations (see Refs. 9–11 and references therein). The transformation up to which the Painlevé classification was done is given by

$$T=\alpha(t), \quad Q=\frac{\beta(t)q+\gamma(t)}{\delta(t)q+\tau(t)}.$$

The question which arises is this. Is there an overlap between the Lie classification and the Painlevé classification? The first step in answering this question would be to determine the symmetries of the Painlevé equations. Alternatively, one may attempt to reduce some of the Lie equations to the Painlevé ones. This certainly is not a trivial task. For example, how would one go about reducing Eq. (45e) to a Painlevé equation since (45e) contains a transcendental function. If one cannot perform such a reduction, then this would mean that the Painlevé classification is inexhaustive and needs supplementation. Moreover, the Painlevé classification was achieved under the restrictive point transformation given above so one would expect incompleteness. However, the Lie classification also requires supplementation in the case of equations possessing no symmetry as we cannot write representatives of such equations. Even in the case of equations having one symmetry the Lie classification is found to be too general. To remedy this requires further investigation. Nevertheless, it should be pointed out that a preliminary investigation shows that the Painlevé classification does provide representatives for equations having zero or one symmetry.

Recent investigations (Kamran *et al.*[12], Kamran and Shadwick[13]) using Cartan's equivalence method have studied the equivalence of differential equations of the form $\ddot{q}=F(\dot{q},q,t)$ under the restricted point transformation

$$T=\phi(t), \quad Q=\psi(t,q).$$

This was motivated by the Painlevé classification and as such should be viewed against the Painlevé background discussed above.

ACKNOWLEDGMENTS

We thank Professor W. Sarlet and Dr. F. Cantrijn for fruitful discussions from which we benefitted. F. M. also wishes to thank Professor P. Winternitz, Professor R. T. Sharp and Dr. J. van der Jeugt for illuminating discussions on Lie algebras.

F. M. thanks the CSIR and the University of the Witwatersrand for financial support. We thank the referee for constructive suggestions which undoubtedly have enhanced the presentation of the paper.

[1] S. Lie, *Differentialgleichungen* (Chelsea, New York, 1967).
[2] W. Sarlet, F. M. Mahomed, and P. G. L. Leach, J. Phys. A **20**, 277 (1987).
[3] F. M. Mahomed and P. G. L. Leach, Quaest. Math. **12**, 121 (1989).
[4] J. Patera, R. T. Sharp, P. Winternitz, and H. Zassenhaus, J. Math. Phys. **17**, 986 (1976).
[5] J. Patera and P. Winternitz, J. Math. Phys. **18**, 1449 (1977).
[6] P. G. L. Leach and F. M. Mahomed, J. Math. Phys. **29**, 1807 (1988).
[7] C. E. Wulfman and B. G. Wybourne, J. Phys. A **9**, 507 (1976).
[8] P. Turkowski, J. Math. Phys. **29**, 2139 (1988).
[9] E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956).
[10] R. Graham, D. Roekaerts, and T. Tél, Phys. Rev. A **31**, 3364 (1985).
[11] W.-H. Steeb and N. Euler, *Nonlinear Evolution Equations and Painlevé Test* (World Scientific, Singapore, 1988).
[12] N. Kamran, K. G. Lamb, and W. F. Shadwick, J. Diff. Geom. **22**, 139 (1985).
[13] N. Kamran and W. F. Shadwick, *Quantum Gravity, Field Theory and Strings, Springer Lecture Notes in Physics*, Vol. 246 (Springer, Berlin, 1986).

A geometric approach to quantum vortices

Vittorio Penna
Dipartimento di Fisica Teorica, Università di Torino, Via P. Giuria 1, 10100 Torino, Italy
Mauro Spera
Dipartimento di Matematica, Il Università di Roma, Via Orazio Raimondo, 00173 Roma, Italy and Dipartimento di Metodi e Modelli Matematici, Università di Padova, Via Belzoni 7, 35131 Padova, Italy

(Received 15 November 1988; accepted for publication 28 June 1989)

In this paper a geometrical description is given of the theory of quantum vortices first developed by Rasetti and Regge [Physica A **80**, 217 (1975)] relying on the symplectic techniques of Marsden and Weinstein [J. Phys. D **7**, 305 (1983)], and Kirillov–Kostant–Souriau geometric quantization. The RR-current algebra is interpreted as the natural Hamiltonian algebra associated to a certain coadjoint orbit of the group $G = \mathrm{SDiff}(R^3)$, the KKS prequantization condition of which is related to the Feynman–Onsager relation. This orbit is also shown to possess a G-invariant Kaehler structure, whence, in principle, it is possible to quantize it in a natural way.

I. INTRODUCTION

In this paper we investigate the possibility of carrying out the Kirillov–Kostant–Souriau (KKS) geometric quantization program for a rotational perfect fluid in R^3 and in particular for a perfect fluid with vortices.

Three-dimensional vortices arise in nature as topological excitations in a superfluid medium (e.g., 4-He) and furnish the only known example of an extended quantum object. The phase space of this system is T^*G, with $G = \mathrm{SDiff}(R^3)$, the "Lie group" (in the sense of Ref. 1) of measure preserving diffeomorphisms of R^3, which rapidly approach the identity at infinity, the "Lie algebra" of which, $\mathscr{G}$, consists of smooth divergence-free vector fields on R^3 rapidly vanishing at infinity. This allows us to avoid unphysical divergences.

The fluid motion is governed by the right-invariant Hamiltonian

$$H = \frac{1}{2}\int_{R^3} d^3x|\mathbf{v}(\mathbf{x})|^2,$$

with $\mathbf{v}(\mathbf{x}, t)\in\mathscr{G}$ the velocity field obeying Euler's equation which, in terms of the vorticity field $\mathbf{w} = \mathrm{curl}\,\mathbf{v}$, reads

$$-\frac{\partial \mathbf{w}}{\partial t} = [\mathbf{w}, \mathbf{v}]$$

([,] is (minus) the usual Lie brackets of vector fields). We shall be mainly concerned with the (singular) case in which $\mathbf{w}$ is a distribution supported on a smooth closed (possibly knotted) curve Γ on R^3 (or a finite collection thereof)

$$\mathbf{w}(\mathbf{x}) = k\oint_\Gamma \delta(\mathbf{x}-\mathbf{y})d\mathbf{y},$$

$k\in R$ being the strength of $\mathbf{w}$.

The canonical quantization scheme has been developed by Rasetti and Regge[2,3] using quantum field theoretic methods. Their discussion abuts at the consideration of the currents (in the sense of Refs. 4–6)

$$\begin{aligned} J(b) &= \int_{R^3} \mathbf{v}\cdot\mathbf{b}\, d^3x \\ &= \text{(in the singular case)} \\ &= \oint_\Gamma k\,\mathbf{B}\, d\mathbf{y}, \end{aligned}$$
$$\mathbf{b} = \mathrm{curl}\,\mathbf{B}\in\mathscr{G}.$$

These are found to give a Lie algebra representation of $\mathscr{G}$ with respect to the Dirac-bracket operation and provide a complete description of the system.

Thus the problem of constructing a quantum theory of vortices becomes equivalent to that of finding a suitable unitary representation for SDiff (R^3), and this is a difficult problem (also see Refs. 4–6).

Difficulties arise when one looks for representations embodying the topological features of vortices. For an unknotted vortex line a many-particle representation is found that allows us to portray a macroscopic vortex as an assembly of microscopic vortices glued to each other.[3] Otherwise one has to find an explicit description of the Casimir operators, related to the topological invariants of the vortex, considered as a knot.[2,3]

A different way of describing classically the rotational fluid is given by employing the coadjoint orbit picture,[7] which we recall in Sec. II: the Euler equation given above expresses the fact that the fluid motion takes place on a single coadjoint orbit M (labeled by the vorticity field at a fixed time) which, according to KKS theory, is endowed with a natural symplectic structure B.

In this framework the RR current algebra appears as the natural Hamiltonian algebra pertaining to a coadjoint orbit of a Lie group (Theorem 1).

In order to apply the KKS method, the De Rham cohomology class $[B]$ induced by B must be integral, i.e., it should belong to $H^2(M, Z)$. In the case of a vortex line we

 0022-2488/89/122778-07$02.50

find that this fact is connected to the Feynman–Onsager quantization condition:

$$k = Nh/m, \quad N \in Z,$$

where h is Planck's constant and m is the 4-He atom mass. In the general case it is related to the topological charge of Kuznetsov and Mikhailov (see Refs. 8 and 9).

The representation space for G then consists of square integrable (with respect to a suitable metric) sections of a line bundle L on M having first Chern class $C_1(L) = [B]$.

Nevertheless, in order to obtain an irreducible representation, one has to "polarize," i.e., to reduce the representation space; in the case of a vortex line, and this is the main result of this paper, M is Kaehlerian (Theorem 2), whence the line bundle becomes holomorphic and we only have to focus our attention to square integrable holomorphic sections thereof.

The principal ingredient of the construction is the determination of a (local) Kaehler potential. At this point we recall that a preliminary analysis of the G.Q. procedure for quantum vortices has also been differently developed by Goldin, Menikoff, and Sharp.[10] In particular they discuss the problem of finding a real polarization claiming its nonexistence (in particular) in the case of vortex filaments (closed or not): this seems to be consistent with our analysis since we have used a "partial" holomorphic polarization, see Sec. III.

The layout of this paper is the following: after briefly reviewing in Sec. II the KKS theory and quantization of Kaehler manifolds we apply it to rotational fluids in Sec. III following Ref. 7 and we elucidate the geometrical meaning of the RR current algebra; the Kaehler structure is displayed (in the vortex line case) in Sec. IV; finally geometric quantization is outlined in Sec. V.

II. COADJOINT ORBITS OF LIE GROUPS AND GEOMETRIC QUANTIZATION OF KAEHLER MANIFOLDS

In this section we give a condensed survey of the Kirillov–Kostant–Souriau (KKS) theory tailored for our purposes; for a more extensive treatment we refer the reader to Refs. 11–13 and also to Refs. 14–16. A Lie group G acts on its Lie algebra $\mathscr{G}$ through the adjoint representation Ad [if G is a matrix group $\mathrm{Ad}(g)X = gXg^{-1}$, $X \in \mathscr{G}$, $g \in G$], which infinitesimalizes to an action (also called adjoint and denoted by ad) of $\mathscr{G}$ on itself explicity given by

$$\mathrm{ad}(u)v := [u, v], \quad u, v \in \mathscr{G}. \tag{2.1}$$

It induces the so-called coadjoint action ad* of $\mathscr{G}$ on $\mathscr{G}^*$ (the dual of $\mathscr{G}$) via the position:

$$\langle \mathrm{ad}^*(u)(f), v\rangle = -\langle f, [u, v]\rangle, \tag{2.2}$$

$u, v \in \mathscr{G}$, $f \in \mathscr{G}^*$, $\langle\ ,\ \rangle$ the pairing between $\mathscr{G}$ and $\mathscr{G}^*$. At group level we have the coadjoint action Ad* of G on $\mathscr{G}^*$:

$$\langle \mathrm{Ad}^*(g)(f), v\rangle := \langle f, \mathrm{Ad}(g^{-1})v\rangle, \tag{2.3}$$

which, of course, infinitesimalizes to ad* with the above definitions. The pairing $\langle\ ,\ \rangle$ becomes G-invariant:

$$\langle \mathrm{Ad}^*(g)(f), \mathrm{Ad}(g)v\rangle = \langle f, v\rangle, \quad v \in \mathscr{G}, \quad f \in \mathscr{G}^*.$$

Given $f_0 \in \mathscr{G}^*$ the orbit $M_{f_0} := \{\mathrm{Ad}^*(g)(f_0), g \in G\}$ is a homogeneous manifold:

$$M_{f_0} \simeq G/G_{f_0} \tag{2.4}$$

with G_{f_0} the isotropy group of f_0, i.e.,

$$G_{f_0} = \{g \in G / \mathrm{Ad}^*(g)f_0 = f_0\}. \tag{2.5}$$

Here, G_{f_0} is a closed subgroup of G, thus a Lie group. Clearly,

$$M_{f_0} \simeq M_{\mathrm{Ad}^*(g)f_0}, \quad g \in G.$$

The isotropy algebra $\mathscr{G}_{f_0}$ (the Lie algebra of G_{f_0}) is then characterized as follows:

$$\mathscr{G}_{f_0} = \{u \in \mathscr{G} / \langle f_0, [u, v]\rangle = 0, \quad \forall\, v \in \mathscr{G}\}. \tag{2.6}$$

Clearly,

$$\mathscr{G}_{\mathrm{Ad}^*(g)f_0} \simeq \mathrm{Ad}(g)\mathscr{G}_{f_0}, \quad g \in G.$$

Here, M_{f_0} is actually a symplectic manifold, i.e., it is endowed with a closed, nondegenerate two-form B (the KKS form) defined as

$$B(u_f, v_f) = B(\mathrm{ad}^*(u)f, \mathrm{ad}^*(v)f) := \langle f, [u, v]\rangle, \tag{2.7}$$

$f \in M_{f_0}$, $u, v \in \mathscr{G}$. The claimed properties of B are easily verified, the second one being evident from the very definition of M_{f_0}. Moreover B is G-invariant:

$$(\mathrm{Ad}^*(g))^* B = B. \tag{2.8}$$

Thus G acts on M_{f_0} via canonical transformations. The action ad* is Hamiltonian, i.e., ($\lrcorner$ denotes contraction)

$$u_f \lrcorner B = (d\lambda_u)_f, \tag{2.9}$$

where $\lambda_u(f) = \langle f, u\rangle$, $f \in M_{f_0}$. In fact,

$$\begin{aligned}(d\lambda_u)_f(v_f) &= \lambda_u(\mathrm{ad}^*(v)f) = \langle \mathrm{ad}^*(v)f, u\rangle = \langle f, [u, v]\rangle \\ &= B(u_f, v_f) = (u_f \lrcorner B)(v_f).\end{aligned}$$

Defining Poisson brackets through the position:

$$\{\lambda_u, \lambda_v\}(f) = B(u_f, v_f) \tag{2.10}$$

we have trivially

$$\{\lambda_u, \lambda_v\}(f) = \lambda_{[u, v]}(f). \tag{2.11}$$

We shall call the Lie algebra $(\Lambda, \{\ ,\ \})$, $\Lambda = \{\lambda_u, u \in \mathscr{G}\}$ the current algebra.

Since B is closed, it determines an element $[B]$ in $H^2(M_{f_0}, R)$, the second De Rham (or Čech) cohomology group. If (and only if) $[B]$ is integral, namely,

$$[B] \in H^2(M_{f_0}, Z),$$

there exists a Hermitian line bundle (L, h) on M_{f_0}, equipped with a connection ∇ compatible with the metric h, i.e.,

$$X(h(s, s')) = h(\nabla_X s, s') + h(s, \nabla_X s'), \tag{2.12}$$

where s, s' are smooth sections of L, X a vector field on M_{f_0}, such that (the pull-back to M_{f_0} of) its curvature form Ω is equal to $-2\pi i B$, whence the first Chern class $C_1(L)$ is equal to $[B]$. We recall that

$$\Omega(X, Y) = [\nabla_X, \nabla_Y] - \nabla_{[X, Y]}, \quad X, Y \in \mathfrak{X}(M_{f_0}). \tag{2.13}$$

Kostant's theorem[12] asserts that $[B]$ is integral iff the Lie algebra map (infinitesimal character)

$$\chi_{f_0} : \mathscr{G}_{f_0} \to iR$$

given by

$$\chi_{f_0}(u) := 2\pi i \langle f_0, u\rangle \tag{2.14}$$

extends to a one-dimensional representation χ of $\mathcal{G}_{f_0}$. When this is the case, the representation space for the group G is given by $\mathcal{H}(L)$ the space of square integrable sections of L with respect to h [Hilbert space completion of $\Gamma(L)$, the space of smooth sections of L] with

$$\|s\|^2 = \int_{M_{f_0}} h(s,s)(f) B^n/n!, \quad s \in \mathcal{H}(L), \quad x \in M_{f_0}, \tag{2.15}$$

where $n = \dim M_{f_0}/2$ and $B^n/n!$ is the natural Liouville volume on (M_{f_0}, B).

At the infinitesimal level $\mathcal{G}$ acts on (a suitable dense domain of) $\mathcal{H}(L)$ via the formula

$$(\tilde{u}\cdot s)(f) := (\nabla_{u_f} s)(f) + 2\pi i \lambda_u(f) s(f) \tag{2.16}$$

(the condition $[\tilde{u}, \tilde{v}] = \widetilde{[u, v]}$ is easily verified). In particular, if $u \in \mathcal{G}_{f_0}$, $\nabla_{u_{f_0}} \equiv 0$ and

$$(\tilde{u}\cdot s)(f_0) = 2\pi i \lambda_u(f_0) s(f_0), \tag{2.17}$$

and by hypothesis, this expression integrates to a one-dimensional representation of $\mathcal{G}_{f_0}$, so $\mathcal{H}(L)$ can be thought of as being obtained from this representation via the standard Mackey's procedure (see Ref. 11).

In the general case, the above construction does not yield an irreducible representation of G. Kostant's notion of polarization furnishes a mechanism that overcomes this problem. We shall illustrate this notion only in one case, namely, when M_{f_0} is endowed with a (G-invariant) Kaehler structure. This means that M_{f_0} is a complex manifold and in terms of any system of local complex coordinates $[z = (z_1,\ldots, z_n)]$

$$B = \frac{i}{2}\sum_{i,j} h_{ij}(z) dz^i \wedge d\bar{z}^j, \tag{2.18}$$

with $H = (h_{ij})$, $i, j = 1,\ldots, n$ Hermitian and positive; this being the case, it is always possible to find locally a function F on M_{f_0} (the Kaehler potential) such that

$$i\partial\bar{\partial}F = B. \tag{2.19}$$

Then G acts on M_{f_0} through biholomorphic transformations. Given the integrality condition, L becomes a holomorphic line bundle. The metric h is given locally by

$$\begin{aligned} h(s,s')(f) &= \exp(-F(f))\,\overline{s(f)}s'(f) \\ &\equiv h(f)\,\overline{s(f)}s'(f), \end{aligned} \tag{2.20}$$

(s, s' sections of L) and the connection ∇ takes locally the form:

$$\begin{aligned} \nabla_{\bar{Z}} s &= (\bar{Z}\cdot s)(f) s_0, \\ \nabla_Z s &= (Z\cdot s)(f) s_0 - 2\pi(Z\cdot F) s_0, \end{aligned} \tag{2.21}$$

where $s = s(f)s_0$, s_0 is a local nowhere vanishing holomorphic section (local frame), Z ($\bar{Z}$) is a type $(1, 0)$ ($(0, 1)$) vector field, i.e., a vector field of the form

$$X - iJX \quad (X + iJX), \tag{2.22}$$

$X \in \mathfrak{X}(M_{f_0})$ and J the complex structure on M_{f_0} (see below and Refs. 17 and 18). The natural candidate for $\mathcal{H}'(L)$ is then the space of square integrable holomorphic sections, i.e., those sections in $\mathcal{H}(L)$ for which

$$\nabla_{\bar{Z}} s = 0 \tag{2.23}$$

$\forall Z$ since, according to the "fundamental lemma of Hermitian geometry" (see, e.g., Ref. 17), ∇ defined above is the only connection compatible with h and the holomorphic structure of L. $\mathcal{H}'(L)$ should not be empty, of course: in the compact case according to Kodaira's theory, this is not the case if $C_1(L)$ is sufficiently "positive."[19]

According to the Borel–Weil–Bott theory, the irreducible representations of compact semisimple Lie groups arise in this way. Such "holomorphic" representations are the most natural from the point of view of the theory of generalized coherent states.[15,16,20,21] Explicitly, following Ref. 20, the (projective) representation of G on $\mathcal{H}'(L)$ is exhibited by the formula (valid on a trivialization of L on an open dense set in M):

$$(T_g s)(z) = e^{-\psi(g^{-1}, z)} s(g^{-1}z), \tag{2.24}$$

where ψ is an analytic function of z at fixed g, fulfilling

$$2\,\mathrm{Re}\,\psi(g, z) = F(gz, \overline{gz}) - F(z, \bar{z}), \tag{2.25}$$

with its imaginary part determined up to a two-cocycle of G.

Before turning to the applications, we recall the notion of almost complex structure[18] on a smooth manifold M; an almost complex structure J is a smooth section in Γ (End TM) fulfilling $J^2 = -I$, i.e.,

$$\begin{aligned} &J: x \to J_x, \\ &J: T_x M \to T_x M, \\ &J_x^2 = -I_x \quad (\text{identity on } T_x M). \end{aligned} \tag{2.26}$$

The Nijenhujs tensor N of J is defined through the formula

$$\begin{aligned} 4N(X, Y) = {}& [X, Y] + J[JX, Y] \\ & + J[X, JY] - [JX, JY], \end{aligned} \tag{2.27}$$

where $X, Y \in \mathfrak{X}(M)$. The Newlander–Nirenberg theorem[18] asserts that J is integrable, i.e., M becomes a complex manifold iff $N = 0$. This condition is also equivalent to the following: the vector fields of type $(1, 0)$ ($(0, 1)$), defined through (2.22), should form Lie algebras. In this case J is called a complex structure.

III. THE GEOMETRIC STRUCTURE OF THE RASETTI–REGGE THEORY

The group G which is to be considered when dealing with the theory of perfect incompressible fluids is SDiff (M), the group of measure preserving diffeomorphisms of a smooth Riemannian manifold (M, g) (the measure being induced from the Riemannian volume.[1,7]) In this paper we shall be concerned with $G = \mathrm{SDiff}(R^3)$ (R^3 equipped with the usual Euclidean metric). G is actually not a Lie group in the usual sense[1]; nevertheless, following Ref. 7 we shall imitate the steps taken in the finite-dimensional situation. The Lie algebra $\mathcal{G}$ of G consists of the smooth divergence-free vector fields of R^3:

$$\mathcal{G} = \left\{ \mathbf{v} = (v_1, v_2, v_3)/\mathrm{div}\,\mathbf{v} = \sum_{i=1}^{3} \frac{\partial v_i}{\partial x^i} = 0, \quad (x^i) \in R^3 \right\}.$$

The Lie bracket [,] is given by (minus) the usual bracket between vector fields.

In view of the identity

$$\mathrm{curl}(\mathbf{v}\times\mathbf{u}) = [\mathbf{v},\mathbf{u}] = (\mathbf{u}\cdot\nabla)\mathbf{v} - (\mathbf{v}\cdot\nabla)\mathbf{u} \tag{3.1}$$

($\times$ denoting vector product in R^3) and since div curl $= 0$, it is checked that $\mathscr{G}$ is indeed a Lie algebra.

Here, $\mathscr{G}$ can also be identified with the Lie algebra of right invariant vector fields on G via the map:

$$\mathbf{v}(x) \to [\eta \to v_\eta \equiv \mathbf{v}(\eta(x))]$$

$\eta \in G$, $V_\eta \in T_\eta\, G$, $T_1\, G = \mathscr{G}$; this isomorphism expresses the transition from an Eulerian to a Lagrangian point of view. The adjoint action of $\mathscr{G}$ on $\mathscr{G}$ is given by

$$\mathrm{ad}(\mathbf{u}_1)\mathbf{u}_2 = [\mathbf{u}_1, \mathbf{u}_2], \quad \mathbf{u}_1, \mathbf{u}_2 \in \mathscr{G}.$$

Following Ref. 7 a coadjoint orbit of G will be specified by fixing the velocity field $\mathbf{v}$ regarded as an element in $\mathscr{G}^*$.

Equivalently, we can use $\mathbf{w} = \mathrm{curl}\,\mathbf{v}$ (the vorticity field) or, in terms of forms, the two-form w on R^3 such that $*w$ has the same coefficients as $\mathbf{w}$ (everything is in Cartesian coordinates), where $*$ is the Hodge operator in R^3. The coadjoint action on forms naturally becomes Lie derivation.[7]

We recall the following:

Theorem (Marsden–Weinstein[7]): The Kirillov form on M_{w} is given by

$$\begin{aligned} B_{\mathrm{w}}(L_{\mathrm{u}_1}\mathrm{w}, L_{\mathrm{u}_2}\mathrm{w}) &= \langle \mathrm{v}, [\mathrm{u}_1, \mathrm{u}_2]\rangle \\ &= \int_{R^3} \mathbf{v}\cdot[\mathbf{u}_1,\mathbf{u}_2] \\ &= \int_{R^3} (\mathbf{u}_1\times\mathbf{u}_2)\cdot\mathbf{w}, \quad (\mathrm{w} = \mathrm{curl}\,\mathbf{v}). \end{aligned} \tag{3.2}$$

Proof: Taking into account the general Kirillov formula and the above discussion immediately yields (3.2).

Let us supplement the above result by the following:

Proposition 1:

$$\mathscr{G}_w = \{\mathbf{u} \in \mathscr{G} / \mathbf{u}(x) = \alpha(x)\mathbf{w}(x), \quad \nabla\alpha\cdot\mathbf{w} = 0\}, \tag{3.3}$$

whenever $\mathbf{w}(x) \neq 0$.

Proof: Using (3.2) we have, if $\mathbf{u}_1 \in \mathscr{G}$

$$0 = \int_{R^3} \mathbf{w}\cdot(\mathbf{u}_1\times\mathbf{u}_2)d^3x, \quad \forall\, \mathbf{u}_2 \in \mathscr{G},$$

whence, whenever $\mathbf{w}(x)\neq 0$, $\mathbf{u}_1(x) = \alpha(x)\cdot\mathbf{w}(x)$ follows [the condition $\nabla\alpha(x)\cdot\mathbf{w}(x) = 0$ is an immediate consequence of the requirement div $\mathbf{u}_1 = 0$].

Let us specialize the above discussion to the case when w is a distribution supported on a closed oriented (possibly knotted) smooth curve Γ in R^3 (a classical vortex line; a finite collection of such curves can be dealt similarly)

$$\mathbf{w}(\mathrm{x}) = k\oint_\Gamma \delta^3(\mathbf{x}-\mathbf{y})d\mathbf{y}, \tag{3.4}$$

where $k \in R$ is the "strength" of the vortex. Then a point in M_{w} can also be represented by a curve obtained from Γ by the action of G. From the calculation in proposition 1 we easily obtain:

Proposition 2 (see also Ref. 7, Sec. 10): (i) In the case of a vortex line, the Kirillov form reads

$$\begin{aligned} B_{\mathrm{w}}(L_{\mathrm{u}_1}\mathrm{w}, L_{\mathrm{u}_2}\mathrm{w}) &= -k\oint \hat{t}\cdot(\mathbf{u}_1\times\mathbf{u}_2)ds \\ &= -k\oint \mathbf{u}_1(\hat{t}\times\mathbf{u}_2)ds, \end{aligned} \tag{3.5}$$

with $\hat{t}$ the unit tangent vector to Γ in a generic point of it.

(ii)

$$\mathscr{G}_w = \{\mathbf{u}\in\mathscr{G}/\mathbf{u}/\!/\hat{t} \text{ on } \Gamma\}. \tag{3.6}$$

We shall further discuss (3.5) later on. We can now state one of the main results of this paper.

Theorem 1: (i) The current algebra Λ pertaining to M_{w} consists of the functions (Rasetti–Regge currents)

$$\lambda_{\mathrm{u}}(\mathrm{v}') = \int_{R^3} \mathbf{v}'\cdot\mathbf{u}\ d^3x, \tag{3.7}$$

where $\mathbf{u}\in\mathscr{G}$, $\mathbf{v}' \in M_{\mathrm{w}}$.

(ii) The Euler equation can be written in terms of the R–R currents language as

$$\frac{\partial}{\partial t}\lambda_{\mathrm{w}}(\xi_t) = -\{\lambda_{\mathrm{w}}(\xi_t), \lambda_{\hat{\mathrm{v}}_t}(\xi_t)\}, \tag{3.8}$$

where $\lambda_{\hat{\mathrm{v}}_t}$ is the time-dependent Hamiltonian governing the motion, $\xi_t = \mathrm{Ad}^*_{\eta_t}(\xi_0)$, $\eta_t \in G$, and holds for any $\mathbf{w}\in\mathscr{G}$.

(iii) In the case of a vortex line:

$$\lambda_{\mathrm{u}}(\mathrm{v}') = \oint_\Gamma \mathbf{a}\cdot d\mathbf{y}, \quad \mathbf{u} = \mathrm{curl}\,\mathbf{a}. \tag{3.9}$$

(iv) Given any smooth oriented closed curve Γ_1 distinct from Γ, let

$$\mathbf{w}_1 = \oint_{\Gamma_1} \delta^3(\mathbf{x}-\mathbf{y})d\mathbf{y},$$

then

$$\lambda_{\mathrm{w}_1}(\mathrm{v}) = k\,\mathscr{L}(\Gamma,\Gamma_1), \tag{3.10}$$

where $\mathscr{L}(\Gamma,\Gamma_1)$ is the (Gauss) linking number. We are formally treating $\mathbf{w}$ as an element of $\mathscr{G}$.

(v) In all cases

$$I := \int_{R^3} \mathbf{w}\cdot\mathbf{v}\ d^3x \tag{3.11}$$

is a constant of motion (Kelvin's theorem).

Proof: (i) and (iii) are straightforward, given the preceding discussion. Thus we have recovered the R–R current algebra in a completely intrinsic way.

(ii) The (vorticity) form of the Euler equation reads:

$$-\frac{\partial\mathbf{w}_t}{\partial t} = [\mathbf{w}_t, \mathbf{v}_t] \quad ([\mathbf{w}_t, \mathbf{v}_t] = \mathrm{curl}(\mathbf{w}_t\times\mathbf{v}_t)),$$

where $\mathrm{w}_t = \mathrm{Ad}_{\eta_t}(\mathrm{w})$ and $\mathbf{v}_t$ is the right logarithmic derivative of η_t, the Euler flow (see Ref. 22). Set $\hat{\mathrm{v}}_t = \mathrm{Ad}_{\eta_t^{-1}}(\mathrm{v}_t)$. Then, for any $\xi\in\mathscr{G}^*$,

$$\int_{R^3} \xi(x)\frac{\partial}{\partial t}\mathrm{w}_t(x)\ d^3x = \int_{R^3}\xi(x)\cdot[\mathbf{w}_t(x), \mathbf{v}_t(x)]\ d^3x \tag{3.12}$$

and resorting to (2.3)

$$\frac{\partial}{\partial t}\int_{R^3}\mathrm{Ad}^*_{\eta_t}(\xi)\cdot\mathbf{w}\ d^3x = \int_{R^3}\xi\cdot\mathrm{Ad}_{\eta_t}([\mathbf{w},\mathbf{v}_t])\ d^3x, \tag{3.13}$$

i.e., using the definition of P. B.

$$\frac{\partial}{\partial t}\lambda_{w}(\xi_t) = \{\lambda_{\hat{v}_t}(\xi_t), \lambda_{w}(\xi_t)\}.$$

(iv) A direct computation yields

$$\lambda_{w_1}(v) = \int_{R^3} d^3x \ \mathbf{v}\cdot\mathbf{w}_1 = \oint_{\Gamma_1} \mathbf{v}\cdot d\mathbf{y} \quad (\mathbf{w} = \text{curl } \mathbf{v}). \tag{3.14}$$

Now

$$\mathbf{v} = k \text{ curl} \oint_{\Gamma} \frac{d\mathbf{y}}{4\pi|\mathbf{x}-\mathbf{y}|} \tag{3.15}$$

(x not on Γ); substituting

$$\begin{aligned}\lambda_{w_1}(v) &= k\oint_{\Gamma_1} d\mathbf{x} \text{ curl} \oint_{\Gamma} \frac{d\mathbf{y}}{4\pi|\mathbf{x}-\mathbf{y}|} \\ &= k\oint_{\Gamma_1}\oint_{\Gamma} \frac{(\mathbf{x}-\mathbf{y})}{4\pi|\mathbf{x}-\mathbf{y}|^3}\cdot(d\mathbf{x}\wedge d\mathbf{y}) \\ &= k\cdot\mathscr{L}(\Gamma,\Gamma_1). \end{aligned}\tag{3.16}$$

(v) If Γ moves under the action of the velocity field (3.15), the constancy of I follows directly from the G-invariance of the pairing $\langle\ ,\ \rangle$.

We must also observe at this point that, by a theorem of Ref. 9, I, which is meaningful as an integral on S^3, for $\mathbf{w}\in\mathscr{G}$, is, up to a scalar, the Hopf invariant (see Ref. 23) of the map $f: S^3\to S^2$ given by (n-field representation[8]):

$f(x) = \hat{n}(x)\in S^2$, $x\in R^3$(stereographic coordinates of S^3)

$|\hat{n}|^2 = 1, \quad w_j(x) = 2k\varepsilon_{jab}\, \hat{n}\cdot(\partial_a \hat{n}\times\partial_b \hat{n}).$

Remark: The topological meaning of I can also be obtained by interpreting v as a connection one-form, with $\mathbf{w}$ its curvature two-form so that I is the Chern Simons action for an Abelian gauge theory.[24,25] In this language $\lambda_{w_1}(v)$ gives the exponent of the holonomy of $\mathbf{v}$ around Γ_1.

IV. THE KAEHLER STRUCTURE OF VORTEX COADJOINT ORBITS

We begin with the following important observation:

Theorem 2: M_w is a homogeneous Kaehler manifold. A local Kaehler potential is given by the formula:

$$F(\Gamma) = \oint_{\Gamma} \log(1+|z(s)|^2)ds, \tag{4.1}$$

if $\{z(s)\}_{s\in\Gamma}$ are suitable local complex coordinates of Γ.

Proof: Recalling Proposition 2, we realize that the tangent space $T_\Gamma M_w$ is given by smooth periodic functions

$\mathbf{u}: s\in\Gamma\to\mathbf{u}(s)\in R^3,$

(s denoting the arc length computed from an arbitrary point of Γ) with $\mathbf{u}(s)$ situated in the normal plane to Γ in s. The symplectic form becomes, for $\mathbf{u}_1, \mathbf{u}_2\in T_\Gamma M_w$ (temporarily omitting $-k$)

$$\begin{aligned}B_\Gamma(u_1,u_2) &= \oint_\Gamma \hat{t}(s)\cdot(\mathbf{u}_1(s)\times\mathbf{u}_2(s))ds \\ &= \oint_\Gamma \mathrm{w}(s)(\mathbf{u}_1(s),\mathbf{u}_2(s))ds,\end{aligned}\tag{4.2}$$

where $\mathrm{w}(s)$ is a copy of the symplectic form on the sphere $S^2\simeq P^1$, the complex projective line. Then $\mathrm{w}(s)$ becomes the Kaehler (Fubini–Study) form on P^1:

$$\begin{aligned}\mathrm{w}(s) &= \frac{dz(s)\wedge d\bar{z}(s)}{(1+|z(s)|^2)^2} \\ &= \partial\bar{\partial}\log(1+|z(s)|^2),\end{aligned}\tag{4.3}$$

where $z(s)$ is a local complex coordinate given by stereographic projection whence

$$B_\Gamma = \delta\bar{\delta}\, F(\Gamma) \tag{4.4}$$

if

$$\delta := \int_\Gamma ds\frac{\delta}{\delta z(s)}dz(s), \quad \bar{\delta} := \int_\Gamma ds \frac{\delta}{\delta\bar{z}(s)}d\bar{z}(s).$$

The theorem is proven once we ascribe a precise meaning to the coordinates $\{z(s)\}$. This is done as follows. Given a smooth curve Γ_0 in R^3, it is completely determined by fixing its length L_0, the position of a point $P_0\in\Gamma_0$ and giving its tangent vector field

$s_0\in[0,L_0]\to\hat{t}(s_0)\in R^3$

(P_0 has $s_0 = 0, L_0$). Since $\|\hat{t}(s_0)\| = 1$, Γ_0 can be associated to a curve

γ_0: $s_0\in[0,L\,]\to\gamma_0(s_0)\in S^2.$

Now, $\forall\, s_0$, consider the complex coordinate system corresponding to stereographic projection from the antipodal point to $\gamma_0\ (s_0)$: the ensuing complex coordinate of $\gamma_0(s_0)$ will be zero. Thus the reference curve is associated to the zero function

$s_0\in[0,L_0]\to z(s_0)\equiv 0\in C.$

Let us now examine the action of $\mathscr{G}$ at Γ. Let $\xi|_\Gamma$ be represented by a function

$[0,L\,]\ni s\to\xi(s)\in R^3,$

$\xi(s) = \alpha(s)\hat{b}(s)+\gamma(s)\hat{t}(s)+\beta(s)\hat{n}(s),$

with $(\hat{t},\hat{n},\hat{b})$ the Frenet trihedron (assume it everywhere defined).

Here, $\varepsilon\,\xi\,|_\Gamma$, ε infinitesimally small, carries Γ in a nearby curve $\Gamma+\delta\Gamma\equiv\Gamma'$. An easy computation involving Frenet formulas yields the position of Γ', its shape and its arc length as follows:

$$\begin{aligned}\mathbf{r}'(s) &= \mathbf{r}(s)+\varepsilon\,\xi(s), \\ s' &= s+\varepsilon\int_0^s(\dot{\gamma}(s)-\beta(s)k(s))ds, \\ \hat{t}'(s') &= \hat{t}(s)+\varepsilon\{(\dot{\alpha}(s)+\beta(s)\tau(s))\hat{b}(s) \\ &\quad+(\dot{\beta}(s)-\alpha(s)\tau(s)+\gamma(s)k(s))\hat{n}(s)\},\end{aligned}\tag{4.5}$$

where an overdot $\equiv d/ds$, and τ and k denote as usual the torsion and the curvature of Γ, respectively. Let $[0,L]\ni s\to\mathbf{r}(s)\in R^3$ the position vector of Γ (s arc length). The position vector of $\Gamma' = \Gamma+\delta\Gamma$ is then

$$\mathbf{r}'(s) = \mathbf{r}(s)+\varepsilon\{\alpha(s)\hat{b}(s)+\beta(s)\hat{n}(s)+\gamma(s)\hat{t}(s)\}.$$

Thus

$$\begin{aligned}\dot{\mathbf{r}}'(s) &= \hat{t}(s)+\varepsilon\{(-\beta k+\dot{\gamma})\hat{t}+(\dot{\beta}-\alpha\,\tau+\gamma\,k)\hat{n} \\ &\quad+(+\beta\,\tau+\dot{\alpha})\,\hat{b}\},\end{aligned}$$

$$\frac{ds'}{ds} = \|\dot{\mathbf{r}}(s)\|$$
$$= 1 + \varepsilon(-\beta k + \dot{\gamma})(+\text{higher-order terms})$$
$$\Rightarrow s' = s + \varepsilon \int_0^s (\dot{\gamma} - \beta k) ds$$

and

$$\frac{ds}{ds'} = 1 - \varepsilon(\dot{\gamma} - \beta k),$$

whence

$$\hat{t}'(s') = \frac{d\mathbf{r}'(s')}{ds'} = \frac{d\mathbf{r}'(s)}{ds} \frac{ds}{ds'} = (4.5).$$

Equation (4.5) shows how Γ' depends on $\gamma(s)$, the component of ξ parallel to Γ. This component must be ignored since it belongs to the isotropy algebra $\mathscr{G}_{w_\Gamma}$: this amounts to setting $\gamma \equiv 0$ ($\rightarrow \dot{\gamma} \equiv 0$) in (4.5) and we get the explicit coadjoint action of $\mathscr{G}$ at Γ. If also $\beta \equiv 0$, there is no length variation (this happens for instance in the self-induction approximation[7]).

Actually the position of Γ' can be reconstructed from $\mathbf{r}'(0) = \varepsilon\{\alpha(0)\,\hat{b}(0) + \beta(0)\,\hat{n}(0)\}$ and from $\hat{t}'(s')$.

Interpreting a finite transformation $g \in G$ as an infinite sequence of infinitesimal transformation and throwing the isotropy algebra component at every step, we get for $\Gamma = g\,\Gamma$

$$\begin{aligned} &\hat{t}'(s') = R(s) \cdot \hat{t}(s), \quad s' = s'(s), \\ &\mathbf{r}'(0) = \mathbf{r}'(\mathbf{r}(0)), \end{aligned} \tag{4.6}$$

where $R(s) \in \mathrm{SO}(3)$, s' is a monotonic smooth function of s, and $\mathbf{r}'(0)$ is a smooth function of $\mathbf{r}(0)$. This is possible for the elements $g \in G$ embedable in flows. In terms of the complex coordinate introduced above, (4.6) reads as

$$\begin{aligned} &\mathbf{r}'(0) = \mathbf{r}'(\mathbf{r}(0)) = \mathbf{r}'(\mathbf{r}_0(0)), \\ &z'(s') = \frac{a(s)z(s) + b(s)}{\bar{a}(s) - \bar{b}(s)z(s)}, \\ &s' = s'(s) = s'(s_0), \end{aligned} \tag{4.7}$$

with

$$\begin{bmatrix} a(s) & b(s) \\ -b(s) & a(s) \end{bmatrix} \in \mathrm{SU}(2)$$

representing $R(s)$ (determined up to ± 1).

If we disregard the "hidden" variable $\mathbf{r}'(0)$ which does not appear in the symplectic form B, we may interpret our dynamical variables as smooth, closed, oriented loops on $S^2 \simeq P^1$.

This is the precise sense in which M_w acquires a Kaehler structure.

Remark: Changing P_0 on Γ_0 amounts at a change of coordinates. A change of a reference frame in R^3 yields no change in the complex description.

Our use of this sort of "partial" complex polarization seems to fit with the analysis of Ref. 10, where the actual position of Γ is an obstacle at finding a real or complex polarization, for two-dimensional point vortices.

Remark: In passing we notice that the endomorphism

$$J: \Gamma \rightarrow J_\Gamma : T_\Gamma M_w \rightarrow T_\Gamma M_w$$

given by

$$(J_\Gamma \mathbf{u})(P) = \hat{t}(P) \times \mathbf{u}(P), \; P \in \Gamma \tag{4.8}$$

is the complex structure of M_w (since it is obviously integrable) pertaining to the complex charts of M_w given above.

V. OUTLINE OF GEOMETRIC QUANTIZATION

In this section we turn to geometric quantization: we have to tackle the question of the integrality of $[B]$. Using heuristically Kostant's theorem we shall find it related to the Feynman–Onsager relation,[2,3] namely, we have the following theorem.

Theorem 3: (i) If $[B_w]$ is integral, then k is quantized:

$$k = N h/m, \; N \in Z (F\text{–}0 \text{ relation}). \tag{5.1}$$

(ii) $B_w \wedge B_w \wedge \cdots \wedge B_w$ (p times) induces the generator of

$$H^{2p}(\Omega^+(S^2)) \simeq Z. \tag{5.2}$$

Proof: (i) If the map (compare Sec. I)

$$\chi_w : \mathscr{G}_w \rightarrow R,$$

$$\chi_w(\mathrm{u}) := 2\pi \langle \mathrm{v}, \mathrm{u} \rangle = 2\pi \int_{R^3} \mathbf{v} \cdot \mathbf{u} \; d^3x,$$

is going to be lifted to a one-dimensional representation of G_w then, in particular, for $\mathbf{u} = \mathrm{w}_1$ we have that

$$\mathrm{w}_1 \rightarrow 2\pi i \int_{R^3} \mathbf{v} \cdot \mathrm{w}_1 \, d^3x = 2\pi i k \, \mathscr{L}(\Gamma, \Gamma_1) \tag{5.3}$$

should lift to a group homomorphism $S^1 \rightarrow S^1$ if $\mathbf{w}_1$ is thought of to induce rotations on Γ_1 which is possible only if k is quantized, $k \in Z$. Introducing physical units $k = N h/m$, $N \in Z$, we have that the F–0 relation follows.

We also notice that the topological quantization à la Kuznetsov and Mikhailov[8] obtained by passing from R^3 to S^3 amounts to KKS quantization.[26]

In general, the map χ of Sec. I should be defined, if $\mathrm{u} \in \mathscr{G}_w$:

$$\chi(\exp \mathrm{u}) := \exp\left(2\pi i \int_{R^3} \mathbf{v} \cdot \mathbf{u} \, d^3x\right). \tag{5.4}$$

Using heuristically the Baker–Campbell–Hausdorff formula (see Refs. 22 and 27) we may check that it is well defined if $\mathbf{u} \in \mathscr{G}_w$, $j = 1, 2$,

$$\begin{aligned} &\chi(\exp \mathrm{u}_1 \exp \mathrm{u}_2) \\ &\quad = \chi(\exp(\mathrm{u}_1 + \mathrm{u}_2 + [\mathrm{u}_1, \mathrm{u}_2] \\ &\qquad + ([\mathrm{u}_1[\mathrm{u}_1, \mathrm{u}_2]] + [\mathrm{u}_2, [\mathrm{u}_2, \mathrm{u}_1]] + \cdots) + \cdots)) \\ &\quad = \chi(\exp \mathrm{u}_1)\chi(\exp \mathrm{u}_2). \end{aligned} \tag{5.5}$$

However, we must say at this point that exp is not even locally surjective (see Ref. 1), i.e., not every $g \in G$ in any neighborhood of the identity can be written as

$$\exp X = g$$

for some $X \in \mathscr{G}$ (so it would be embedable in a flow).

(ii) The Kaehler form B is a closed, not exact two-form since otherwise the Kaehler form on $S^2 \simeq P^1$ (see Theorem 2) would be exact:

$$
\begin{aligned}
B_w &= \oint_\Gamma w(s)\,ds \\
&= D\vartheta = (\delta + \bar{\delta})\vartheta \\
&= \oint_\Gamma ds\, d\vartheta(s) \\
&\Rightarrow w(s) = d\vartheta(s)
\end{aligned} \tag{5.6}
$$

globally on S^2, which is absurd.

If we restrict B_w to $\Omega^+(S^2)$, the space of non-self-intersecting oriented loops on S^2 we get a nonzero representative of

$$H^2(\Omega^+(S^2)) \simeq Z \quad \text{(see Ref. 28)}.$$

Since $H^{2p}(\Omega^+(S^2)) \simeq Z$ as well, (ii) is proven.

Once we have the integrality of $[B]$, the Kaehler potential, and the R–R currents, we have all the ingredients needed to build up, at least formally, representations of G à la KKS, via the formula (2.16) or (2.24) of Sec. II.

Some remarks are now in order. Reinserting $-k$ in the formulas, we see that in order to get a nontrivial representation space we have to ensure that N is nonpositive. This can always be achieved eventually by changing the orientation on S^2. The reason can be traced back to the fact that a holomorphic line bundle $\mathscr{L}$ on P possesses holomorphic sections if

$$\{C_1(\mathscr{L}),\, [P^1]\}$$

[evaluation of $C_1(\mathscr{L})$ on the fundamental two-cycle of P^1] is non-negative (see Ref. 19).

Let us observe that, using Theorem 1, and formally regarding the vorticity fields belonging to curves Γ as elements in $\mathscr{G}_w$, we see that they can effectively be used as representation labels: this is consistent with the R–R point of view: the topology of knot determines the representation. It goes without saying that the formal measure

$$d\mu(\Gamma) = \text{``}\frac{1}{\mathscr{Z}} e^{-f(\Gamma)} B^\infty(\Gamma)\text{''}$$

(with $\mathscr{Z}$ "some renormalization constant") should be given a definite mathematical meaning. There also seem to exist deep connections between RR theory and Witten's work.[24] We hope to tackle these problems in a later study.

ACKNOWLEDGMENTS

We gratefully thank Professor M. Rasetti for inspiration, encouragement, and extensive, fruitful discussions. We are also indebted to Professor T. Regge, Professor J. Rawnsley, and Professor F. Mercuri for their interest in our work and enlightening discussions. Thanks are also due to the referee for his critical remarks on a previous version of this paper.

Work of V.P. supported by INFM, sezione di Torino. Work of M.S. supported by Ministero della Pubblica Istruzione and CNR–GNAFA.

[1] D. Ebin and J. Marsden, Ann. Math. **92**, 102 (1970).
[2] M. Rasetti and T. Regge, Physica A **80**, 217 (1975).
[3] M. Rasetti and T. Regge, "Quantum Vortices," in *Highlights of Condensed-Matter Theory*, edited by F. Bassani *et al.* (Compositori, Bologna, 1985).
[4] G. Goldin, J. Math. Phys. **12**, 462 (1971).
[5] G. Goldin, R. Menikoff, and D. H. Sharp, J. Math. Phys. **21**, 650 (1980).
[6] G. Goldin, R. Menikoff, and D. H. Sharp, Phys. Rev. Lett. **51**, 2246 (1983).
[7] J. Marsden and A. Weinstein, J. Phys. D **7**, 305 (1983).
[8] E. A. Kuznetsov and A. V. Mikhailov, Phys. Lett. A **77**, 37 (1980).
[9] G. E. Volovik and V. P. Mineev, Zh. Eksp. Teor. Fiz. **72**, 2256 (1977).
[10] G. A. Goldin, R. Menikoff, and D. H. Sharp, *Diffeomorphisms Groups and Quantized Vortex Filaments, Lecture Notes in Physics*, Vol. 278 (Springer, New York, 1987), p. 360; *Quantized Vortex Filaments in Incompressible Fluids, Lecture Notes in Physics*, Vol. 278 (Springer, New York, 1987), p. 363; Phys. Rev. Lett. **58**, 2162 (1987).
[11] A. A. Kirillov, *Elements of the Theory of Representations* (Springer, Berlin, 1975), Chap. 15.
[12] B. Kostant, *Quantization and Unitary Representations, Lectures in Modern Analysis and Applications III, Lecture Notes in Mathematics*, Vol. 170 (Springer, Berlin, 1970).
[13] J. M. Souriau, *Structure des Systemes Dynamiques* (Dunod, Paris, 1970).
[14] E. Witten, Commun. Math. Phys. **114**, 1 (1988).
[15] J. H. Rawnsley, Q. J. Math. Oxford **28**, 403 (1977).
[16] E. Onofri, J. Math. Phys. **17**, 401 (1976).
[17] S. Chern, *Complex Manifolds without Potential Theory* (Springer, Berlin, 1979).
[18] A. L. Besse, *Einstein Manifolds* (Springer, Berlin, 1987).
[19] P. Griffiths and J. Harris, *Principles of Algebraic Geometry* (Wiley, New York, 1978).
[20] A. M. Perelomov, *Generalized Coherent States and Their Applications* (Springer, Berlin, 1986).
[21] M. Rasetti, Int. J. Theor. Phys. **5**, 377 (1972); **13**, 425 (1973).
[22] J. Milnor, *Remarks on Infinite-Dimensional Lie Groups*, edited by B. S. DeWitt and R. Stora, Les Houches, Session XL 1983, Relativity, Groups and Topology II (Elsevier, Amsterdam, 1984).
[23] J. H. C. Whitehead, Proc. Natl. Acad. Sci. (USA) **33**, 117 (1947).
[24] E. Witten, "Quantum Fields Theory and the Jones Polynomial," to appear in Proceedings of I. A. M. P. Congress, Swansea, July, 1988, edited by I. Davies, B. Simon, and A. Truman.
[25] A. M. Polyakov, Mod. Phys. Lett. A **3**, 325 (1988).
[26] V. Penna, Ph. D. thesis, Università di Torino, 1989.
[27] R. Gilmore, *Lie Groups, Lie Algebras and Their Applications* (Wiley, New York, 1974).
[28] B. Doubrovnic, S. Novikov, A. Fomenko, *Geometrie contemporaine*, III partie (Mir, Moscow, 1984); S. P. Novikov, Russian Math. Surveys **37**(5), 1 (1982).

New aspects of the path integrational treatment of the Coulomb potential

D. P. L. Castrigiano and F. Stärk
Institut für Mathematik der Technischen Universität München, D-8000 München 2, Arcisstrasse 21, Federal Republic of Germany

(Received 18 October 1988; accepted for publication 21 June 1989)

The well-known treatment of the path integral for the Coulomb potential, by means of the Kustaanheimo–Stiefel transformation and a time transformation, is made more transparent by reducing the problem to the equality of two measures on the space of paths. For this equality two proofs are given: an elementary computational one and a short one recurring to general features of stochastic processes. It is shown that the time transformation is a special case of the well-known time change for a continuous local martingale, by which the process is changed to a Brownian motion and which is determined by its quadratic variation.

I. INTRODUCTION

In this paper we present a clear-cut derivation of the propagator for the Coulomb potential by path integration in configuration space. We use well-known tools, see Refs. 1 and 2: the Kustaanheimo–Stiefel transformation (KS) and a path-dependent time transformation (T). By them the Coulomb path integral is converted to a Gaussian one. (This approach does not make explicit use of the spherical symmetry of the system, cf. Ref. 3.)

The purpose of this paper is to show that the procedure performed in the literature[1,2] can be made much more transparent by reducing the whole problem to the equality of two measures on the space of paths. For this equality we give two proofs, one by an elementary computation and a short one by a general argument from the theory of stochastic processes. In the course of the following consideration three important points are emphasized, which remain hidden in the literature.

(i) *The two roles of the KS transformation.* As a mapping from $\mathbb{R}^4$ onto $\mathbb{R}^3$ it gives rise to a transformation from the space of paths in $\mathbb{R}^4$ to that of paths in $\mathbb{R}^3$. It also gives rise to new, i.e., non-Cartesian, coordinates for $\mathbb{R}^4$ (although it is not one to one). See Sec. III A.

(ii) *The time transformation T as a global transformation of the space of paths into itself.* It is a special case of the well-known time change for a continuous local martingale, by which the process is changed to a Brownian motion, and which is determined by its quadratic variation. See Sec. III B and Sec. V B.

(iii) *The actual path integrational content of the procedure.* This is the previously mentioned equality of two measures, see Eq. (11) and Sec. V.

II. STATING THE PROBLEM

The propagator $\rho(y,y',t)$ for the Coulomb potential is the fundamental solution of the diffusion equation

$$\left(\frac{\partial}{\partial t}-\frac{1}{2m}\Delta-\frac{e^2}{|y|}\right)\rho(y,y',t)=0,$$
$$\rho(y,y',0)=\delta^3(y-y'). \tag{1}$$

It is given by the Feynman–Kac path integral, which for the general solution $w(y,t)$ of (1) with the initial distribution $w(y)\equiv w(y,0)$, reads

$$w(y,t)=\int_{\mathscr{C}_y^3} w(\psi(t))\exp\left[\int_0^t \frac{e^2}{|\psi(s)|}\,ds\right]dW_y^3(\psi). \tag{2}$$

Here $\mathscr{C}_y^3=\{\psi:[0,\infty[\,\to\mathbb{R}^3\,|\,\psi \text{ continuous with } \psi(0)=y\}$ is the space of paths and W_y^3 denotes the Wiener measure on it. The problem consists in evaluating (2).

III. TRANSFORMATIONS

A. The KS transformation

The KS transformation is a mapping k from $\mathbb{R}^4$ onto $\mathbb{R}^3$ given by $y=k(x)$ with

$$\begin{aligned} y_1&=x_1^2-x_2^2-x_3^2+x_4^2,\\ y_2&=2x_1x_2-2x_3x_4,\\ y_3&=2x_1x_3+2x_2x_4. \end{aligned} \tag{3}$$

Particularly interesting in the sequel is its property

$$|y|=x^2. \tag{4}$$

Obviously k induces the *transformation of paths*

$$K:\ \mathscr{C}_x^4\to\mathscr{C}_y^3,\quad K\varphi:=k\circ\varphi,\quad \text{with } y=k(x) \tag{5}$$

for any $x\in\mathbb{R}^4$, and where $\mathscr{C}_x^4$ denotes the space of paths in $\mathbb{R}^4$ starting at x.

The KS *coordinates* $(y,\varphi)\in\mathbb{R}^3\times[-\pi,\pi[$ of $x\in\mathbb{R}^4$ are given by

$$\begin{aligned} &x_1=(1/2\sigma)(y_2\cos\varphi+y_3\sin\varphi),\quad x_2=\sigma\cos\varphi,\\ &x_3=\sigma\sin\varphi,\quad x_4=(1/2\sigma)(y_3\cos\varphi-y_2\sin\varphi), \end{aligned} \tag{6}$$

with $\sigma:=[\tfrac{1}{2}(|y|-y_1)]^{1/2}$.

It is easy to see that $x=x(y,\varphi)$ runs exactly through the set of inverse images of y under k if φ runs through $[-\pi,\pi[$. In this context the coordinates (6) are interesting because of the integration formula

$$\int_{\mathbb{R}^4} g(x)\,d^4x=\int_{\mathbb{R}^3}\int_{-\pi}^{\pi} g(x(y,\varphi))\,d\varphi\,\frac{d^3y}{8|y|}. \tag{7}$$

 0022-2488/89/122785-04$02.50

B. The T transformation

To each path $\psi\in\mathscr{C}_x^3$, a new time is associated by

$$u:=u_\psi(t):=4\int_0^t|\psi(s)|ds. \tag{8}$$

For almost all ψ one has $\dot u>0$ and $u(\infty)=\infty$, see Ref. 4. In particular, one gets for the inverse mapping of (8) the implicit relation

$$t=t_\psi(u)=\frac{1}{4}\int_0^u\frac{ds}{|\psi(t_\psi(s))|}. \tag{9}$$

It induces the transformation of paths, the *new-time* transformation

$$T:\mathscr{C}_y^3\to\mathscr{C}_y^3,\quad T\psi:=\psi\circ t_\psi. \tag{10}$$

IV. COULOMB PROPAGATOR

The main result of this paper is contained in the following formula:

$$W_y^3=TK(W_x^4),\quad \text{with } y=k(x), \tag{11'}$$

where TK is the composition of the transformations (10) and (5), and where the right side denotes the image of the Wiener measure on $\mathscr{C}_x^3$ under TK. If L is any functional on $\mathscr{C}_y^3$, then (11') is restated equivalently by the formula

$$\int_{\mathscr{C}_y^3}L(\psi)dW_y^3(\psi)=\int_{\mathscr{C}_x^4}L(TK\varphi)dW_x^4(\varphi). \tag{11''}$$

We shall give two proofs of (11) in Sec. V.

In this section let us take (11) for granted. Using (11) it is easy the convert (2) to a Gaussian path integral without performing any path integration. Indeed, applying (11") to (2) and using the relation (9) one has with the abbreviation $\chi:=K\varphi$,

$$\begin{aligned}
w(y,u)&=\int_{\mathscr{C}_x^4}w(\chi(t_\chi(u)))\exp(4e^2t_\chi(u))dW_x^4(\varphi)\\
&=\int_0^\infty dr\int_{\mathscr{C}_x^4}\delta(t_\chi(u)-r)w(\chi(r))e^{4e^2r}dW_x^4(\varphi)\\
&\overset{(8),(9)}{=}\int_0^\infty dr\int_{\mathscr{C}_x^4}\delta(u-u_\chi(r))4|\chi(r)|\\
&\quad\times w(\chi(r))e^{4e^2r}dW_x^4(\varphi)\\
&\overset{(4)}{=}\frac{1}{2\pi}\int_{-\infty}^{\infty}dE\,e^{iEu}\int_0^\infty dr\\
&\quad\times\left[\int_{\mathscr{C}_x^4}\tilde w(\varphi(r))\exp\left(\int_0^r V_E(\varphi(s))ds\right)dW_x^4(\varphi)\right].
\end{aligned} \tag{12}$$

Here, as announced, the potential

$$V_E(x):=-4iEx^2+4e^2 \tag{13}$$

is quadratic and, therefore, can be treated by path integration. Dependent on the initial distribution $w(y)$, one obtains

$$\tilde w(x):=4x^2w(k(x)) \tag{14}$$

as the initial distribution for (13).

By the Feynman–Kac formula the remaining path integral in (12) yields the solution of the diffusion equation for the (generalized) oscillator potential (13) with the initial distribution (14). It is well known how to compute the propagator for (13) by path integration; it follows

$$\begin{aligned}
\rho(E;x,x',r)&=\frac{2iEm\exp(4e^2r)}{\pi^2sh^2(r\sqrt{8iE/m})}\\
&\quad\times\exp\{-\sqrt{iEm/2}[(x-x')^2\,\mathrm{cth}(r\sqrt{2iE/m})\\
&\quad+(x+x')^2\mathrm{th}(r\sqrt{2iE/m})]\}.
\end{aligned} \tag{15}$$

So in the case $w(y)=\delta^3(y-y')$ the path integral in (12) yields explicitly

$$\begin{aligned}
w(E;y,y',r)&\equiv\int4x'^2\delta^3(k(x')-y')\rho(E;x,x',r)d^4x'\\
&=\frac{2iEm\exp(4e^2r)}{\pi sh^2(r\sqrt{8iE/m})}\exp[-\sqrt{2iEm}\\
&\quad\times\mathrm{cth}(r\sqrt{8iE/m})(|y|+|y'|)]\\
&\quad\times I_0\left[\frac{\sqrt{8iEm}}{sh(r\sqrt{8iE/m})}\left(\frac{1}{2}(|y|\,|y'|+yy')\right)^{1/2}\right].
\end{aligned} \tag{16}$$

As expected, the integral in (16) does not depend on x but only on its image $y\equiv k(x)$ under the KS-transformation. The integral is not hard to do with the aid of the integration formula (7). Here I_0 denotes the modified Bessel function of the first kind. Combining (12) and (16) one obtains for the Coulomb propagator

$$\rho(y,y',u)=\frac{1}{2\pi}\int_{-\infty}^{\infty}dE\,e^{iEu}\int_0^\infty dr\,w(E;y,y',r). \tag{17}$$

For a further evaluation of (17) see the partial-wave decomposition of the Green's function in Ref. 2.

Remark: Because of the factor $\exp(4e^2r)$ the integrals in (17) do not exist in the Lebesgue sense. So there is the question about the meaning of (17). For instance, it is not hard to show by means of Fubini's theorem and the theorem of monotone convergence that

$$\begin{aligned}
\rho(y,y',u)&=\lim_{\varepsilon\downarrow0}\frac{1}{2\pi}\int_{-\infty}^{\infty}dE\,e^{iEu}\\
&\quad\times\int_0^\infty dr\,w(E;y,y',r)e^{-\varepsilon r^2}.
\end{aligned}$$

V. EQUALITY OF $W_{k(x)}^3$ and $TK(W_x^4)$.

It remains to show this equality, which constitutes the path integral content of our derivation of the Coulomb propagator. We give two proofs of it.

A. Computational proof

It suffices to show (11") for cylindrical functionals,

$$L(\psi)=F(\psi(u_1),\psi(u_2),\ldots,\psi(u_m)), \tag{18}$$

where $0<u_1<u_2<\cdots<u_m$ and F is any continuous and bounded real-valued function on $(\mathbb{R}^3)^m$. We first consider the case $m=1$ and then generalize the result to an arbitrary m. Applying the Fourier transformation, as in (12), we obtain for the right side of (11")

$$\int_{\mathscr{C}_x^4} L(TK\varphi)dW_x^4(\varphi)$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} dE\, e^{iEu}\int_0^{\infty} dr\Big(\int_{\mathscr{C}_x^4} 4|\chi(r)|F(\chi(r))$$
$$\times \exp(-iEu_\chi(r))dW_x^4(\varphi)\Big), \quad (19)$$

with $u = u_1$ and $\chi = K\varphi$. In order to evaluate the path integral $I \equiv I(E;x,r)$ in (19), one passes to a discrete version I_N of it approximating $u_\chi(r)$ by

$$4\frac{r}{N}\sum_{n=1}^{N-1}\left|\varphi\left(\frac{r}{N}n\right)\right|^2. \quad (20)$$

Let

$$p^{(d)}(a,b,s) := \left(\frac{m}{2\pi s}\right)^{d/2}\exp\left(-\frac{m}{2s}|a-b|^2\right)$$

denote the free propagator in d dimensions. Then by definition

$$I_N = \int_{\mathbf{R}^4}\cdots\int_{\mathbf{R}^4} d^4x_1\cdots d^4x_{N-1}\, d^4x'\, 4x'^2F(k(x'))$$
$$\times\exp\Big(-i4E\frac{r}{N}(x_1^2+\cdots+x_{N-1}^2)\Big)\cdot p^{(4)}\Big(x,x_1,\frac{r}{N}\Big)$$
$$\times p^{(4)}\Big(x_1,x_2,\frac{r}{N}\Big)\cdots p^{(4)}\Big(x_{N-1},x',\frac{r}{N}\Big). \quad (21)$$

The integrals with respect to the variables $x_1,...,x_{N-1}$ are Gaussian and can be done. It follows by induction that

$$I_N = \frac{N^2}{4\pi^2r^2}\left(1-\frac{a_N}{a_{N-1}}\right)\int_{\mathbf{R}^4} d^4x'\, 4x'^2F(k(x'))$$
$$\times\exp\left\{\left[\frac{N}{r}\left(\frac{1}{2}-a_N\right)+i4E\frac{r}{N}(1-a_N)\right](x^2+x'^2)\right.$$
$$\left.+\frac{N}{r}\left(1-\frac{a_N}{a_{N-1}}\right)^{1/2}xx'\right\}, \quad (22)$$

where

$$a_N := \frac{1-q^{N+1}}{1-q}, \quad \text{with } q := \frac{1+\sqrt{1-4v}}{1-\sqrt{1-4v}}$$
$$\text{and } v := \left[2+\frac{i8Er^2}{N^2}\right]^{-2}. \quad (23)$$

The path integral I in (19) is the limit of I_N for $N\to\infty$. Since

$$a_N/N \underset{N\to\infty}{\to} (e^\delta-1)/\delta \text{ with } \delta := 4r\sqrt{2iE}, \quad (24)$$

one obtains

$$I(E;x,r) = \int_{\mathbf{R}^4} 4x'^2F(k(x'))\rho_0(E;x,x',r)d^4x', \quad (25)$$

where ρ_0 is given by (15) in the case of zero charge $e=0$. Thus applying the definition of the Wiener measure to the left side of (11″), it remains to show [cf. (19)] that

$$\int F(y')p^{(3)}(y,y',u)d^3y'$$
$$= \frac{1}{2\pi}\int_{-\infty}^{\infty} dE\, e^{iEu}\int_0^{\infty} dr\, I(E;x,r), \quad (26)$$

with $y = k(x)$. This equation contains only ordinary integrals and relates the propagator of the three-dimensional free particle to that of a four-dimensional oscillator. Of course, (26) is the special case of (17) for vanishing charge e. It is verified by elementary computations using (7).

Now, we accomplish the proof of (11″) considering the case of a general functional (18). Then the right side of (19) is replaced with

$$\int_{-\infty}^{\infty}\frac{dE_1}{2\pi}e^{iE_1u_1}\int_0^{\infty}dr_1\int_{-\infty}^{\infty}\frac{dE_2}{2\pi}e^{iE_2u_2}\int_{r_1}^{\infty}dr_2\cdots\int_{-\infty}^{\infty}\frac{dE_m}{2\pi}e^{iE_mu_m}\int_{r_{m-1}}^{\infty}dr_m$$
$$\times\left[\int_{\mathscr{C}_x^4}4|\chi(r_1)|\cdots4|\chi(r_m)|F(\chi(r_1),\cdots,\chi(r_m))\exp(-iE_1u_\chi(r_1)-\cdots-iE_mu_\chi(r_m))dW_x^4(\varphi)\right]. \quad (27)$$

Introducing $\varepsilon_j := \Sigma_{\iota=j}^m E_\iota$ and $s_j := r_j - r_{j-1}, j = 1,...,m$, $r_0 := 0$ one obtains $\Sigma_{j=1}^m E_ju_\chi(r_j) = \Sigma_{j=1}^m\varepsilon_j4\int_{r_{j-1}}^{r_j}|\varphi(s)|^2\,ds$and, hence, by a computation analogous to that for $m = 1$,

$$\int_{\mathbf{R}^4}\cdots\int_{\mathbf{R}^4}d^4x_1\cdots d^4x_m\,4x_1^2\cdots4x_m^2F(k(x_1),\cdots,k(x_m))\rho_0(\varepsilon_1;x,x_1,s_1)\rho_0(\varepsilon_2;x_1,x_2,s_2)\cdots\rho_0(\varepsilon_m;x_{m-1},x_m,s_m) \quad (28)$$

for the path integral. Now $\varepsilon_1,\cdots,\varepsilon_m$ and $s_1,...,s_j$ are chosen as new integration variables, replacing $E_1,...,E_m$ and $r_1,...,r_m$, respectively. By (28) and by this change of variables (27) becomes

$$\int_{\mathbf{R}^4}\cdots\int_{\mathbf{R}^4}d^4x_1\cdots d^4x_m\cdot\prod_{j=1}^m\left\{4x_j^2\int_{-\infty}^{\infty}\frac{\varepsilon_j}{2\pi}\int_0^{\infty}ds_j\exp[i\varepsilon_j(u_j-u_{j-1})]\rho_0(\varepsilon_j;x_{j-1},x_j,s_j)\right\}\cdot F(k(x_1),\cdots,(kx_m)) \quad (29)$$

with $u_0 := 0$ and $x_0 := x$. Here the variables x_m up to x_1 are partly integrated out applying (26). We end up with

$$\int_{\mathbf{R}^3}\cdots\int_{\mathbf{R}^3}F(y_1,...,y_m)p^{(3)}(y,y_1,u_1)p^{(3)}(y_1,y_2,u_2-u_1)$$
$$\cdots p^{(3)}(y_{m-1},y_m,u_m-u_{m-1})d^3y_1\cdots d^3y_m \quad (30)$$

and $y = k(x)$. This is the left side of (11″) by definition of the Wiener measure.

B. General argument

For an alternative proof of (11) we use the language of stochastic processes.

Starting point is the four-dimensional Brownian motion $(X_t)_{t\geqslant 0}$ realized as the natural process $X_t\varphi := \varphi(t)$ on the probability space $(\mathscr{C}_x^4, W_x^4)$. This gets transformed to KX by means of the KS transformation: $KX_t\varphi := k(X_t\varphi) = K\varphi(t)$, still on $(\mathscr{C}_x^4, W_x^4)$. Of course, KX is equal by law to the natural process $Y_t\psi := \psi(t)$ on $(\mathscr{C}_y^3, K(W_x^4))$, $y = k(x)$. By the Ito formula, see, e.g., Ref. 6, Sec. 5, the jth component $(j = 1,2,3)$ of the stochastic differential of KX is given by $d(KX_t)_j = m^{-1/2} M_j(t) dX_t$, where $M_j = \nabla(KX)_j$, i.e., explicitly

$$M_1 = 2(X_1, -X_2, -X_3, X_4),$$
$$M_2 = 2(X_2, X_1, -X_4, -X_3), \tag{31}$$
$$M_3 = 2(X_3, X_4, X_1, X_2).$$

Hence it follows that the quadratic variation $\langle (KX)_j \rangle_t$ is given by $\int_0^t M_j(s)^2\, ds$ (cf. Ref. 6, Sec. 5), which by (31) becomes $4\int_0^t X_s^2 ds$ independently of j. Transferred to Y this result reads, because of (4),

$$\langle Y_j \rangle_t = 4\int_0^t |Y_s| ds, \tag{32}$$

independent of $j = 1,2,3$. Evaluated for an individual path ψ the right side of (32) yields exactly (8).

It is not by chance that (32) and (8) coincide. Indeed, the quadratic variation of a continuous local martingale constitutes an *intrinsic clock*. By means of it the martingale can be time changed to a Brownian motion. Explicitly,

$$(Y_{t(u)})_{u\geqslant 0}, \quad \text{with } t(u) \text{ given by (9)} \tag{33}$$

is indistinguishable from three-dimensional Brownian motion. The proof follows by the same kind of reasoning as in Ref. 6, Sec. 9.3, where the one-dimensional case is treated. From this the assertion (11) follows immediately.

We should emphasize that this relation between the new-time transformation T and the intrinsic clock is revealed in Ref. 7, see also Ref. 8.

[1]In this context these transformations have been applied for the first time in I. H. Duru and H. Kleinert, "Solution of the path integral for the H atom," Phys. Lett. B **84**, 185 (1979); "Quantum mechanics of H atom from path integrals," Fortschr. Phys. **30**, 401 (1982).

[2]Our treatment is more related to R. Ho and A. Inomata, "Exact-path-integral treatment of the hydrogen atom," Phys. Rev. Lett. **48**, 231 (1982).

[3]An alternative method to do the path integral for the H atom, which should be mentioned here, takes advantage of the spherical symmetry of the system. By it the full path integral is reduced to the one-dimensional radial part of it: A. Inomata, "Alternative exact-path-integral treatment of the hydrogen atom," Phys. Lett. A **101**, 253 (1984); F. Steiner, "Space-time transformations in radial path integrals," Phys. Lett. A **106**, 356 (1984); "Exact path integral treatment of the hydrogen atom," **106**, 363 (1984); In particular, because of the literature cited therein see also F. Steiner, in *Path Integrals from meV to MeV*, edited by M. C. Gutzwiller, A. Inomata, I. R. Klauder, and L. Streit (World Scientific, Singapore, 1986).

[4]This is true, since for a d-dimensional $(d\geqslant 3)$ Brownian motion every prescribed point (different from the starting point) is nonattainable and every compact subset is transient.

[5]F. Oberhettinger and L. Badii, *Tables of Laplace Transforms* (Springer, Berlin, 1973).

[6]K. L. Chung, R. J. Williams, *Introduction to Stochastic Integration* (Birkhäuser, Boston, 1983).

[7]Ph. Blanchard and M. Sirugue, "Treatment of some singular-potentials by change of variables in Wiener integrals," J. Math. Phys. **22**, 1372 (1981).

[8]A. Young and C. DeWitt-Morette, "Time Substitutions in Stochastic Processes as a Tool in Path Integration," Ann. Phys. **169**, 140 (1986).

On asymptotic expansions of twisted products

Ricardo Estrada, José M. Gracia-Bondía, and Joseph C. Várilly
Escuela de Matemática, Universidad de Costa Rica, San José, Costa Rica

(Received 9 August 1988; accepted for publication 14 June 1989)

The series development of the quantum-mechanical twisted product is studied. The series is shown to make sense as a moment asymptotic expansion of the integral formula for the twisted product, either pointwise or in the distributional sense depending on the nature of the factors. A condition is given that ensures convergence and is stronger than previously known results. Possible applications are examined.

I. INTRODUCTION

The self-contained approach to quantum mechanics on the flat phase spaces $\mathbb{R}^{2n}$ has been growing since the seminal paper by Moyal[1] 40 years ago. In this approach, the Weyl correspondence[2] between operators and their symbols fades into the background and the main role is assumed by the *twisted product* of two symbols, corresponding to the usual composition of operators.

The paradoxical fact is that two different definitions of twisted product coexist more or less peacefully in the literature. The first one, historically speaking, probably goes back to Groenewold's[3] paper. With the notational conventions

$$u:=(q,p)\in\mathbb{R}^{2n},\quad \alpha:=(\alpha_1,\dots,\alpha_{2n})\in\mathbb{N}^{2n},$$

$$|\alpha|:=\alpha_1+\cdots+\alpha_{2n},\quad \partial^\alpha:=\frac{\partial^{\alpha_1}}{\partial u_1^{\alpha_1}}\cdots\frac{\partial^{\alpha_{2n}}}{\partial u_{2n}^{\alpha_{2n}}},$$

$$\hat{\partial}^\alpha f=(-1)^{\alpha_{n+1}+\cdots+\alpha_{2n}}$$

$$\times\frac{\partial^{\alpha_1}}{\partial u_{n+1}^{\alpha_1}}\cdots\frac{\partial^{\alpha_n}}{\partial u_{2n}^{\alpha_n}}\frac{\partial^{\alpha_{n+1}}}{\partial u_1^{\alpha_{n+1}}}\cdots\frac{\partial^{\alpha_{2n}}}{\partial u_n^{\alpha_{2n}}},$$

it is written

$$f\times_\hbar g=\sum_{|\alpha|=0}^{\infty}\left(\frac{i\hbar}{2}\right)^{|\alpha|}\frac{1}{\alpha!}\partial^\alpha f\hat{\partial}^\alpha g. \tag{1}$$

This form was used in the influential papers by Bayen *et al.*[4] It has a geometric appeal, because the terms in the development are classical invariants.[5] It also lends itself to semiclassical considerations. Indeed, one can obtain (1) by the following heuristic argument: try to build a quantum mechanical product with the ordinary product and the classical Poisson bracket $\{\cdot,\cdot\}_P$ as raw materials. One writes

$$f*_\hbar g:=fg+(i\hbar/2)\{f,g\}_P,$$

which include the first two terms in (1). The product $*_\hbar$ fails to be associative, by a term of order $\hbar^2$. The natural correction to the previous definition includes the third term in (1); the new product fails to be associative by a term of order $\hbar^3$, and so on.

The second definition probably appeared in print for the first time in Ref. 6, although it was already implicit in a remarkable paper by Baker,[7] and is

$$(f\times_\hbar g)(u)=(\pi\hbar)^{-2n}\int_{\mathbb{R}^{2n}}\int_{\mathbb{R}^{2n}}f(u+v)g(u+w)$$
$$\times\exp\left(\frac{2i}{\hbar}\sigma[v,w]\right)d^{2n}v\,d^{2n}w. \tag{2}$$

In the formula, the symplectic form σ is defined on $\mathbb{R}^{2n}$ by

$$\sigma[v,w]=\sigma[(q,p),(q',p')]:=\sum_{i=1}^{n}q_ip'_{i+n}-p'_iq_{i+n}. \tag{3}$$

The most direct route to (2) is by group-theoretical considerations.[8] The rather obvious shortcoming of the first definition is that it is not clear what can be made of series such as (1) in general. Moreover, at least one of the factors in (1) has to be smooth; thus the validity of that formula is *a priori* restricted.

Formula (2) appears at first sight even more restrictive, as it seems to demand stringent integrability conditions. However, the important equality

$$\int_{\mathbb{R}^{2n}}(f\times_\hbar g)(u)d^{2n}u=\int_{\mathbb{R}^{2n}}f(u)g(u)d^{2n}u \tag{4}$$

makes the extension of (2) by the methods of duality (i.e., distribution theory) a relatively easy task. The strategy to follow was sketched by Antonets[9] and carried out in detail by two of the present authors in Ref. 10. Hereinafter, when we refer to (2) or to the integral form of the twisted product, we understand this extension, which we call the standard or Antonets extension; necessary details will be given later.

The main advantage of (2) is, then, that it can be extended to a large class of symbols, each step in the extension process being given a precisely defined functional-analytic meaning.

Formula (1) is handy when one of the factors to be twisted–multiplied is a polynomial. We note that formula (2) gives, of course, the same result in that case; nothing other than integration by parts is involved and, in fact, if integration by parts is formally carried on indefinitely in (2), we recover (1). On the other hand, there is no way that beautiful (and useful) consequences of (2) such as

$$\delta\times_\hbar\delta=(\pi\hbar)^{-2n}, \tag{5}$$

where δ is the Dirac measure concentrated at the origin of phase space, can be extracted from (1).

The purpose of this paper is to establish some rigorous general relations between (1) and (2). Results concerning relations of this kind are very scarce in the literature. Voros[11]

 0022-2488/89/122789-08$02.50

proved that if both f and g are GLS symbols[12,13] of class r and s, respectively (see the definition in Sec. II), then the difference between (2) and the sum of the n first terms in (1) is a GLS symbol of the class $r+s-2n$. In Ref. 10, it is proved that, under rather favorable circumstances, (1) converges to the standard extension of (2). A similar weaker result is found in Ref. 14. Kammerer also mentions a couple of results of this type without proof in Refs. 15 and 16.

Formula (1) looks like, and is, an asymptotic development of (2). By and large, questions relative to the correctness and the meaning of semiclassical expansions in the framework of phase-space quantum mechanics boil down to the link between the two formulas. The range of applicability is very wide, as such semiclassical phase-space expansions have found frequent use in nuclear physics,[17] scattering theory,[18] transport theory, and so on.

Maybe what has deterred potential researchers of the aforementioned link is that application of the usual methods of stationary phase[19] seems awkward here. Fortunately, there is no need for that, Kanwal and one of us[20,21] have developed a new distributional theory of asymptotic expansions. This theory encompasses many classical results for asymptotic developments of integrals and turns out to be very well suited to our task.

One of the advantages of the chosen framework is that a sense can be given to the asymptotic development of (2) even when one of the factors is a distribution; and thus a comparison between (1) and (2) can be made under fairly general conditions for (1) to make sense. More can be said when smoothness or analyticity are present. Roughly speaking, smoothness guarantees that (1) is the asymptotic expansion of (2) as $\hbar\downarrow 0$ in the classical sense and analyticity gives coincidence of both expressions.

In Sec. II, the Antonets extension and other function and distribution spaces employed in the sequel are introduced.

In Sec. III the distributional theory for asymptotic expansions is developed and the results on the relation between the two definitions of twisted product are proved.

In Sec. IV an application of the formulas of the previous section to computing expansions for the quantum evolution of states is given. The proof of Theorem 4—a mathematically interesting result on its own—is deferred to the Appendix.

II. SUITABLE SPACES OF FUNCTIONS AND DISTRIBUTIONS

It is a characteristic of distributional asymptotic expansions that their validity depends on the space of distributions in which the expansion is made.[21] So we first briefly discuss the relevant spaces.

We will write $x\in\mathbb{R}^n$ and, as before, $\alpha:=(\alpha_1,\ldots\alpha_n)\in\mathbb{N}^n$, $|\alpha|:=\alpha_1+\cdots+\alpha_n$, $x^\alpha:=x_1^{\alpha_1}\cdots x_n^{\alpha_n}$, and $\alpha!:=\alpha_1!\cdots\alpha_n!$.

The Schwartz space[22] $\mathscr{S}=\mathscr{S}(\mathbb{R}^n)$ is the space of smooth functions h on $\mathbb{R}^n$ such that $x^\alpha\,\partial^\beta h(x)$ vanishes at infinity for all α and β. Its dual space $\mathscr{S}'=\mathscr{S}'(\mathbb{R}^n)$ is the space of tempered distributions of $\mathbb{R}^n$. The space of all smooth functions on $\mathbb{R}^n$ is denoted $\mathscr{E}=\mathscr{E}(\mathbb{R}^n)$ and its dual is the space $\mathscr{E}'$ of distributions of compact support on $\mathbb{R}^n$.

Intermediate spaces of smooth functions are $\mathscr{O}_M$ and $\mathscr{O}_C$.[23] Here $\mathscr{O}_M$ is defined as the space of smooth functions that, together with all of their derivatives, are of polynomial growth at infinity; $\mathscr{O}_C$ is the subspace of $\mathscr{O}_M$ for which the bounding polynomials are of the same degree for all derivatives. More precisely, for any real number r we can define $\mathscr{O}_r$ as[24] the set of smooth functions f for which $\partial^\alpha f(x)=O(|x|^r)$ as $|x|\to\infty$ for all $\alpha\in\mathbb{N}^n$; then $\mathscr{O}_C=\bigcup_{r\in\mathbb{R}}\mathscr{O}_r$.

Two other useful subspaces are the space $\mathscr{K}$ consisting of those smooth functions f for which, for some r, $\partial^\alpha f(x)=O(|x|^{r-|\alpha|})$, for all $\alpha\in\mathbb{N}^n$ (for n even, $\mathscr{K}$ is called the space of "GLS symbols,"[11-13] and r is called the "class" of the symbol f); and [10] the space $\mathscr{O}_T$ of those smooth functions f for which, for some r, $\partial^\alpha f(x)=O(|x|^{r+|\alpha|})$ for all $\alpha\in\mathbb{N}^n$. Clearly $\mathscr{K}\subset\mathscr{O}_C\subset\mathscr{O}_T\subset\mathscr{O}_M$.

The Paley–Wiener–Schwartz theorem yields another space of smooth functions, $\mathscr{O}_{\exp}$. If $\mathscr{F}$ denotes the Fourier transformation, $\mathscr{F}g(t)=\int_{\mathbb{R}^n}\exp(-ix\cdot t)g(x)d_nx$, then $\mathscr{O}_{\exp}(\mathbb{R}^n):=\mathscr{F}(\mathscr{E}'(\mathbb{R}^n))$ consists of smooth functions on $\mathbb{R}^n$ that extend to entire functions on $\mathbb{C}^n$ of exponential type.[19,22] Let us write $\mathscr{Z}$ for the space of smooth functions that extend to entire functions on $\mathbb{C}^n$; then, of course, $\mathscr{O}_{\exp}\subset\mathscr{Z}\subset\mathscr{E}$.

The Fourier transformation takes each of the spaces $\mathscr{O}_M$ and $\mathscr{O}_C$ onto the dual of the other: $\mathscr{F}(\mathscr{O}_M)=\mathscr{O}'_C$, $\mathscr{F}(\mathscr{O}_C)=\mathscr{O}'_M$: see Refs. 25 and 26. The duals $\mathscr{E}'$, $\mathscr{O}'_C$, and $\mathscr{O}'_M$ are all spaces of tempered distributions.

If $f,g\in\mathscr{S}$, we write

$$\langle f,g\rangle:=\int_{\mathbb{R}^n}f(x)g(x)d^nx.$$

If $\langle T,h\rangle=\langle T(x),h(x)\rangle$ denotes the evaluation of a distribution $T\in\mathscr{S}'$ on a test function $h\in\mathscr{S}$, we can regard $\mathscr{S}$ as a subset of $\mathscr{S}'$ and the $\langle\cdot,\cdot\rangle$ notations are consistent. Now, the twisted product of two functions in $\mathscr{S}$ also belongs to $\mathscr{S}$. From (2) it is easy to verify the *tracial property* (4) of the twisted product. As an immediate consequence of (4),

$$\langle f\times_\hbar g,h\rangle=\langle f,g\times_\hbar h\rangle=\langle g,h\times_\hbar f\rangle$$
$$=\int_{\mathbb{R}^{2n}}(f\times_\hbar g\times_\hbar h)(u)d^{2n}u. \tag{6}$$

The Antonets extension of the twisted product is now defined, in two steps, as follows: first, if $T\in\mathscr{S}'(\mathbb{R}^{2n})$ and $f\in\mathscr{S}(\mathbb{R}^{2n})$, $T\times_\hbar f$ and $f\times_\hbar T$ are defined as tempered distributions by

$$\langle T\times_\hbar f,h\rangle:=\langle T,f\times_\hbar h\rangle,\quad\langle f\times_\hbar T,h\rangle:=\langle T,h\times_\hbar f\rangle, \tag{7}$$

for $h\in\mathscr{S}(\mathbb{R}^{2n})$; clearly this notation is consistent with the inclusion $\mathscr{S}\subset\mathscr{S}'$ by (6). It turns out[10] that the distributions $T\times_\hbar f$ and $f\times_\hbar T$ are, in fact, smooth functions lying in the space $\mathscr{O}_T(\mathbb{R}^{2n})$.

The second step is to introduce the multiplier algebras $\mathscr{M}_L^\hbar$ (resp. $\mathscr{M}_R^\hbar$) of those $T\in\mathscr{S}'(\mathbb{R}^{2n})$ for which $T\times_\hbar f\in\mathscr{S}(\mathbb{R}^{2n})$ [resp. $f\times_\hbar T\in\mathscr{S}(\mathbb{R}^{2n})$], for all $f\in\mathscr{S}(\mathbb{R}^{2n})$. The twisted product is extended to these spaces by defining

$$\langle T \times_\hbar S, f\rangle := \langle T, S \times_\hbar f\rangle, \quad \langle R \times_\hbar T, f\rangle := \langle T, f \times_\hbar R\rangle,$$

whenever $T \in \mathscr{S}'(\mathbb{R}^{2n})$, $S \in \mathscr{M}_L^\hbar$, $R \in \mathscr{M}_R^\hbar$, $f \in \mathscr{S}(\mathbb{R}^{2n})$. The intersection $\mathscr{M}^\hbar := \mathscr{M}_L^\hbar \cap \mathscr{M}_R^\hbar$ is then a *-algebra of distributions called the Moyal *-algebra. [The involution is complex conjugation, since $(f \times_\hbar g)^* = g^* \times_\hbar f^*$ in general.] Since we can also extend (7) by writing

$$\langle T \times_\hbar S, f\rangle := \langle S, f \times_\hbar T\rangle, \quad \langle R \times_\hbar T, f\rangle = \langle R, T \times_\hbar f\rangle,$$

whenever $R, S \in \mathscr{O}'_T(\mathbb{R}^{2n})$, we obtain that $\mathscr{O}'_T(\mathbb{R}^{2n}) \subset \mathscr{M}^\hbar$ and, as a consequence, that $\mathscr{E}'(\mathbb{R}^{2n}) \subset \mathscr{M}^\hbar$ and $\mathscr{O}'_M(\mathbb{R}^{2n}) \subset \mathscr{M}^\hbar$. (The necessary topological details to complete this formal argument are given in Ref. 10.)

It should be noted that the spaces $\mathscr{M}^\hbar$ for different values of the parameter $\hbar > 0$ are all distinct. Indeed,[27] it is straightforward to check[28] that $f_a(q,p) := \exp(\frac{1}{2} iaq \cdot p)$ lies in $\mathscr{M}^\hbar$ if $a > 0$, $a \neq \hbar$, but that $f_\hbar \notin \mathscr{M}^\hbar$.

The Fourier transformation on $\mathbb{R}^{2n}$ permutes the $\mathscr{M}^\hbar$ spaces; indeed, $\mathscr{F}(\mathscr{M}^\hbar) = \mathscr{M}^{4/\hbar}$ so that $\mathscr{M}^2$ is Fourier invariant. From this we obtain that $\mathscr{O}_{\exp}(\mathbb{R}^{2n}) \subset \mathscr{M}^\hbar$ and $\mathscr{O}_C(\mathbb{R}^{2n}) \subset \mathscr{M}^\hbar$, for all $\hbar > 0$. In particular, the twisted product is defined via the Antonets extension if at least one factor belongs to either $\mathscr{O}_{\exp}(\mathbb{R}^{2n})$ or $\mathscr{O}_C(\mathbb{R}^{2n})$ and the other belongs to $\mathscr{S}'(\mathbb{R}^{2n})$.

We do not have $\mathscr{O}_T(\mathbb{R}^{2n}) \subset \mathscr{M}^\hbar$, since $f_\hbar \in \mathscr{O}_T(\mathbb{R}^{2n})$. Nor is $\mathscr{O}_T(\mathbb{R}^{2n})$ an algebra: Lassner and Lassner[29] have given an example of two functions f and g in $\mathscr{O}_T(\mathbb{R}^{2n})$ for which $f \times_\hbar g$ is defined and not in $\mathscr{E}(\mathbb{R}^{2n})$. However, Figueroa[30] has recently shown that $\mathscr{O}_C(\mathbb{R}^{2n})$ is an algebra under the twisted product, of which the space $\mathscr{K}(\mathbb{R}^{2n})$ of GLS symbols is a subalgebra. Also, he has shown that if $f \in \mathscr{O}_M(\mathbb{R}^{2n})$, $g \in \mathscr{O}_C(\mathbb{R}^{2n})$, then $f \times_\hbar g \in \mathscr{O}_M(\mathbb{R}^{2n})$.

Finally, it should be noted that $\mathscr{O}_{\exp}(\mathbb{R}^{2n})$ is an algebra under the twisted product: indeed, it has been shown in Ref. 10 that the series (1) converges in $\mathscr{O}_{\exp}(\mathbb{R}^{2n})$ if both factors belong to $\mathscr{O}_{\exp}(\mathbb{R}^{2n})$.

III. ASYMPTOTIC EXPANSIONS

Many authors have used the techniques of generalized functions to study asymptotic expansions (see Refs. 31–33, for example.) As is shown in Ref. 21, the asymptotic expansion of generalized functions is closely related to classical problems of the asymptotic development of integrals. In many cases, the power of distribution theory permits one to obtain the asymptotic development in a very simple fashion; and, in fact, that is what happens in the present case.

One of the simplest, and yet more useful, asymptotic expansions of generalized functions is the "moment asymptotic expansion."[21] It can be written as

$$T(\lambda x) \sim \sum_{|\alpha|=0}^{\infty} \frac{(-1)^{|\alpha|}\mu_\alpha\, \partial^\alpha \delta(x)}{\alpha!\lambda^{|\alpha|+n}} \quad \text{as } \lambda \to \infty, \tag{8}$$

or, more precisely,

$$T(\lambda x) = \sum_{|\alpha|=0}^{N} \frac{(-1)^{|\alpha|}\mu_\alpha\, \partial^\alpha \delta(x)}{\alpha!\lambda^{|\alpha|+n}} + O(\lambda^{-N-n-1}) \quad \text{as } \lambda \to \infty. \tag{9}$$

where the μ_α are the moments of T, namely,

$$\mu_\alpha = \langle T(x), x^\alpha\rangle.$$

The interpretation of the asymptotic expansion of generalized functions is in the weak or distributional sense. Thus (8) or (9) means that for any test function h we have

$$\langle T(\lambda x), h(x)\rangle = \sum_{|\alpha|=0}^{N} \frac{\mu_\alpha\, \partial^\alpha h(0)}{\alpha!\lambda^{|\alpha|+n}} + O(\lambda^{-N-n-1}) \quad \text{as } \lambda \to \infty. \tag{10}$$

The moment asymptotic expansion will be valid in some spaces of generalized functions, but not in others.[21] Generally speaking, this asymptotic expansion holds if the distribution T is of "rapid decay" at infinity. More specifically, we have the following lemma.

Lemma: Let $\mathscr{X}$ be a space of smooth functions on $\mathbb{R}^n$ that contains all polynomials, and let $\mathscr{X}_q$ denote the vector subspace of those $h \in \mathscr{X}$ that vanish at 0, together with all derivatives of order less than q. Write $A_\lambda h(x) = h(x/\lambda)$ and suppose, moreover, that the topology of $\mathscr{X}$ is determined by seminorms p for which $p(A_\lambda h) = O(\lambda^{-q})$ as $\lambda \to \infty$, for all $h \in \mathscr{X}_q$. Then

$$T(\lambda x) = \sum_{|\alpha|=0}^{q} \frac{(-1)^{|\alpha|}\mu_\alpha\, \partial^\alpha \delta(x)}{\alpha!\lambda^{|\alpha|+n}} + O(\lambda^{-q-n-1}) \quad \text{as } \lambda \to \infty, \tag{11}$$

for all $T \in \mathscr{X}'$.

Proof: Choose $h \in \mathscr{X}$ and let P_q be the Taylor polynomial of order q of h at 0. Then

$$\langle T(\lambda x), P_q(x)\rangle = \sum_{|\alpha|=0}^{q} \frac{\mu_\alpha\, \partial^\alpha h(0)}{\alpha!\lambda^{|\alpha|+n}}$$

and $h_1 := h - P_q \in \mathscr{X}_{q+1}$. Moreover, for some seminorm p, we have

$$|\langle T(\lambda x), h_1(x)\rangle| = \lambda^{-n}|\langle T(x), h_1(\lambda^{-1}x)\rangle| \leqslant \lambda^{-n} C p(A_\lambda h_1) = O(\lambda^{-q-n-1}). \quad \blacksquare$$

Take, for example, $\mathscr{X} = \mathscr{E}$. It suffices to consider the family of seminorms given by

$$p_{\alpha,R}(h) := \sup\{|\partial^\alpha h(x)| : |x| \leqslant R\},$$

for $\alpha \in \mathbb{N}^n$ and $R > 0$. If $h \in \mathscr{E}_q$, $|x|^{-q}h(x)$ remains bounded as $|x| \to 0$, so $|h(x)| \leqslant C|x|^q$, for $|x| \leqslant 1$, and so

$$p_{0,R}(A_\lambda h) \leqslant \sup_{|x|\leqslant R} C|x/\lambda|^q \leqslant CR^q\lambda^{-q}, \quad \text{for } \lambda \geqslant R,$$

$$p_{\alpha,R}(A_\lambda h) \leqslant \lambda^{-|\alpha|} p_{0,R}(A_\lambda(\partial^\alpha h)) \leqslant C'\lambda^{-|\alpha|}\lambda^{-q+|\alpha|} = C'\lambda^{-q}, \quad \text{for } \lambda \geqslant R.$$

For $\mathscr{X} = \mathscr{O}_r$, write $g(t) := 1$ if $0 \leqslant t \leqslant 1$, $g(t) := t^{-r}$ if $t \geqslant 1$. Then a suitable family of seminorms for a topology of $\mathscr{O}_r$ is given by

$$p_\alpha(h) := \sup\{g(|x|)|\partial^\alpha h(x)| : x \in \mathbb{R}^n\}.$$

It is straightforward to verify that $p_\alpha(A_\lambda h) = O(\lambda^{-q})$ for $h \in \mathscr{X}_q$.

If $T \in \mathscr{O}'_C$, then $T \in \mathscr{O}'_r$, for some r, and so the expansion (11) is valid for T. On the other hand, if $T \in \mathscr{O}'_M$, $h \in \mathscr{O}_M$, then $\mathscr{F}h \in \mathscr{O}'_C$, $\mathscr{F}^{-1}T \in \mathscr{O}_C$, and so

$$\langle T(\lambda x), h(x)\rangle = \langle \mathscr{F}^{-1}T(y), \mathscr{F}h(\lambda y)\rangle$$

shows that (11) likewise holds for $T\in\mathscr{O}'_M$.

In summary, the distributional asymptotic expansion (10) is valid for h in one of the spaces $\mathscr{E}$, $\mathscr{O}_M$, or $\mathscr{O}_C$, and T in $\mathscr{E}'$, $\mathscr{O}'_M$, $\mathscr{O}'_C$, respectively. However, the expansion does *not* hold if, say, $T\in\mathscr{S}'$ and $h\in\mathscr{S}$.

The same basic technique provides partial moment expansions for functions or distributions of two variables (both in $\mathbb{R}^n$, say). In this case we write

$$T(\lambda x,y)\sim\sum_{|\alpha|=0}^{\infty}\frac{(-1)^{|\alpha|}\,\partial^\alpha\delta(x)\mu_\alpha(y)}{\alpha!\lambda^{|\alpha|+n}}\quad\text{as }\lambda\to\infty,\tag{12}$$

where the partial moments μ_α are distributions whose action on a test function g is given by

$$\langle\mu_\alpha(y),g(y)\rangle:=\langle T(x,y),x^\alpha g(y)\rangle_{x,y}.$$

(Hereinafter we subscript the symbol $\langle\cdot,\cdot\rangle$ to indicate the variables over which functions or distributions are to be integrated, whenever two or more variables are involved.)

Let us write $(T\circ g)(x):=\langle T(x,y),g(y)\rangle_y$. In order for (12) to hold for a test function $h(x,y)=f(x)g(y)$, one sees from the above lemma that f should belong to one of the spaces $\mathscr{E}$, $\mathscr{O}_M$, or $\mathscr{O}_C$, while $T\circ g$ should belong to $\mathscr{E}'$, $\mathscr{O}'_M$, or $\mathscr{O}'_C$, respectively. In such cases it is easy to show, by appropriate norm estimates, that (12) continues to hold for h in the completed (projective) tensor product of the spaces to which f and g must belong.

In particular, let us consider the Fourier kernel

$$T(x,y)=e^{ix\cdot y},\quad x,y\in\mathbb{R}^n.$$

In this case we find that

$$(T\circ g)(x)=\int_{\mathbb{R}^n}e^{ix\cdot y}g(y)d^ny=(2\pi)^n\mathscr{F}^{-1}g(x)$$

and the partial moments are given by

$$\langle\mu_\alpha,g\rangle=\int_{\mathbb{R}^n}\int_{\mathbb{R}^n}e^{ix\cdot y}x^\alpha g(y)d^nx\,d^ny$$
$$=(2\pi)^n\langle(-i)^{|\alpha|}\,\partial^\alpha\delta(y),g(y)\rangle$$

or

$$\mu_\alpha=(2\pi)^n(-i)^{|\alpha|}\,\partial^\alpha\delta.$$

Thus the partial moment expansion of the Fourier kernel takes the form

$$e^{i\lambda x\cdot y}\sim(2\pi)^n\sum_{|\alpha|=0}^{\infty}\frac{i^{|\alpha|}\,\partial^\alpha\delta(x)\partial^\alpha\delta(y)}{\alpha!\lambda^{|\alpha|+n}}\quad\text{as }\lambda\to\infty,$$

and the expansion is valid for test functions h in either of the spaces $\mathscr{E}(\mathbb{R}^n)\,\hat{\otimes}\,\mathscr{O}_{\exp}(\mathbb{R}^n)$ or $\mathscr{O}_M(\mathbb{R}^n)\,\hat{\otimes}\,\mathscr{O}_C(\mathbb{R}^n)$.

Replacing y by Ay, where A is a nonsingular matrix, allows one to develop an asymptotic expansion for $\exp(i\lambda x\cdot Ay)$. In particular, if $v,w\in\mathbb{R}^{2n}$ and if J is the matrix

$$J=\begin{pmatrix}O & I_n\\ -I_n & O\end{pmatrix},$$

so that $v\cdot Jw=\sigma[v,w]$ as in (3), we obtain

$$e^{i\lambda\sigma[v,w]}\sim(2\pi)^{2n}\sum_{|\alpha|=0}^{\infty}\frac{i^{|\alpha|}\,\partial^\alpha\delta(v)\hat{\partial}^\alpha\delta(w)}{\alpha!\lambda^{|\alpha|+2n}}\quad\text{as }\lambda\to\infty,$$

where we have used the fact that $\partial^\alpha\delta(Jw)=\hat{\partial}^\alpha\delta(w)$.

If we now use this asymptotic expansion with $\lambda=2/\hbar$ in the integral (2) that defines the twisted product, we immediately obtain that, as $\hbar\to0$,

$$(f\times_\hbar g)(u)=(\pi\hbar)^{-2n}\int_{\mathbb{R}^{2n}}\int_{\mathbb{R}^{2n}}f(u+v)g(v+w)\exp\Big(\frac{2i}{\hbar}\sigma[v,w]\Big)d^{2n}v\,d^{2n}w$$
$$\sim\Big\langle\sum_{|\alpha|=0}^{\infty}\Big(\frac{i\hbar}{2}\Big)^{|\alpha|}\frac{1}{\alpha!}\partial^\alpha\delta(v)\hat{\partial}^\alpha(w),f(u+v)g(u+w)\Big\rangle_{v,w}.$$

Therefore, we have established the following theorem.

Theorem 1: Let f and g be smooth functions on $\mathbb{R}^{2n}$ that satisfy one of the following conditions: (a) $f\in\mathscr{E}(\mathbb{R}^{2n})$, $g\in\mathscr{O}_{\exp}(\mathbb{R}^{2n})$; or (b) $f\in\mathscr{O}_M(\mathbb{R}^{2n})$, $g\in\mathscr{O}_C(\mathbb{R}^{2n})$. Then

$$(f\times_\hbar g)(u)=\sum_{|\alpha|=0}^{N}\Big(\frac{i\hbar}{2}\Big)^{|\alpha|}\frac{1}{\alpha!}\partial^\alpha f(u)\hat{\partial}^\alpha g(u)+O(\hbar^{N+1})\quad\text{as }\hbar\to0,\tag{13}$$

for every $u\in\mathbb{R}^{2n}$. ■

Since $g\times_\hbar f=(f^*\times_\hbar g^*)^*$, the expansion also holds for $f\in\mathscr{O}_{\exp}$, $g\in\mathscr{E}$, or for $f\in\mathscr{O}_C$, $g\in\mathscr{O}_M$. Note that the integral form (2) of the twisted product is defined in the second case in a distributional sense, via the Antonets extension, since $\mathscr{O}_C$ is contained in $\mathscr{M}^\hbar$. Note, also, that (13) in particular applies when both f and g belong to $\mathscr{K}(\mathbb{R}^{2n})$.

Clearly, pointwise expansions of the twisted product such as (13) can hold only if both factors are smooth functions. If we consider distributional asymptotic expansions, one of the factors can be a generalized function, as long as the other remains smooth. That (13) need not hold in any sense when both f and g are generalized functions is illustrated by the example (5).

Theorem 2: Let $f\in\mathscr{S}(\mathbb{R}^{2n})$ and $T\in\mathscr{S}'(\mathbb{R}^{2n})$. Then (13) holds in the distributional sense, that is,

$$\langle f\times_\hbar T,h\rangle=\sum_{|\alpha|=0}^{N}\Big(\frac{i\hbar}{2}\Big)^{|\alpha|}\frac{1}{\alpha!}\langle\partial^\alpha f\hat{\partial}^\alpha T,h\rangle+O(\hbar^{N+1})\quad\text{as }\hbar\to0,$$

for every $h\in\mathscr{S}(\mathbb{R}^{2n})$.

Proof: Let R be the distribution on $\mathbb{R}^{4n}$ given by

$$R(v,u):=\langle T(u+w),e^{i\sigma[v,w]}\rangle_w$$
$$=\langle\tau_{-u}T(w),\exp(-iw\cdot Jv)\rangle_w$$

$$= \mathscr{F}(\tau_{-u}T)(Jv)$$
$$= \exp(iu \cdot Jv)(\mathscr{F}T)(Jv),$$

where $\tau_{-u}T$ is the translate of T by u. Then

$$(R \circ h)(v) = \int_{\mathbb{R}^{2n}} \exp(iu \cdot Jv) h(u) d^{2n}u \, \mathscr{F}T(Jv)$$
$$= \mathscr{F}T(Jv)\mathscr{F}\check{h}(-Jv) = \mathscr{F}(T * \check{h})(Jv).$$

[Here $\check{h}(u) = h(-u)$.] Since convolution of a distribution in $\mathscr{S}'$ with a test function in $\mathscr{S}$ lies[23] in $\mathscr{O}_C$, we find that $R \circ h \in \mathscr{F}(\mathscr{O}_C) = \mathscr{O}'_M$. Thus the moment asymptotic expansion (12) of $R(\lambda v, u)$ holds for test functions in $\mathscr{O}_M(\mathbb{R}^{2n}) \hat{\otimes} \mathscr{S}(\mathbb{R}^{2n})$ and thus for test functions in $\mathscr{S}(\mathbb{R}^{2n}) \hat{\otimes} \mathscr{S}(\mathbb{R}^{2n}) = \mathscr{S}(\mathbb{R}^{4n})$. In particular, if we set $g(v,u) = f(u+v)h(u)$, then $g \in \mathscr{S}(\mathbb{R}^{4n})$.

The partial moments of R given by

$$\langle \mu_\alpha(u), k(u) \rangle = \langle R(v,u), v^\alpha k(u) \rangle_{u,v}$$
$$= \langle T(u+w), e^{i\sigma[v,w]} v^\alpha k(u) \rangle_{u,v,w}$$
$$= (2\pi)^{2n} \langle T(u+w), (-i)^{|\alpha|} \hat{\partial}^\alpha \delta(w) k(u) \rangle_{u,w}$$
$$= (2\pi)^{2n} i^{|\alpha|} \langle \hat{\partial}^\alpha T(u), k(u) \rangle,$$

and so

$$\langle f \times_\hbar T, h \rangle = (\pi\hbar)^{-2n} \langle R((2/\hbar)v, u), g(v,u) \rangle_{u,v}$$
$$\sim (\pi\hbar)^{-2n} \sum_{|\alpha|=0}^{\infty} \left(\frac{\hbar}{2}\right)^{|\alpha|+2n} \frac{(-1)^{|\alpha|}}{\alpha!} \langle \partial^\alpha \delta(v) \mu_\alpha(u), f(u+v)h(u) \rangle_{u,v}$$
$$\sim (2\pi)^{-2n} \sum_{|\alpha|=0}^{\infty} \left(\frac{\hbar}{2}\right)^{|\alpha|} \frac{1}{\alpha!} \langle \partial^\alpha f(u) \mu_\alpha(u), h(u) \rangle$$
$$\sim \sum_{|\alpha|=0}^{\infty} \left(\frac{i\hbar}{2}\right)^{|\alpha|} \frac{1}{\alpha!} \langle \partial^\alpha f(u) \hat{\partial}^\alpha T(u), h(u) \rangle,$$

as required. ■

Finally, let us examine the question of the convergence of the twisted product expansion. Asymptotic series may be divergent and even if convergent its sum is, in general, different from the function that was developed. A striking example in the present context is the following. If f and g both belong to $\mathscr{D}(\mathbb{R}^{2n})$ but their supports do not meet, then the twisted product $f \times_\hbar g$ extends to an entire analytic function and thus is generically nonzero on $\mathbb{R}^{2n}$; however, $f \times_\hbar g \sim 0$ to all orders as $\hbar \to 0$.

Even so, the asymptotic development of the twisted product becomes a convergent series in some cases. The basic step for obtaining convergence results is the following. Suppose $h \in \mathscr{Z}(\mathbb{R}^n)$; then, since h extends to an entire function in $\mathbb{C}^n$, its Taylor series

$$h(x) = \sum_{|\alpha|=0}^{\infty} \frac{1}{\alpha!} \partial^\alpha h(0) x^\alpha$$

converges in the topology of $\mathscr{E}(\mathbb{R}^n)$. Since $\langle T(\lambda x), x^\alpha \rangle = \lambda^{-|\alpha|-n} \mu_\alpha$, we obtain the convergent series

$$\langle T(\lambda x), h(x) \rangle = \sum_{|\alpha|=0}^{\infty} \frac{\partial^\alpha h(0) \mu_\alpha}{\alpha! \lambda^{|\alpha|+n}},$$

for every λ, if $T \in \mathscr{E}'(\mathbb{R}^n)$.

Use of this result immediately gives the following theorem.

Theorem 3: Let $f \in \mathscr{Z}(\mathbb{R}^{2n})$, $g \in \mathscr{O}_{\exp}(\mathbb{R}^{2n})$ (or vice versa). Then

$$(f \times_\hbar g)(u) = \sum_{|\alpha|=0}^{\infty} \left(\frac{i\hbar}{2}\right)^{|\alpha|} \frac{1}{\alpha!} \partial^\alpha f(u) \hat{\partial}^\alpha g(u), \qquad (14)$$

for every $\hbar > 0$, the convergence being uniform on compact subsets of $\mathbb{R}^{2n}$. ■

This theorem strengthens the previous result of Ref. 10, where the convergence was proved when both f and g belong to $\mathscr{O}_{\exp}(\mathbb{R}^{2n})$. In Ref. 12, under the same hypothesis of Ref. 10, only convergence in $\mathscr{S}'(\mathbb{R}^{2n})$ was proved.

Note that, as $\mathscr{Z} \not\subset \mathscr{S}'$, the twisted product (14) is not defined via the Antonets extension in general.

To summarize, using moment distributional asymptotics, we have developed a rigorous methodology for giving precise meaning to the "folk wisdom" concerning the semiclassical expansion of twisted products. In this context, to strive for maximum generality is perhaps pointless. The interested reader, we believe, will have no difficulty in employing variants of this methodology to get statements tailored to his particular needs.

IV. A PHYSICAL APPLICATION AND OUTLOOK

Let ρ_0 (a generalized Wigner function) represent a quantum state. We shall assume that ρ_0 belongs to $\mathscr{S}$; this is at any rate the case with the explicit examples below.[34]

Let H be the classical Hamiltonian governing the dynamics of a quantum system. We shall assume that H is real, smooth, and with bounded derivatives of order $\geqslant 2$. [In particular, if $H = p^2/2m + V(x)$, then $|\partial^\alpha V| \leqslant C_\alpha$ for $|\alpha| \geqslant 2$.] This ensures (i) the operator associated to H by the Weyl correspondence[2] is essentially self-adjoint; (ii) thus, there exists a one-parameter group $U^\hbar(t)$ of distributions in phase space corresponding to the unitary evolution group [moreover, $U^\hbar(t) \in \mathscr{M}^\hbar$ for all t[35]]; and (iii) the solution of the classical Hamilton equations associated to H yields a globally defined group of diffeomorphisms (ψ_t) of the phase space.

The "classical evolution" of ρ_0 is given by

$$\rho_c(t) = \rho_0 \circ \psi_{-t},$$

where ρ_c satisfies the differential equation

$$\frac{\partial \rho_c}{\partial t} = \{H, \rho_c\}_{\mathrm{P}}, \quad \rho_c(0) = \rho_0.$$

The quantum solution is given by

$$\rho_q^\hbar(t) = U^\hbar(t) \times_\hbar \rho_0 \times_\hbar \overline{U}^\hbar(t).$$

If we define the *Moyal bracket* $\{\cdot,\cdot\}_{\mathrm{M}}^\hbar$ by the formula

$$\{f,g\}_{\mathrm{M}}^\hbar = -(i/\hbar)(f \times_\hbar g - g \times_\hbar f),$$

then $\rho_q^\hbar$ verifies the equation

$$\frac{\partial \rho_q^\hbar}{\partial t} = \{H, \rho_q^\hbar\}_{\mathrm{M}}^\hbar, \quad \rho_q^\hbar(0) = \rho_0.$$

We can derive asymptotic formulas for the Moyal bracket from the previous results. Under the hypothesis of Theorem 2, for instance, we have

$$\{f,T\}_{\mathrm{M}}^\hbar - \{f,T\}_{\mathrm{P}} = O(\hbar^2) \tag{15}$$

in the distributional sense as $\hbar \to 0$.

Now we derive an expansion for $\rho_q^\hbar$ in terms of the classical dynamics. Introduce, for $0 \leqslant t' \leqslant t$,

$$\Phi_1^\hbar(t,t') := U^\hbar(t') \times_\hbar \rho_c(t-t') \times_\hbar \overline{U}^\hbar(t').$$

Note that $\Phi_1^\hbar(t,0) = \rho_c(t)$ and $\Phi_1^\hbar(t,t) = \rho_q^\hbar(t)$. If we denote

$$\rho_1^\hbar(y_1) := (1/\hbar^2)[\{H, \rho_c(y_1)\}_{\mathrm{M}}^\hbar - \{H, \rho_c(y_1)\}_{\mathrm{P}}],$$

then we can write

$$\rho_q^\hbar(t) = \rho_c(t) + \hbar^2 \int_0^t dt'\, U^\hbar(t') \times_\hbar \rho_1^\hbar(t-t') \times_\hbar \overline{U}^\hbar(t'),$$

Now we "classically evolve" $\rho_1^\hbar$ in turn. Let, for $0 \leqslant t'' \leqslant t'$

$$\Phi_2^\hbar(t,t',t'') := U^\hbar(t'') \times_\hbar (\rho_1(t-t') \circ \psi_{t''-t'}) \times_\hbar \overline{U}^\hbar(t'').$$

If we denote

$$\rho_2^\hbar(y_1,y_2) := (1/\hbar^2)[\{H, \rho_1^\hbar(y_1) \circ \psi_{-y_2}\}_{\mathrm{M}}^\hbar - \{H, \rho_1^\hbar(y_1) \circ \psi_{-y_2}\}_{\mathrm{P}}],$$

we then obtain

$$\rho_q^\hbar(t) = \rho_0 \circ \psi_{-t} + \hbar^2 \int_0^t dt'\, \rho_1^\hbar(t-t') \circ \psi_{-t'} + \hbar^4 \int_0^t dt' \int_0^{t'} dt''\, U^\hbar(t'') \times_\hbar \rho_2^\hbar(t-t', t'-t'') \times_\hbar \overline{U}^\hbar(t'').$$

Continuing in this vein, after k iterations we get the expression

$$\begin{aligned}
\rho_q^\hbar(t) = {} & \rho_0 \circ \psi_{-t'} + \hbar^2 \int_0^t dt'\, \rho_1^\hbar(t-t') \circ \psi_{-t'} + \hbar^4 \int_0^t dt' \int_0^{t'} dt''\, \rho_2^\hbar(t-t', t'-t'') \circ \psi_{-t''} \\
& + \cdots + \hbar^{2k} \int_0^t dt \int_0^{t'} dt'' \cdots \int_0^{t^{(k-1)}} dt^{(k)} \rho_k^\hbar(t'-t', t'-t'', \ldots, t^{(k-1)} - t^{(k)}) \circ \psi_{-t^{(k)}} \\
& + \hbar^{2k+2} \int_0^t dt' \int_0^{t'} dt'' \cdots \int_0^{t^{(k)}} dt^{(k+1)} U^\hbar(t^{(k+1)})\, {}_\hbar \rho_{k+1}^\hbar (t-t', t'-t'', \ldots, t^{(k)} - t^{(k+1)})_\hbar \overline{U}^\hbar(t^{(k+1)}),
\end{aligned} \tag{16}$$

where ρ_0 is given, and the $\rho_j^\hbar$ are defined recursively by

$$\rho_j^\hbar(y_1,\ldots,y_j) := (1/\hbar^2)[\{H, \rho_{j-1}^\hbar(y_1,\ldots,y_{j-1}) \circ \psi_{-y_k}\}_{\mathrm{M}}^\hbar - \{H, \rho_{j-1}^\hbar(y_1,\ldots,y_{j-1}) \circ \psi_{-y_k}\}_{\mathrm{P}}].$$

In principle, (16) is an asymptotic expansion for the quantum evolution of a state. However, in general, ρ_0 *will depend on* $\hbar$ in a fairly singular way, which can introduce negative powers of $\hbar$, thus spoiling the expansion; not unrelatedly, one usually finds that $\lim_{\hbar \to 0} \rho_0^\hbar$ does not lie in $\mathscr{S}$.

We have, for instance, for the translates of the fundamental state of the (one-dimensional) harmonic oscillator (with mass and frequency 1),

$$\rho_{q_0,p_0}^{\mathrm{HO},\hbar}(q,p) = (1/\pi\hbar) \exp[-[(p-p_0)^2 + (q-q_0)^2]/\hbar].$$

For the translates of the fundamental state of the (one-dimensional) Morse oscillator with the same characteristics[36,37] one has

$$\rho_{q_0,p_0}^{\mathrm{MO},\hbar}(q,p) = \frac{2}{\pi\hbar\Gamma(2/(\alpha^2\hbar) - 1)} \times \left(\frac{2}{\alpha^2\hbar} e^{-\alpha(q-q_0)}\right)^{2/(\alpha^2\hbar)-1} \times K_{2i(p-p_0)/\alpha\hbar}\left(\frac{2}{\alpha^2\hbar} e^{-\alpha(q-q_0)}\right),$$

where α is a nonzero real parameter. Although $\rho_{q_0,p_0}^{\mathrm{HO},\hbar}$ and $\rho_{q_0,p_0}^{\mathrm{MO},\hbar}$ both lie in $\mathscr{S}$ for any $\hbar > 0$, they both tend to $\delta(q-q_0)\delta(p-p_0)$ as $\hbar \to 0$, and this is not a Schwartz function. In general, we expect classical states to be elements only of $\mathscr{O}'_M$.

Under these conditions, in order to get a rigorous meaning for the classical limit, one would have to impose strong and unnatural restrictions on the Hamiltonian,[9] which we do not wish to do. This does not gainsay the fact that the very first correction term of (16) can sometimes give very accurate approximations.[36]

These difficulties seem to vanish when we consider thermally averaged states. The Gibbs state for a collection of harmonic oscillators in phase space at temperature T is given by

$$\rho^{\mathrm{HO},T,\hbar}(q,p):=\frac{1}{\pi\hbar}\tanh\Big(\frac{\hbar}{2kT}\Big)\times\exp\Big[-\frac{q^2+p^2}{\hbar}\tanh\Big(\frac{\hbar}{2kT}\Big)\Big]. \tag{17}$$

Here k is Boltzmann's constant. Note that this state is classical (i.e., non-negative) for all T and that as $\hbar\to 0$,

$$\rho^{\mathrm{HO},T,\hbar}(q,p)\to(2\pi kT)^{-1}\exp[-\tfrac{1}{2}(q^2+p^2)/kT],$$

which lies in $\mathscr{S}$.

If we take $\rho^{\mathrm{HO},T,\hbar}$ as the initial state ρ_0, then the asymptotic development (16) makes perfect sense.

Indeed, we expect classical limits of quantum states that retain smoothness to be elements of the Schwartz space $\mathscr{S}$, in view of the following result (which may be thought of as a kind of Tauberian theorem).

Theorem 4: $\mathscr{O}_M\cap\mathscr{O}'_M=\mathscr{S}$ and $\mathscr{O}_C\cap\mathscr{O}'_C=\mathscr{S}$.

This important mathematical result was established only recently by Ortner and Wagner,[38] using the theory of the distribution spaces $\mathscr{D}'_{L^p}$. We give in the Appendix a different but quite elementary proof.

The interesting point raised by some recent investigations[39,40] is that there exist *large families* of quantum (mixed, of course) states that are also classical states and not necessarily Gaussian in general. It is tempting to interpret them, in keeping with Wigner's seminal intent[41]—to wit, the calculation of quantum corrections to the classical distribution functions in thermodynamic equilibrium—as thermally averaged states of some sort [the machine that two of us built in Ref. 39 to produce such quantum-classical states is able to turn out states of the form (17)]. It is also apparent that the introduction of mixed states gives a new dimension to the subject of semiclassical mechanics, as the introduction of incoherent superpositions of states tends to wash out oscillatory behavior; this point of view is reinforced by examples in finite-dimensional quantum mechanics (see Sec. 5.6 of Ref. 42). We have conjectured that in that context (16) constitutes a valid expansion; the question is currently under examination.

APPENDIX: PROOF OF THEOREM 4

For the proof of Theorem 4, we first remark that, on account of the Fourier invariance of $\mathscr{S}$, it is enough to establish that $\mathscr{O}_C\cap\mathscr{O}'_C=\mathscr{S}$.

Lemma 1: Let $f\in\mathscr{O}_C\cap\mathscr{O}'_C$, with $f\geqslant 0$. Then

$$\int_{\mathbb{R}^n}|x|^r f(x)\,d^n x<\infty,\quad\text{for all } r\geqslant 1.$$

Proof: Let $h\in\mathscr{D}(\mathbb{R})$ be a compactly supported test function satisfying $h(t)=1$ for $|t|\leqslant 1$, and $h'(t)\leqslant 0$ for $t\geqslant 0$. Define $h_m\in\mathscr{D}(\mathbb{R}^n)$ by $h_m(x):=h(|x|/m)$. Then $h_m\to 1$ in $\mathscr{O}_C$ as $m\to\infty$. Thus

$$\langle f,1\rangle=\lim_{m\to\infty}\langle f,h_m\rangle=\lim_{m\to\infty}\int_{\mathbb{R}^n}f(x)h_m(x)\,d^n x=\int_{\mathbb{R}^n}f(x)\,d^n x$$

using the monotone covergence theorem. Hence $f\in L^1(\mathbb{R}^n)$. If we now replace $f(x)$ by $|x|^{2m}f(x)$, which also lies in $\mathscr{O}_C\cap\mathscr{O}'_C$, we find that $|x|^{2m}f\in L^1(\mathbb{R}^n)$, for all $m=1,2,\ldots$, and the result follows. ■

Lemma 2: Let $f\in C^1(\mathbb{R}^n)$ with $f\geqslant 0$, and suppose there are constants $K>0$, $M>0$ such that

$$\Big|\frac{\partial f}{\partial x_i}(x)\Big|\leqslant K|x|^M,\quad\text{for } |x|\geqslant 1.$$

Let $m>Mn$. Then for each $\epsilon>0$ and each $C>0$, we can find $A>0$ so that if $|a|>A$ and $f(a)\geqslant 2\epsilon$, then

$$\int_{|x-a|<|a|^{-2M}}|x|^{2m}f(x)\,d^n x\geqslant C.$$

Proof: If $|x-a|<|a|^{-2M}$, choose $y=(1-\theta)a+\theta x$ with $0\leqslant\theta\leqslant 1$ so that $f(x)-f(a)=\nabla f(y)\cdot(x-a)$. Then

$$f(x)=f(a)+\nabla f(y)\cdot(x-a)\geqslant 2\epsilon-K|y|^M|a|^{-2M}\geqslant 2\epsilon-K(|a|+|a|^{-2M})^M|a|^{-2M}.$$

Now

$$(|a|+|a|^{-2M})^M|a|^{-2M}\to 0\quad\text{as } |a|\to\infty;$$

so

$$(|a|+|a|^{-2M})^M|a|^{-2M}<\epsilon/K$$

if $|a|>A_1$ for A_1 large enough. Thus $f(x)\geqslant\epsilon$ whenever $|a|\geqslant A_1$, $f(a)\geqslant 2\epsilon$ and $|x-a|<|a|^{-2M}$. Now

$$\int_{|x-a|<|a|^{-2M}}|x|^{2m}f(x)\,d^n x\geqslant\epsilon(|a|-|a|^{-2M})^{2m}|a|^{-2Mn}B_n,$$

where B_n is the volume of the unit ball in $\mathbb{R}^n$; the right-hand side exceeds C for $|a|\geqslant A$ with A large enough, since

$$(|a|-|a|^{-2M})^{2m}|a|^{-2Mn}\to\infty\ \text{as } |a|\to\infty. \qquad ■$$

Proof of Theorem 4: Let $g\in\mathscr{O}_C\cap\mathscr{O}'_C$. Then if $\alpha\in\mathbb{N}^n$, $m\in\mathbb{N}$, the function $f(x):=|x|^{2m}|\partial^\alpha g(x)|^2$ also belongs to $\mathscr{O}_C\cap\mathscr{O}'_C$ and is non-negative. Let $\epsilon>0$ and use Lemma 2 with

$$C:=1+\int_{\mathbb{R}^n}|x|^{2m}f(x)\,d^n x$$

to obtain the existence of $A>0$ such that if $|a|>A$, then $f(a)<2\epsilon$. We conclude that $\lim_{x\to\infty}f(x)=0$ and consequently $\lim_{x\to\infty}|x|^m\,\partial^\alpha g(x)=0$. ■

[1] J. E. Moyal, Proc. Cambridge Philos. Soc. **45**, 99 (1949).
[2] H. Weyl, *Gruppentheorie und Quantenmechanik* (Hirzel, Leipzig, 1928).
[3] H. J. Groenewold, Physica **12**, 405 (1946).
[4] F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer, Ann. Phys. (NY) **111**, 61, 111 (1978).
[5] R. Howe, J. Funct. Anal. **38**, 188 (1980).
[6] J. C. T. Pool, J. Math. Phys. **7**, 66 (1966).
[7] J. Baker, Phys. Rev. **109**, 2198 (1958).
[8] J. J. Sławianowski, in *The Uncertainty Principle and Foundations of Quantum Mechanics* (Wiley, New York, 1977), p. 147.

[9]M. A. Antonets, Lett. Math. Phys. **2**, 241 (1978); Teor. Mat. Fiz. **38**, 331 (1979).
[10]J. M. Gracia-Bondía and J. C. Várilly, J. Math. Phys. **29**, 869 (1988).
[11]A. Voros, J. Funct. Anal. **29**, 104 (1978).
[12]A. Grossmann, G. Loupias, and E. M. Stein, Ann. Inst. Fourier (Grenoble) **18**, 343 (1968).
[13]F. J. Narcowich, J. Math. Phys. **28**, 2873 (1987).
[14]D. Arnal and J. C. Cortet, J. Geom. Phys. **3** (1986).
[15]J.-B. Kammerer, C. R. Acad. Sci. Paris **298**, 59 (1984).
[16]J.-B. Kammerer, J. Math. Phys. **27**, 529 (1986).
[17]B. Grammaticos and A. Voros, Ann. Phys. (NY) **123**, 359 (1979).
[18]P. Carruthers and F. Zachariasen, Rev. Mod. Phys. **55**, 245 (1983).
[19]L. Hörmander, *The Analysis of Partial Differential Operators I* (Springer, Berlin, 1983).
[20]R. Estrada and R. P. Kanwal, Complex Variables **9**, 31 (1987).
[21]R. Estrada and R. P. Kanwal, *A Distributional Theory for Asymptotic Expansions*, to appear in Proc. Roy. Soc. London A, 1990.
[22]R. P. Kanwal, *Generalized Functions: Theory and Technique* (Academic, New York, 1983).
[23]J. Horváth, *Topological Vector Spaces and Distributions I* (Addison–Wesley, Reading, MA, 1966).
[24]H. Bremermann, *Distributions, Complex Variables, and Fourier Transforms* (Addison–Wesley, Reading, MA, 1965).
[25]L. Schwartz, *Théorie des Distributions* (Hermann, Paris, 1966).
[26]A. Grothendieck, *Produits tensoriels topologiques et espaces nucléaires* (Am. Math. Soc., Providence, RI, 1955).
[27]This example was pointed out to us by P. Wagner.
[28]J. C. Várilly, E. de Faría, and J. M. Gracia-Bondía, Cienc. Tec. (C.R.) **10**, 81 (1986).
[29]G. Lassner and G. A. Lassner, "Qu*-algebras and twisted product," BiBoS preprint, Bielefeld, 1987.
[30]H. Figueroa, "Function algebras under the twisted product", Boletim da Sociedade Paranaense de Matemática, 1989.
[31]R. A. Handelsman and N. Bleistein, SIAM J. Math. Anal. **4**, 519 (1973).
[32]P. Durbin, J. Inst. Math. Appl. **23**, 181 (1979).
[33]D. S. Jones, *Generalized Functions* (Cambridge U.P., Cambridge, 1982).
[34]If we suppose that $\rho_0 \in (\mathscr{M}_L^\hbar)'$ and $\rho_0 \in (\mathscr{M}_R^\hbar)'$, then necessarily $\rho_0 \in \mathscr{S}$; but this is not strictly necessary for having a quantum state. In any case, elements of $(\mathscr{M}_L^\hbar)'$ or $(\mathscr{M}_R^\hbar)'$ are necessarily smooth.
[35]D. Robert, *Autour de l'approximation semiclassique* (Birkhäuser, Boston, 1987).
[36]H. W. Lee and M. O. Scully, J. Chem. Phys. **77**, 4604 (1982).
[37]J. P. Dahl and M. Springborg, J. Chem. Phys. **88**, 4535 (1988).
[38]N. Ortner and P. Wagner, "Applications of weighted $\mathscr{D}'_{L^p}$ spaces to the convolution of distributions," Bull. Polish Acad. Sci. Math. **37** (1989).
[39]J. M. Gracia-Bondía and J. C. Várilly, Phys. Lett. A **128**, 20 (1988).
[40]F. J. Narcowich and R. F. O'Connell, Phys. Lett. A **133**, 167 (1988).
[41]E. P. Wigner, Phys. Rev. **40**, 749 (1932).
[42]J. C. Várilly and J. M. Gracia-Bondía, Ann. Phys. (NY) **190**, 107 (1989).

Angular reduction in multiparticle matrix elements

D. R. Lehman and W. C. Parke
Physics Department, The George Washington University, Washington, DC 20052

(Received 19 April 1989; accepted for publication 19 July 1989)

A general method for reduction of coupled spherical harmonic products is presented. When the total angular coupling is zero, the reduction leads to an explicitly real expression in the scalar products of the unit vector arguments of the spherical harmonics. For nonscalar couplings, the reduction gives Cartesian tensor forms for the spherical harmonic products; tensors built from the physical vectors in the original expression. The reduction for arbitrary couplings is given in closed form, making it amenable to symbolic manipulation on a computer. The final expressions do not depend on a special choice of coordinate axes, nor do they contain azimuthal quantum number summations, or do they have complex tensor terms for couplings to a scalar; consequently, they are easily interpretable from the properties of the physical vectors they contain.

I. INTRODUCTION

A common occurrence in quantum mechanical calculations for multiparticle systems is the product of several spherical harmonics coming from the operators and eigenstates of particle or cluster wave functions. For example, in three-body models of ^{6}Li, the quadrupole form factor begins with up to five spherical harmonics coupled to zero total angular momentum, each with a different argument (two each in the initial and final states, one in the quadrupole operator). There are a variety of ways to evaluate transition amplitudes and expectation values involving these products. This paper will present an alternative that can be applied to arbitrary tensor couplings. When those tensors are built from physical vectors in the problem, the method leads to scalar couplings expressed as polynomials of the scalar products of those vectors.

Several methods for handling a series of spherical harmonic couplings have been suggested in the literature. One technique applies when only three are coupled, and makes use of the freedom of choice for the orientation of the spatial axis system. One of the spherical harmonic argument vectors is aligned with the azimuthal quantization z axis, and another pair defines the xz plane. (See, for example, the paper by Balian and Brezin.[1]) Putting coplanar vectors all in the xy plane also simplifies the explicit form of the spherical harmonics. In either case, a sum over azimuthal quantum numbers remains for scalar expressions. An extension of the above method takes advantage of the three-dimensional character of the underlying space. The vector argument within any spherical harmonic in a product expression is written in terms of any three independent vectors in the problem. Those spherical harmonics with an argument direction determined by a pair of vectors can be expanded as a product of spherical harmonics in each of these vectors.[2] Spherical harmonics with the same argument are then combined. The result of the reduction will be a sum over products of no more than three spherical harmonics in three different solid angles. The technique described above can then be applied to these remaining spherical harmonics.

For the case of a pair of coupled spherical harmonics with angular arguments determined by two different unit vectors $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$, each spherical harmonic with high angular indices (l_1 and l_2) coupled to a total angular momentum of low angular index (such as $L = 1, 2,$ or 3), it is possible[3] to express the coupled pair in terms of a basis set of pair-coupled spherical harmonics each with minimal angular index, times Legendre functions of argument $\hat{\mathbf{a}}\cdot\hat{\mathbf{b}}$. Such results will turn out to be special cases of the method given in the following.

In this paper, we wish to present a general method for the reduction of products of spherical harmonics which we have been using for some years.[4] When the total angular coupling is zero, the reduction leads to an explicitly real expression in the scalar dot products of the vector arguments of the original spherical harmonics. For nonscalar couplings, the reduction gives Cartesian tensor forms for the spherical harmonic products; tensors built from the physical vectors in the original problem. The advantages of the method are the following: (1) The result is readily interpretable from the known properties of the physical vectors it contains. (2) No special choice of coordinate axes are needed. (3) The final expression contains no azimuthal quantum number summations and no complex terms for couplings to a scalar. (4) The reduction for arbitrary couplings can be given in closed form, making it easily programmable in a computer calculation. As there are no spherical harmonic origin-shift expansions, numerical convergence problems associated with this reexpansion are avoided. Section II introduces how the reduction of the scalar couplings of spherical harmonics can lead to simple results in terms of the corresponding vector dot-product expression. In Sec. III, we set up a method for transforming between Cartesian and spherical tensors. Section IV gives the general results for expanding the coupling of Cartesian tensors into an irreducible tensor sum. A by-product of this work is a general formula for the Cartesian Clebsch–Gordan coefficients. Section V shows how the Cartesian coupling can reduce arbitrarily coupled spherical harmonics with different arguments, using a few simple rules. Finally, Sec. VI gives our conclusions.

II. EXAMPLES OF SCALAR COUPLING REDUCTIONS

As a way of introducing the general scheme for Cartesian recoupling, consider the following expression:

$$[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times[\mathbf{Y}^{[1]}(\hat{\mathbf{c}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{d}})]^{[2]}]^{[0]}. \quad (1)$$

We use here the angular coupling notation of Fano and Racah,[5] i.e.,

$$[A^{[l_1]}\times B^{[l_2]}]^{[l_3]}_{m_3} = \sum_{m_1,m_2}\langle l_1m_1l_2m_2|l_3m_3\rangle A^{[l_1]}_{m_1}B^{[l_2]}_{m_2}. \quad (2)$$

The phases for the contrastandard $\mathbf{Y}^{[l]}_m$ spherical harmonics are fixed by

$$\mathbf{Y}^{[l]}_m = (-i)^l\mathbf{Y}_{lm} \quad (3)$$

which insures that the $\mathbf{Y}^{[l]}_m$ behave as the eigenstates of L^2 and L_z under conjugation and time reversal according to[6]

$$\psi^{[j]*}_m = (-1)^{j+m}\psi^{[j]}_{-m}. \quad (4)$$

This phase choice also has the advantage of eliminating explicit phase factors in matrix element angular recoupling algebra.[6,7]

It is well known that the $\mathbf{Y}_{lm}(\hat{\mathbf{a}})$ can be constructed from the irreducible Cartesian tensors of rank l built from the unit vector $\hat{\mathbf{a}}$. For example,

$$\begin{aligned}\mathbf{Y}^{[1]}_0(\hat{\mathbf{a}}) &= +(-i)(\hat{1}/\sqrt{4\pi})a_3,\\ \mathbf{Y}^{[1]}_{\pm1}(\hat{\mathbf{a}}) &= \mp(-i)(\hat{1}/\sqrt{4\pi})(1/\sqrt{2})(a_1\pm ia_2),\end{aligned} \quad (5)$$

and

$$\begin{aligned}\mathbf{Y}^{[2]}_0(\hat{\mathbf{a}}) &= +(-i)^2(\hat{2}/\sqrt{4\pi})a_{33},\\ \mathbf{Y}^{[2]}_{\pm1}(\hat{\mathbf{a}}) &= \mp(-i)^2(\hat{2}/\sqrt{4\pi})\sqrt{\tfrac{2}{3}}(a_{13}\pm ia_{23}),\\ \mathbf{Y}^{[2]}_{\pm2}(\hat{\mathbf{a}}) &= +(-i)^2(\hat{2}/\sqrt{4\pi})\sqrt{\tfrac{1}{6}}(a_{11}-a_{22}\pm 2ia_{12})\end{aligned} \quad (6)$$

where $\hat{l}$ is used for $\sqrt{2l+1}$, and

$$a_{ij}\equiv\tfrac{3}{2}(a_ia_j-(\tfrac{1}{3})\delta_{ij}). \quad (7)$$

The irreducible Cartesian tensors such as a_{ij} constructed from direct products of a unit vector with itself are normalized to make contraction with that vector give the corresponding next lower rank tensor. The expressions above for the spherical harmonics can be verified using

$$\hat{\mathbf{a}} = \sin\theta\cos\phi\,\hat{\mathbf{e}}_x+\sin\theta\sin\phi\,\hat{\mathbf{e}}_y+\cos\theta\hat{\mathbf{e}}_z \quad (8)$$

and then comparing to standard forms for the spherical harmonics[8] written in terms of the spherical angles (θ,ϕ).

Similarly, the angular components of $[\mathbf{Y}^{[1]}(\hat{\mathbf{c}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{d}})]^{[2]}$ can be expressed in terms of the components of the irreducible Cartesian tensor

$$Q(c,d)_{ij}\equiv\tfrac{3}{4}(c_id_j+c_jd_i-(\tfrac{2}{3})(\mathbf{c}\cdot\mathbf{d})\delta_{ij}). \quad (9)$$

The identity

$$[\mathbf{Y}^{[1]}(\hat{\mathbf{c}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{c}})]^{[2]}_m = \frac{\hat{1}\cdot\hat{1}}{\sqrt{4\pi}}\left|\begin{pmatrix}1&1&2\\0&0&0\end{pmatrix}\right|\mathbf{Y}^{[2]}_m(\hat{\mathbf{c}}) \quad (10)$$

is employed to determine the constant factor in the result

$$[\mathbf{Y}^{[1]}(\hat{\mathbf{c}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{d}})]^{[2]}_m = \sqrt{(2\cdot3)/(5\cdot4\pi)}\begin{cases}+(-i)^2\dfrac{\hat{2}}{\sqrt{4\pi}}Q(c,d)_{33} & (m=0)\\ \mp(-i)^2\dfrac{\hat{2}}{\sqrt{4\pi}}\sqrt{\dfrac{2}{3}}(Q(c,d)_{13}\pm iQ(c,d)_{23}) & (m=\pm1)\\ +(-i)^2\dfrac{\hat{2}}{\sqrt{4\pi}}\dfrac{1}{\sqrt{6}}(Q(c,d)_{11}-Q(c,d)_{22}\pm 2iQ(c,d)_{12}) & (m=\pm2).\end{cases} \quad (11)$$

Using the traceless nature of a_{ij} and $Q(c,d)_{ij}$ in the form

$$a_{11}Q(c,d)_{22}+a_{22}Q(c,d)_{11} = a_{33}Q(c,d)_{33}-a_{11}Q(c,d)_{11}-a_{22}Q(c,d)_{22},$$

one finds

$$[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times[\mathbf{Y}^{[1]}(\hat{\mathbf{c}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{d}})]^{[2]}]^{[0]} = \frac{\hat{2}}{4\pi}\cdot\frac{2}{3}\cdot\sqrt{\frac{2\cdot3}{5\cdot4\pi}}\sum_{i,j}a_{ij}Q(c,d)_{ij}. \quad (12)$$

The last factor, $\Sigma a_{ij}Q(c,d)_{ij}$, is just

$$\tfrac{3}{2}\hat{\mathbf{a}}\cdot\mathbf{Q}(c,d)\cdot\hat{\mathbf{a}} = (\tfrac{3}{2})^2\{(\mathbf{a}\cdot\mathbf{c})(\mathbf{a}\cdot\mathbf{d})-(\tfrac{1}{3})(\mathbf{c}\cdot\mathbf{d})\}. \quad (13)$$

Aligning vector directions to help find the connection between the spherical harmonic recoupling and the corresponding contracted Cartesian tensor products will not work if the couplings have odd parity, such as in the expression

$$[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[1]}\times[\mathbf{Y}^{[2]}(\hat{\mathbf{c}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{d}})]^{[1][0]}].$$

Two of the couplings above produce an axial vector from the direct product of two tensors of rank two. If we define the pseudovector

$$R(a,b)_i\equiv\frac{4}{9}\sum_{jkl}\epsilon_{ijk}a_{jl}b_{kl} = \mathbf{a}\cdot\mathbf{b}(\mathbf{a}\times\mathbf{b})_i \quad (14)$$

(ϵ_{ijk} is the completely antisymmetric tensor in three dimensions with $\epsilon_{123}=1$), then

$$[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[1]}_m = (1/\sqrt{4\pi})\cdot\sqrt{(3\cdot5)/2}\begin{cases}(-i)\sqrt{3/4\pi}R(a,b)_3 & (m=0),\\ \mp(-i)\sqrt{3/4\pi}\,(1/\sqrt{2})(R(a,b)_1\pm iR(a,b)_2) & (m=\pm1)\cdot\end{cases} \quad (15)$$

In this odd parity case, the coefficient in the first line of the expression can be determined by the explicit Clebsch–Gordan recoupling of the spherical harmonics with total azimuthal quantum number $m = 0$. We now write the double pair coupling to zero as

$$[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[1]}\times[\mathbf{Y}^{[2]}(\hat{\mathbf{c}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{d}})]^{[1]}]^{[0]}$$
$$=\left(\frac{\sqrt{3}}{4\pi}\right)\left(\frac{1}{\sqrt{4\pi}}\sqrt{\frac{(3\cdot 5)}{2}}\right)^2$$
$$\times\sum_i R(a,b)_i R(c,d)_i. \tag{16}$$

The summed factor above becomes $(\mathbf{a}\cdot\mathbf{b})(\mathbf{c}\cdot\mathbf{d})(\mathbf{a}\times\mathbf{b}\cdot\mathbf{c}\times\mathbf{d})$.

In Sec. V, we show that expressions such as Eqs. (12) and (16) can be written by inspection for arbitrary couplings.

III. GENERAL TRANSFORMATION BETWEEN IRREDUCIBLE SPHERICAL AND CARTESIAN TENSORS

In this section, we will find a covariant connection between spherical and Cartesian tensor components of arbitrary rank. This will lead to the generalization of the results of Sec. II to arbitrary couplings of spherical harmonics.

To establish our notation, we first review the connection between the generators of rotations and angular momentum. An orthonormal basis $\hat{\mathbf{e}}_i$ in Euclidean three-space can be defined through the infinitesimal displacements in that space by

$$\mathbf{dr} = \sum_i dx_i\,\hat{\mathbf{e}}_i. \tag{17}$$

In a coordinate transformed frame, they become

$$\hat{\mathbf{e}}'_i = \sum_j \frac{dx_j}{dx'_i}\hat{\mathbf{e}}_j. \tag{18}$$

The condition

$$\sum_{i,j}\frac{dx_k}{dx'_i}\frac{dx_l}{dx'_j}\delta_{ij} = \delta_{kl} \tag{19}$$

makes the transformation a rotation. For infinitesimal orthogonal transformations, Eqs. (18) and (19) give

$$\hat{\mathbf{e}}'_i = \sum_j\left(\delta_{ij} + \sum_k \epsilon_{ijk}n_k\,\delta\theta\right)\hat{\mathbf{e}}_j, \tag{20}$$

where $\hat{\mathbf{n}}$ is a unit vector along the axis of rotation in the right-hand sense and $\delta\theta$ is the rotation angle. Taking $\mathscr{R}$ to be an element of the rotation group, an infinitesimal rotation can be represented by

$$\mathscr{R} = \left(\mathscr{I} + i\sum_k S_k n_k\,\delta\theta\right). \tag{21}$$

Comparing with Eq. (20), we can read the generator for infinitesimal rotations of the Cartesian basis vectors to be the well known result:

$$(S_k)_{ij} = -i\epsilon_{ijk}. \tag{22}$$

Apart from Planck's constant, these are a representation for the angular momentum operators for a spin-one field in quantum theory. However, the z component of the angular momentum operator is usually taken as diagonal with elements being the possible measured values of this S_z. The matrix S_z is diagonalized by the unitary matrix

$$(\mathscr{U}_{mi}) = -i\begin{bmatrix} 1/\sqrt{2} & -i/\sqrt{2} & 0\\ 0 & 0 & 1\\ -1/\sqrt{2} & -i/\sqrt{2} & 0\end{bmatrix}, \tag{23}$$

giving

$$\mathscr{U}S_z\mathscr{U}^{-1} = \begin{bmatrix}1 & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & -1\end{bmatrix}. \tag{24}$$

The vector basis set in this contrastandard spherical representation is that given by Danos[6] (note that this differs from Fano and Racah[9]):

$$\hat{\mathbf{e}}^{[1]}_m = \sum_i \mathscr{U}_{mi}\hat{\mathbf{e}}_i,$$
$$\hat{\mathbf{e}}^{[1]}_{\pm 1} = \pm\frac{i}{\sqrt{2}}(\mathbf{e}_x \pm i\mathbf{e}_y),\quad \hat{\mathbf{e}}^{[1]}_0 = -i\hat{\mathbf{e}}_z. \tag{25}$$

Furthermore,

$$\hat{\mathbf{e}}^{[1]\dagger}_m\cdot\hat{\mathbf{e}}^{[1]}_n = \delta_{mn},\quad \sum_m \hat{\mathbf{e}}^{[1]}_m\,\hat{\mathbf{e}}^{[1]}_m{}^\dagger = \mathbb{I}, \tag{26}$$

where $\mathbb{I}$ is the unit dyadic operator.

The arbitrary phase in the unitary transformation has been taken to make the spherical basis vectors conform with the conjugation property of angular momentum eigenstates given in Eq. (4). A contrastandard spherical tensor carries a superscripted square bracket enclosing its rank index. Higher-weight spherical tensors irreducible under the rotation group can be constructed from angular couplings of the vector basis set:

$$\hat{\mathbf{e}}^{[l]}_m = [\hat{\mathbf{e}}^{[1]}\times\hat{\mathbf{e}}^{[1]}\times\hat{\mathbf{e}}^{[1]}\times\cdots(l)\cdots\times\hat{\mathbf{e}}^{[1]}]^{[l]}_m. \tag{27}$$

Individual pairwise couplings on the right-hand side of Eq. (27) taken in any order give the same result. This fact comes from the "stretched" form of the tensor, i.e., it has the highest rank which can be constructed from l vectors of rank 1. Explicitly, the Clebsch–Gordan products in Eq. (27) give

$$\hat{\mathbf{e}}^{[l]}_m = \left[\frac{(l-m)!(l+m)!}{l!(2l-1)!!}\right]^{1/2}\sum_{\substack{m\text{'s from}\\ -1\text{ to }1}}\left[\frac{1}{(1-m_1)!(1+m_1)!\cdots(1-m_l)!(1+m_l)!}\right]^{1/2}\hat{\mathbf{e}}^{[1]}_{m_1}\cdots\hat{\mathbf{e}}^{[1]}_{m_l}. \tag{28}$$

The summation expression in Eq. (28) implicitly depends on m, since the coupled terms on the right-hand side of Eq. (27) must have their azimuthal quantum numbers add to m.

The rank-l tensors $\hat{\mathbf{e}}^{[l]}_m$ satisfy

$$\hat{\mathbf{e}}^{[l]\dagger}_m\cdot\hat{\mathbf{e}}^{[l]}_n = \delta_{mn} \tag{29}$$

and

$$\sum_m \hat{\mathbf{e}}_m^{[l]}\hat{\mathbf{e}}_m^{[l]\dagger} = \mathscr{P}^{[l]}, \tag{30}$$

where $\mathscr{P}^{[l]}$ is a projection operator on rank-l Cartesian tensors which picks out only the irreducible part. The dot products which appear between higher rank tensors imply contraction over all Cartesian tensor indices.

A Cartesian tensor irreducible under the rotation group and of rank-l must be both completely symmetric in its l indices and traceless. Suppose $T_{i_1\cdots i_l}$ is such a tensor. Then a natural connection between this Cartesian tensor and its spherical representation is given by the scalar expression:

$$\mathbb{T} = \sum_{i\text{'s}} T^{\{l\}}_{i_1 i_2\cdots i_l}\hat{\mathbf{e}}_{i_1}\hat{\mathbf{e}}_{i_2}\cdots\hat{\mathbf{e}}_{i_l}$$

$$= \sum_m T_m^{[l]}\hat{\mathbf{e}}_m^{[l]\dagger} = \hat{l}\,[T^{[l]}\times\hat{\mathbf{e}}^{[l]}]^{[0]}. \tag{31}$$

Contrastandard *Cartesian* tensors will be denoted by putting their rank index in curly brackets. With Eq. (31), the transformation coefficients between Cartesian and spherical tensors become

$$\mathscr{U}^{[l]}_{m_{i_1 i_2\cdots i_l}} = \hat{\mathbf{e}}_m^{[l]}\cdot\hat{\mathbf{e}}_{i_l}\cdots\hat{\mathbf{e}}_{i_1}. \tag{32}$$

Thus,

$$T^{\{l\}}_{i_1 i_2\cdots i_l} = \hat{l}\,[T^{[l]}\times\mathscr{U}^{[l]}_{i_1 i_2\cdots i_l}]^{[0]} \tag{33}$$

and

$$T_m^{[l]} = \sum_{i\text{'s}} T^{\{l\}}_{i_1 i_2\cdots i_l}\mathscr{U}^{[l]}_{m_{i_1 i_2\cdots i_l}}. \tag{34}$$

The transformation coefficients satisfy the orthonormality conditions

$$\sum_{i\text{'s}} \mathscr{U}^{[l]\dagger}_{m_{i_1 i_2\cdots i_l}}\mathscr{U}^{[l]}_{n_{i_1 i_2\cdots i_l}} = \delta_{mn}. \tag{35}$$

With

$$\mathscr{U}^{[1]}_{m_3} = (-i)\delta_{m0}, \tag{36}$$

we find from Eq. (28),

$$\mathscr{U}^{[l]}_{m_{3\cdots 3}} = (-i)^l\left[\frac{l!}{(2l-1)!!}\right]^{1/2}\delta_{m0}. \tag{37}$$

The coefficients $\mathscr{U}^{[l]}_{m_{i_1 i_2\cdots i_l}}$ are completely symmetric and traceless in the Cartesian indices i_1 to i_l. Thus, they are irreducible in the space of both their spherical and Cartesian indices.

We use Eq. (37) to set the scale for normalization of Cartesian tensor components relative to spherical ones:

$$T^{\{l\}}_{3\cdots 3} = i^l\left[\frac{l!}{(2l-1)!!}\right]^{1/2} T_0^{[l]}. \tag{38}$$

IV. CARTESIAN TENSOR RECOUPLING

A symmetric and traceless tensor of rank l can be constructed from a unit vector $\hat{\mathbf{a}}$ in the form:

$$a^{\{l\}} = \left[\frac{(2l-1)!!}{l!}\right]\sum_{r=0}^{[l/2]}(-1)^r\left[\frac{(2l-2r-1)!!}{(2l-1)!!}\right]$$

$$\times\{a\cdots(l-2r)\cdots a\,\delta\cdots(r)\cdots\delta\}. \tag{39}$$

These are the Cartesian equivalents of the spherical harmonics, which we will refer to as the Cartesian harmonic tensors. In this expression, as in Eq. (27), the parenthetical value between continuation dots shows the number of repetitions of the factor shown before and after the dots. The l-Cartesian indices have been suppressed on $a^{\{l\}}$ and in each term of the summation. The δ's above are double indexed Kronecker deltas. The curly brackets in the right-hand side of Eq. (39) direct that the terms inside are to be summed over all permutations of the unsymmetrized indices. For a given summation index r, there will be $[l!/((l-2r)!2^r r!)]$ such terms in the symmetrization bracket. Our choice for normalization of $a^{\{l\}}$ leads to (all vectors, even when not marked with a caret, are unit vectors)

$$\mathbf{a}^{\{l\}}\cdot\hat{\mathbf{b}}\cdots(l-\text{contractions})\cdots\;\hat{\mathbf{b}} = P_l(\hat{\mathbf{a}}\cdot\hat{\mathbf{b}}), \tag{40}$$

where $\hat{\mathbf{b}}$ is a second unit vector and $\mathbf{P}_l(\hat{\mathbf{a}}\cdot\hat{\mathbf{b}})$ is the Legendre polynomial. As an example, Eq. (39) for $l=3$ becomes

$$a^{\{3\}}_{i_1 i_2 i_3} = \tfrac{5}{2}a_{i_1}a_{i_2}a_{i_3} - \tfrac{1}{2}(a_{i_1}\delta_{i_2 i_3} + a_{i_2}\delta_{i_3 i_1} + a_{i_3}\delta_{i_1 i_2}). \tag{41}$$

We will represent Cartesian harmonic tensors constructed from a unit vector ($\hat{\mathbf{a}}$) by the corresponding lower case latin letter (a). By using Eq. (38) and

$$\mathbf{Y}_0^{[l]}(\hat{\mathbf{a}}) = ((-i)^l/\sqrt{4\pi})\hat{l}P_l(\hat{\mathbf{a}}\cdot\hat{\mathbf{e}}_3) \tag{42}$$

it follows that the irreducible Cartesian tensors defined by Eq. (39) are related to the Cartesian transformed spherical harmonics by

$$\mathbf{Y}^{\{l\}}(\hat{\mathbf{a}}) = \frac{\hat{l}}{\sqrt{4\pi}}\left[\frac{l!}{(2l-1)!!}\right]^{1/2} a^{\{l\}}. \tag{43}$$

Now consider the coupling of two irreducible tensors of rank l_1 and l_2. The result can be decomposed into a sum of irreducible tensors from rank l_1+l_2 to $|l_1-l_2|$. This summation is well known in the case of spherical tensors, giving a Clebsch–Gordan series. The irreducible Cartesian tensors following from this decomposition must again be completely symmetric and traceless. By explicitly constructing symmetric and traceless tensors from the products of two irreducible tensors $A^{\{l_1\}}$ and $B^{\{l_2\}}$, it is straightforward to show that the general form for the irreducible rank l_3 tensor is given by

$$[\mathbf{A}^{\{l_1\}}\times\mathbf{B}^{\{l_2\}}]^{\{l_3\}} = C_{l_1 l_2 l_3}\left[\frac{((l_1-l_2+l_3)/2)!((l_2-l_1+l_3)/2)!}{l_3!}\right]$$

$$\times\sum_{r=0}^{\min[l_1-k,l_2-k]}(-1)^r 2^r\frac{(2l_3-2r-1)!!}{(2l_3-1)!!}\{\mathbf{A}^{\{l_1\}}\cdot(k+r)\mathbf{B}^{\{l_2\}}\delta\cdots(r)\cdots\delta\}, \tag{44}$$

when $l_1+l_2-l_3\equiv 2k$ is even, and by

$$[\mathbf{A}^{\{l_1\}}\times\mathbf{B}^{\{l_2\}}]^{\{l_3\}}=D_{l_1l_2l_3}\left[\frac{((l_1-l_2+l_3-1)/2)!((l_2-l_1+l_3-1)/2)!}{l_3!}\right]\cdot\frac{1}{\sqrt{2}}$$
$$\times\sum_{r=0}^{\min[l_1-k'-1,l_2-k'-1]}(-1)^r2^r\frac{(2l_3-2r-1)!!}{(2l_3-1)!!}\{\epsilon:\mathbf{A}^{\{l_1\}}\cdot(k'+r)\mathbf{B}^{\{l_2\}}\delta\cdots(r)\cdots\delta\}\qquad(45)$$

when $l_1+l_2-l_3\equiv 2k'+1$ is odd.

In these expressions, a single dot followed by a parenthetical value (k) between two Cartesian tensors indicates a tensor contraction of order k. In addition, the colon indicates a double contraction of the form

$$(\epsilon:\mathbf{AB})_{i_1i_2\cdots i_l}=\sum_{j,k}\epsilon_{i_1jk}\mathbf{A}_{ji_2\cdots}\mathbf{B}_{k\cdots i_l}.\qquad(46)$$

In Eq. (44), terms within the curly bracket are summed over permutations of the indices across A,B, and δ, leading to a symmetric tensor with

$$\left[\frac{l_3!}{(l_1-k-r)!(l_2-k-r)!2^rr!}\right]\qquad(47)$$

terms for each r, while for Eq. (45), the symmetrization bracket gives

$$\left[\frac{l_3!}{(l_1-k'-r-1)!(l_2-k'-r-1)!2^rr!}\right]\qquad(48)$$

terms for each r. The factors $C_{l_1l_2l_3}$ and $D_{l_1l_2l_3}$ will be determined by specializing the tensors in Eqs. (44) and (45) to ones constructed from vectors. The square-bracketed coefficients in Eqs. (44) and (45) are the inverse of the number of terms in the first symmetrization bracket of the following summation. The $(1/\sqrt{2})$ factor in Eq. (45) is inserted in anticipation of the concurrence of the C and D coefficients.

For $A^{\{l_1\}}=a^{\{l_1\}}$ and $B^{\{l_2\}}=a^{\{l_2\}}$, the right-hand side of Eq. (44) must be proportional to $a^{\{l_3\}}$. With these substitutions, the summations in Eq. (44) can be performed, giving

$$[a^{\{l_1\}}\times a^{\{l_2\}}]^{\{l_3\}}=C_{l_1l_2l_3}\frac{J_1!!J_2!!J_3!!(J/2)!}{l_1!l_2!(2l_3-1)!!}a^{\{l_3\}},\qquad(49)$$

where $J\equiv l_1+l_2+l_3$ and $J_i\equiv J-2l_i-1$, and $(-1)!!\equiv 1$.

With the spherical harmonic coupling identity

$$[\mathbf{Y}^{[l_1]}(\hat{\mathbf{c}})\times\mathbf{Y}^{[l_2]}(\hat{\mathbf{c}})]^{[l_3]}_m=\frac{\hat{l}_1\hat{l}_2}{\sqrt{4\pi}}\left|\begin{pmatrix}l_1&l_2&l_3\\0&0&0\end{pmatrix}\right|\mathbf{Y}^{[l_3]}_m(\hat{\mathbf{c}}),\qquad(50)$$

together with Eq. (43), we find

$$C_{l_1l_2l_3}=\hat{l}_3\left[\frac{(2l_1)!(2l_2)!(2l_3)!}{(J_1+1)!(J_2+1)!(J_3+1)!(J+1)!}\right]^{1/2}.\qquad(51)$$

For odd $l_1+l_2+l_3$, one can compare relation (45) for $A^{\{l_1\}}=a^{\{l_1\}}$ and $B^{\{l_2\}}=b^{\{l_2\}}$ with the corresponding spherical harmonic coupling. Using the Clebsch–Gordan coefficients for $m_3=0$, it follows (after some tedious algebra) that

$$D_{l_1l_2l_3}=C_{l_1l_2l_3}.\qquad(52)$$

The relations (44) and (45) with (51) and (52) constitute an explicit solution for the Clebsch–Gordan coefficients in an expansion of a product of irreducible tensors in Cartesian form.

V. APPLICATIONS TO SPHERICAL HARMONIC COUPLINGS

We now are in a position to reduce any set of spherical harmonic couplings to Cartesian form. Repeated application of the pairwise coupling formula (44) and (45) will necessarily lead to a Cartesian expression in the original vectors of the problem. For couplings to a scalar, clearly the result will be a polynomial in the scalar products of these vectors, with order no greater than the smaller of the ranks of the two spherical harmonics entering with these vector arguments. If the coupling is to a pseudoscalar, a "box" product (e.g., $\mathbf{a}\cdot\mathbf{b}\times\mathbf{c}$) of three independent vectors must be an overall factor.

The reduction of an arbitrary series of spherical harmonic couplings proceeds as follows: For each spherical harmonic, introduce the rescaling factors shown in Eq. (43). For each pair coupling, write the appropriate Cartesian coupling as in Eq. (44) or (45). Finally, perform the indicated Cartesian tensor contractions, starting with Eq. (39) for each spherical harmonic. In this process, the traceless nature of these tensors greatly simplifies the reduction, since the Kronecker delta's within one such tensor contracted with another irreducible tensor will vanish.

For example,[10] the above method can be used to show

$$[\mathbf{Y}^{[l]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[l]}(\hat{\mathbf{b}})]^{[1]}_m=\frac{(-i)}{4\pi}\left[\frac{3(2l+1)}{l(l+1)}\right]^{1/2}P_l'[\hat{\mathbf{a}}\times\hat{\mathbf{b}}]_m,\qquad(53a)$$

and

$$[\mathbf{Y}^{[l-1]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[l]}(\hat{\mathbf{b}})]^{[1]}_m=\frac{(-i)}{4\pi}\left[\frac{3}{l}\right]^{1/2}\times[P_l'\hat{\mathbf{b}}_m-((l-1)P_{l-2}+\hat{\mathbf{a}}\cdot\hat{\mathbf{b}}P_{l-2}')\hat{\mathbf{a}}_m],\qquad(53b)$$

where $P_l\equiv P_l(\hat{\mathbf{a}}\cdot\hat{\mathbf{b}})$ is the Legendre function of order l and P_l' is its derivative. Similarly, higher-order couplings of the form $[\mathbf{Y}^{[l_1]}(\hat{\mathbf{a}})\times Y^{[l_2]}(\hat{\mathbf{b}})]^{[L]}$ for $L=2,3,\ldots$ can be expressed in terms of the order-L "stretched" even or odd parity couplings of the vectors $\hat{\mathbf{a}}$ and $\hat{\mathbf{b}}$ times Legendre functions and their derivatives. They are most easily derived by expanding the given form in terms of an independent set of stretched couplings with unknown scalar coefficients, then contracting with each tensor of the set to form scalar relations for the coefficients.

In matrix element calculations, spherical harmonic couplings to total angular momentum of zero arise. In these cases, we have found it convenient to introduce a set of rules for generating the final scalar expression given the initial coupling. These rules result from the Cartesian recoupling formalism of the last section and are taken in a form which allows for an easy verification of each step.

The rules are as follows:

Step (1a): For each interior pair coupling of even parity, introduce the Cartesian tensor factor

$$Q_{l_1l_2l_3}(A,B) \equiv \left[\frac{l_1!l_2!(2l_3-1)!!((J_1+1)/2)!((J_2+1)/2)!}{l_3!J_1!!J_2!!J_3!!(J/2)!}\right] \times \sum_{r=0}^{\min[l_1-k,l_2-k]} (-1)^r 2^r \frac{(2l_3-2r-1)!!}{(2l_3-1)!!} \times\{A^{\{l_1\}}\cdot(k+r)B^{\{l_2\}}\delta\cdots(r)\cdots\delta\} \tag{54}$$

coming from the coupling in Eq. (44). As before, $J\equiv l_1+l_2+l_3$, $J_i\equiv J-2l_i-1$, $2k\equiv l_1+l_2+l_3$, $(-1)!!\equiv 1$, and the bracketed terms contain an implicit symmetrization sum, with the number of such terms given by the expression in (47). This rank-l_3 tensor has been normalized so that when $A^{\{l_1\}}=a^{\{l_1\}}$ and $B^{\{l_2\}}=a^{\{l_2\}}$, Q reduces to $a^{\{l_3\}}$. Thus the Q's are a natural generalization of the Cartesian harmonic tensors. Note also that

$$Q_{l_1l_2l_3}(a,a)\cdot(l_3)\hat{\mathbf{b}} = P_{l_3}(\hat{\mathbf{a}}\cdot\hat{\mathbf{b}})$$

and thus

$$Q_{l_1l_2l_3}(a,a)\cdot(l_3)\hat{\mathbf{a}} = 1.$$

Step (1b): For each interior pair coupling of odd parity, introduce the Cartesian tensor factor

$$R_{l_1l_2l_3}(A,B) = \left[\frac{2l_1!l_2!(2l_3-1)!!(J_1/2)!(J_2/2)!}{(l_3-1)!(J_1+1)!!(J_2+1)!!(J_3+1)!!((J+1)/2)!}\right] \times \sum_{r=0}^{\min[l_1-k'-1,l_2-k'-1]} (-1)^r 2^r \frac{(2l_3-2r-1)!!}{(2l_3-1)!!}\{\epsilon{:}A^{\{l_1\}}\cdot(k'+r)B^{\{l_2\}}\delta\cdots(r)\cdots\delta\} \tag{55}$$

coming from the coupling in Eq. (45) ($J\equiv l_1+l_2+l_3, J_i\equiv J-2l_i-1$, and $2k'+1\equiv l_1+l_2+l_3$). As before, the bracketed term contains an implicit symmetrization sum, with the number of such terms given in (48). This tensor has been normalized so that, for $A^{\{l_1\}}=a^{\{l_1\}}$, $B^{\{l_2\}}=b^{\{l_2\}}$ and as the vector $\hat{\mathbf{b}}$ approaches $\hat{\mathbf{a}}$, we have

$$\lim_{b\to a}\frac{|R_{l_1l_2l_3}(a,b)\cdot(l_3-1)\mathbf{a}|}{|\mathbf{a}\times\mathbf{b}|} = 1. \tag{56}$$

Step (2a): For even parity couplings, introduce a factor

$$q_{l_1l_2l_3} \equiv \frac{\hat{l}_1\hat{l}_2}{\sqrt{4\pi}}\left|\begin{pmatrix} l_1 & l_2 & l_3\\ 0 & 0 & 0\end{pmatrix}\right|$$

or

$$q_{l_1l_2l_3} = \frac{\hat{l}_1\hat{l}_2}{\sqrt{4\pi}}\left[\frac{J_1!!J_2!!J_3!!(J/2)!}{((J_1+1)/2)!((J_2+1)/2)!((J_3+1)/2)!(J+1)!!}\right]^{1/2}. \tag{57}$$

Our normalization for Q makes this the same factor which one would ordinarily use in coupling spherical harmonics with identical arguments.

Step (2b): For odd parity couplings, introduce a factor

$$r_{l_1l_2l_3} \equiv \frac{\hat{l}_1\hat{l}_2}{\sqrt{4\pi}}\left[\frac{l_1(l_1+1)l_2(l_2+1)}{2l_3\cdot 2l_3}\right]^{1/2}\left|\begin{pmatrix} l_1 & l_2 & l_3\\ 1 & -1 & 0\end{pmatrix}\right|$$

or

$$r_{l_1l_2l_3} = \frac{\hat{l}_1\hat{l}_2}{2l_3\sqrt{4\pi}} \times\left[\frac{(J_1+1)!!(J_2+1)!!(J_3+1)!!((J+1)/2)!}{(J_1/2)!(J_2/2)!(J_3/2)!J!!}\right]^{1/2}. \tag{58}$$

Step (3): For the final $L\times L$ coupling to 0, use a factor

$$S_L \equiv \frac{\hat{L}}{4\pi}\cdot\frac{L!}{(2L-1)!!} \tag{59}$$

and fully contract the final pair of Cartesian tensors. The factors in Eqs. (12) and (16) have been arranged to exhibit these steps.

As another example, consider the fourfold coupling

$$[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[2]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{c}})\times\mathbf{Y}^{[3]}(\hat{\mathbf{d}})]^{[2]}]^{[0]}. \tag{60}$$

Performing each step from right to left on the above coupling, we have

$$\frac{\sqrt{5}}{4\pi}\cdot\frac{2}{3}\cdot\sqrt{\frac{3\cdot 3}{5\cdot 4\pi}}\sqrt{\frac{2\cdot 5}{7\cdot 4\pi}}\,Q^{222}(a,b){:}Q^{132}(c,d), \tag{61}$$

where

$$Q^{222}(a,b) = \tfrac{3}{4}\{\mathbf{a}\cdot\mathbf{b}(\mathbf{ab}+\mathbf{ba}-(\tfrac{2}{3})\mathbf{a}\cdot\mathbf{b}\,\delta) - \tfrac{2}{3}(\mathbf{aa}-(\tfrac{1}{3})\delta) - \tfrac{2}{3}(\mathbf{bb}-(\tfrac{1}{3})\delta)\} \tag{62}$$

and

$$Q^{132}(c,d) = \tfrac{5}{2}\{\mathbf{c}\cdot\mathbf{d}(\mathbf{dd}-(\tfrac{1}{3})\delta) + \tfrac{1}{5}(\mathbf{cd}+\mathbf{dc}-(\tfrac{2}{3})\mathbf{c}\cdot\mathbf{d}\,\delta)\}. \tag{63}$$

The last contraction to a scalar is simplified by noting that all contractions of the Kronecker deltas in $Q^{222}(a,b)$ with $Q^{132}(c,d)$ must vanish. The surviving terms for $Q^{222}(a,b){:}Q^{132}(c,d)$ are

$$\tfrac{3}{4}\{-5(\mathbf{a\cdot d})^2(\mathbf{c\cdot d})+15(\mathbf{a\cdot b})(\mathbf{c\cdot d})(\mathbf{a\cdot d})(\mathbf{b\cdot d})$$
$$-5(\mathbf{b\cdot d})^2(\mathbf{c\cdot d})-3(\mathbf{a\cdot b})^2(\mathbf{c\cdot d})+2(\mathbf{c\cdot d})$$
$$+2(\mathbf{a\cdot c})(\mathbf{a\cdot d})-3(\mathbf{a\cdot b})(\mathbf{a\cdot c})(\mathbf{b\cdot d})$$
$$-3(\mathbf{a\cdot b})(\mathbf{a\cdot d})(\mathbf{b\cdot c})+2(\mathbf{b\cdot c})(\mathbf{b\cdot d})\}. \quad (64)$$

Inserting in Eq. (61) gives the final answer for the fourfold coupling shown in the Appendix as Eq. (A9).

Evidently the above procedure will work for spherical harmonics of arbitrarily high rank and argument, coupled to each other any number of times. The algorithm is susceptible to algebraic coding within a reasonably sophisticated algebraic manipulation program.

In the Appendix of this paper, we give results for a selection of spherical harmonic couplings as a reference and as a check of the implementation of our method.

VI. CONCLUSIONS

Although a large body of work covers angular coupling of irreducible tensors, explicit results for the coupling of Cartesian tensors of arbitrary rank have not been available. For many physical applications, using Cartesian coupling has some distinct advantages over the corresponding spherical case. We have shown that the Cartesian coupling of spherical harmonics can be performed in a straightforward manner, following a well defined procedure. The results are relatively simple and easy to interpret. Specifically, a simple algorithm permits one to write down directly a scalar expression for the coupling to zero of any number of spherical harmonics in terms of the unit vectors involved.

Note added in proof: After this manuscript was submitted, R. F. Snider brought to our attention earlier work on irreducible Cartesian tensors that the reader may find useful.[11]

ACKNOWLEDGMENT

The work of the authors is suppported in part by the U. S. Department of Energy under grant No. DE-FG05-86-ER40270.

APPENDIX

In this appendix, we give examples of the reduction described in the paper for some commonly found spherical harmonic couplings to a scalar. The results serve to show the simplicity of the expressions, to exhibit their usefulness for physical interpretations in terms of the initial vector directions contained in the spherical harmonics, and to act as reference. [Note: All vectors on the right-hand side of the equations below are unit vectors.]

$$[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times[\mathbf{Y}^{[1]}(\hat{\mathbf{b}})\times\mathbf{Y}^{[3]}(\hat{\mathbf{c}})]^{[2]}]^{[0]}=\frac{3}{16\pi^{3/2}}\{5(\mathbf{a\cdot c})^2(\mathbf{b\cdot c})-2(\mathbf{a\cdot b})(\mathbf{a\cdot c})-(\mathbf{b\cdot c})\}. \quad (\mathrm{A1})$$

$$[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[2]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[0]}=\frac{5}{8\sqrt{14}\pi^{3/2}}\{9(\mathbf{a\cdot b})(\mathbf{a\cdot c})(\mathbf{b\cdot c})-3(\mathbf{a\cdot b})^2-3(\mathbf{a\cdot c})^2-3(\mathbf{b\cdot c})^2+2\}. \quad (\mathrm{A2})$$

$$[\mathbf{Y}^{[1]}(\hat{\mathbf{a}})\times[\mathbf{Y}^{[1]}(\hat{\mathbf{b}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[1]}]^{[0]}=\frac{\sqrt{3}}{8\sqrt{2}\pi^{3/2}}\{3(\mathbf{a\cdot c})(\mathbf{b\cdot c})-(\mathbf{a\cdot b})\}. \quad (\mathrm{A3})$$

$$[[\mathbf{Y}^{[1]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{b}})]^{[2]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{c}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{d}})]^{[2]}]^{[0]}=\frac{3}{32\sqrt{5}\pi^2}\{3(\mathbf{a\cdot c})(\mathbf{b\cdot d})+3(\mathbf{b\cdot c})(\mathbf{a\cdot d})-2(\mathbf{a\cdot b})(\mathbf{c\cdot d})\}. \quad (\mathrm{A4})$$

$$[[\mathbf{Y}^{[1]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{b}})]^{[2]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{c}})\times\mathbf{Y}^{[3]}(\hat{\mathbf{d}})]^{[2]}]^{[0]}=\frac{3\sqrt{15}}{80\sqrt{2}\pi^2}\{5(\mathbf{a\cdot c})(\mathbf{b\cdot c})(\mathbf{c\cdot d})-(\mathbf{a\cdot c})(\mathbf{b\cdot d})-(\mathbf{b\cdot c})(\mathbf{a\cdot d})-(\mathbf{c\cdot d})(\mathbf{a\cdot b})\} \quad (\mathrm{A5})$$

$$[[\mathbf{Y}^{[1]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[3]}(\hat{\mathbf{b}})]^{[2]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{c}})\times\mathbf{Y}^{[3]}(\hat{\mathbf{d}})]^{[2]}]^{[0]}$$
$$=\frac{3\sqrt{5}}{160\pi^2}\{-10(\mathbf{c\cdot d})(\mathbf{a\cdot d})(\mathbf{b\cdot d})+25(\mathbf{c\cdot d})(\mathbf{b\cdot d})^2(\mathbf{a\cdot b})$$
$$-3(\mathbf{c\cdot d})(\mathbf{a\cdot b})+2(\mathbf{a\cdot c})(\mathbf{b\cdot d})+2(\mathbf{b\cdot c})(\mathbf{a\cdot d})-10(\mathbf{b\cdot c})(\mathbf{b\cdot d})(\mathbf{a\cdot b})\}. \quad (\mathrm{A6})$$

$$[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[1]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{c}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{d}})]^{[1]}]^{[0]}=\frac{3\sqrt{15}}{32\pi^2}(\mathbf{a\cdot b})(\mathbf{a\cdot c})\{(\mathbf{b\cdot d})-(\mathbf{b\cdot c})(\mathbf{a\cdot d})\}. \quad (\mathrm{A7})$$

$$[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[2]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{c}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{d}})]^{[2]}]^{[0]}=\frac{\sqrt{15}}{32\sqrt{7}\pi^2}\{-6(\mathbf{c\cdot d})(\mathbf{a\cdot b})^2+4(\mathbf{c\cdot d})-6(\mathbf{a\cdot c})(\mathbf{a\cdot d})$$
$$+9(\mathbf{a\cdot c})(\mathbf{b\cdot d})(\mathbf{a\cdot b})+9(\mathbf{b\cdot c})(\mathbf{a\cdot d})(\mathbf{a\cdot b})-6(\mathbf{b\cdot c})(\mathbf{b\cdot d})\}. \quad (\mathrm{A8})$$

$$[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[2]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{c}})\times\mathbf{Y}^{[3]}(\hat{\mathbf{d}})]^{[2]}]^{[0]}$$
$$=\frac{3\sqrt{5}}{16\sqrt{14}\pi^2}\{-5(\mathbf{c\cdot d})(\mathbf{a\cdot d})^2+15(\mathbf{c\cdot d})(\mathbf{a\cdot d})(\mathbf{b\cdot d})(\mathbf{a\cdot b})-5(\mathbf{c\cdot d})(\mathbf{b\cdot d})^2-3(\mathbf{c\cdot d})(\mathbf{a\cdot b})^2$$
$$+2(\mathbf{c\cdot d})+2(\mathbf{a\cdot c})(\mathbf{a\cdot d})-3(\mathbf{a\cdot c})(\mathbf{b\cdot d})(\mathbf{a\cdot b})-3(\mathbf{b\cdot c})(\mathbf{a\cdot d})(\mathbf{a\cdot b})+2(\mathbf{b\cdot c})(\mathbf{b\cdot d})\}. \quad (\mathrm{A9})$$

$$[[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[1]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[1]}\times[\mathbf{Y}^{[2]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{e}})]^{[1]}]^{[0]}$$

$$=\frac{15\sqrt{3}}{64\sqrt{2}\pi^{5/2}}(\mathbf{a}\cdot\mathbf{b})(\mathbf{d}\cdot\mathbf{e})\{3(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})-3(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{d})-3(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{e})$$
$$+2(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{e})+3(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{d})-2(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{d})\}. \tag{A10}$$

$$[[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[1]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[2]}\times[\mathbf{Y}^{[2]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{e}})]^{[2]}]^{[0]}$$

$$=\frac{15\sqrt{15}}{64\sqrt{14}\pi^{5/2}}(\mathbf{a}\cdot\mathbf{b})\{-2(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{d})+3(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})(\mathbf{d}\cdot\mathbf{e})+3(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{d})(\mathbf{d}\cdot\mathbf{e})$$
$$-2(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{e})+2(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{d})-3(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{e})(\mathbf{d}\cdot\mathbf{e})-3(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{d})(\mathbf{d}\cdot\mathbf{e})+2(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{e})\}. \tag{A11}$$

$$[[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[2]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[1]}\times[\mathbf{Y}^{[2]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{e}})]^{[1]}]^{[0]}$$

$$=\frac{15\sqrt{15}}{64\sqrt{14}\pi^{5/2}}(\mathbf{d}\cdot\mathbf{e})\{3(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})-3(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{d})+3(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{e})$$
$$-3(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{d})-2(\mathbf{a}\cdot\mathbf{c})(\mathbf{a}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})+2(\mathbf{a}\cdot\mathbf{c})(\mathbf{a}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{d})-2(\mathbf{b}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})+2(\mathbf{b}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{d})\}. \tag{A12}$$

$$[[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[2]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[2]}\times[\mathbf{Y}^{[2]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{e}})]^{[2]}]^{[0]}$$

$$=\frac{25}{448\sqrt{14}\pi^{5/2}}\{36(\mathbf{a}\cdot\mathbf{b})^2(\mathbf{c}\cdot\mathbf{d})^2-108(\mathbf{a}\cdot\mathbf{b})^2(\mathbf{c}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})(\mathbf{d}\cdot\mathbf{e})+36(\mathbf{a}\cdot\mathbf{b})^2(\mathbf{c}\cdot\mathbf{e})^2+72(\mathbf{a}\cdot\mathbf{b})^2(\mathbf{d}\cdot\mathbf{e})^2$$
$$-48(\mathbf{a}\cdot\mathbf{b})^2-108(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{c})(\mathbf{d}\cdot\mathbf{e})^2+72(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{c})-54(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{d})$$
$$+81(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})(\mathbf{d}\cdot\mathbf{e})+81(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{d})(\mathbf{d}\cdot\mathbf{e})-54(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{e})$$
$$-54(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{d})+81(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{e})(\mathbf{d}\cdot\mathbf{e})+36(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{d})$$
$$-54(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{e})(\mathbf{d}\cdot\mathbf{e})+81(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{d})(\mathbf{d}\cdot\mathbf{e})-54(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{e})$$
$$-54(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{d})(\mathbf{d}\cdot\mathbf{e})+36(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{e})+36(\mathbf{a}\cdot\mathbf{c})^2(\mathbf{d}\cdot\mathbf{e})^2-24(\mathbf{a}\cdot\mathbf{c})^2$$
$$+36(\mathbf{a}\cdot\mathbf{c})(\mathbf{a}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{d})-54(\mathbf{a}\cdot\mathbf{c})(\mathbf{a}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})(\mathbf{d}\cdot\mathbf{e})-54(\mathbf{a}\cdot\mathbf{c})(\mathbf{a}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{d})(\mathbf{d}\cdot\mathbf{e})+36(\mathbf{a}\cdot\mathbf{c})(\mathbf{a}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{e})$$
$$-12(\mathbf{a}\cdot\mathbf{d})^2+36(\mathbf{a}\cdot\mathbf{d})(\mathbf{a}\cdot\mathbf{e})(\mathbf{d}\cdot\mathbf{e})-12(\mathbf{a}\cdot\mathbf{e})^2+36(\mathbf{b}\cdot\mathbf{c})^2(\mathbf{d}\cdot\mathbf{e})^2-24(\mathbf{b}\cdot\mathbf{c})^2+36(\mathbf{b}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{d})$$
$$-54(\mathbf{b}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})(\mathbf{d}\cdot\mathbf{e})-54(\mathbf{b}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{d})(\mathbf{d}\cdot\mathbf{e})+36(\mathbf{b}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{e})-12(\mathbf{b}\cdot\mathbf{d})^2$$
$$+36(\mathbf{b}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{e})(\mathbf{d}\cdot\mathbf{e})-12(\mathbf{b}\cdot\mathbf{e})^2-24(\mathbf{c}\cdot\mathbf{d})^2+72(\mathbf{c}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})(\mathbf{d}\cdot\mathbf{e})-24(\mathbf{c}\cdot\mathbf{e})^2-48(\mathbf{d}\cdot\mathbf{e})^2+32\}. \tag{A13}$$

$$[[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[1]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[1]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{e}})]^{[1]}]^{[0]}$$

$$=\frac{3\sqrt{15}}{64\sqrt{2}\pi^{5/2}}(\mathbf{a}\cdot\mathbf{b})\{3(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})-3(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{d})-3(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{e})$$
$$+2(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{e})+3(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{d})-2(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{d})\}. \tag{A14}$$

$$[[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[1]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[2]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{e}})]^{[2]}]^{[0]}$$

$$=\frac{9\sqrt{5}}{64\sqrt{2}\pi^{5/2}}(\mathbf{a}\cdot\mathbf{b})\{(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})+(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{d})-(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{e})-(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{d})\}. \tag{A15}$$

$$[[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[2]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[1]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{e}})]^{[1]}]^{[0]}$$

$$=\frac{15\sqrt{3}}{64\sqrt{14}\pi^{5/2}}\{3(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})-3(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{d})+3(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{e})-3(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{d})$$
$$-2(\mathbf{a}\cdot\mathbf{c})(\mathbf{a}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})+2(\mathbf{a}\cdot\mathbf{c})(\mathbf{a}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{d})-2(\mathbf{b}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})+2(\mathbf{b}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{d})\}. \tag{A16}$$

$$[[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[2]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[2]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{e}})]^{[2]}]^{[0]}$$

$$=\frac{5\sqrt{3}}{448\sqrt{2}\pi^{5/2}}\{-36(\mathbf{a}\cdot\mathbf{b})^2(\mathbf{c}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})+24(\mathbf{a}\cdot\mathbf{b})^2(\mathbf{d}\cdot\mathbf{e})-36(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{c})(\mathbf{d}\cdot\mathbf{e})$$
$$+27(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})+27(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{d})+27(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{e})$$
$$-18(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{e})+27(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{d})-18(\mathbf{a}\cdot\mathbf{b})(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{d})+12(\mathbf{a}\cdot\mathbf{c})^2(\mathbf{d}\cdot\mathbf{e})$$
$$-18(\mathbf{a}\cdot\mathbf{c})(\mathbf{a}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})-18(\mathbf{a}\cdot\mathbf{c})(\mathbf{a}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{d})+12(\mathbf{a}\cdot\mathbf{d})(\mathbf{a}\cdot\mathbf{e})+12(\mathbf{b}\cdot\mathbf{c})^2(\mathbf{d}\cdot\mathbf{e})-18(\mathbf{b}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})$$
$$-18(\mathbf{b}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{d})+12(\mathbf{b}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{e})+24(\mathbf{c}\cdot\mathbf{d})(\mathbf{c}\cdot\mathbf{e})-16(\mathbf{d}\cdot\mathbf{e})\}. \tag{A17}$$

$$[[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[1]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[2]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[3]}(\hat{\mathbf{e}})]^{[2]}]^{[0]}$$
$$=\frac{3\sqrt{15}}{64\pi^{5/2}}(\mathbf{a\cdot b})\{-(\mathbf{a\cdot c})(\mathbf{b\cdot d})(\mathbf{c\cdot e})-(\mathbf{a\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot d})+5(\mathbf{a\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot e})(\mathbf{d\cdot e})$$
$$+(\mathbf{a\cdot d})(\mathbf{b\cdot c})(\mathbf{c\cdot e})+(\mathbf{a\cdot e})(\mathbf{b\cdot c})(\mathbf{c\cdot d})-5(\mathbf{a\cdot e})(\mathbf{b\cdot c})(\mathbf{c\cdot e})(\mathbf{d\cdot e})\}. \tag{A18}$$

$$[[[\mathbf{Y}^{[2]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[2]}(\hat{\mathbf{b}})]^{[2]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[2]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[3]}(\hat{\mathbf{e}})]^{[2]}]^{[0]}$$
$$=\frac{15}{448\pi^{5/2}}\{12(\mathbf{a\cdot b})^2(\mathbf{c\cdot d})(\mathbf{c\cdot e})-30(\mathbf{a\cdot b})^2(\mathbf{c\cdot e})^2(\mathbf{d\cdot e})+12(\mathbf{a\cdot b})^2(\mathbf{d\cdot e})$$
$$-18(\mathbf{a\cdot b})(\mathbf{a\cdot c})(\mathbf{b\cdot c})(\mathbf{d\cdot e})-9(\mathbf{a\cdot b})(\mathbf{a\cdot c})(\mathbf{b\cdot d})(\mathbf{c\cdot e})-9(\mathbf{a\cdot b})(\mathbf{a\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot d})+45(\mathbf{a\cdot b})(\mathbf{a\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot e})(\mathbf{d\cdot e})$$
$$-9(\mathbf{a\cdot b})(\mathbf{a\cdot d})(\mathbf{b\cdot c})(\mathbf{c\cdot e})+6(\mathbf{a\cdot b})(\mathbf{a\cdot d})(\mathbf{b\cdot e})-9(\mathbf{a\cdot b})(\mathbf{a\cdot e})(\mathbf{b\cdot c})(\mathbf{c\cdot d})+45(\mathbf{a\cdot b})(\mathbf{a\cdot e})(\mathbf{b\cdot c})(\mathbf{c\cdot e})(\mathbf{d\cdot e})$$
$$+6(\mathbf{a\cdot b})(\mathbf{a\cdot e})(\mathbf{b\cdot d})-30(\mathbf{a\cdot b})(\mathbf{a\cdot e})(\mathbf{b\cdot e})(\mathbf{d\cdot e})+6(\mathbf{a\cdot c})^2(\mathbf{d\cdot e})+6(\mathbf{a\cdot c})(\mathbf{a\cdot d})(\mathbf{c\cdot e})+6(\mathbf{a\cdot c})(\mathbf{a\cdot e})(\mathbf{c\cdot d})$$
$$-30(\mathbf{a\cdot c})(\mathbf{a\cdot e})(\mathbf{c\cdot e})(\mathbf{d\cdot e})-4(\mathbf{a\cdot d})(\mathbf{a\cdot e})+10(\mathbf{a\cdot e})^2(\mathbf{d\cdot e})+6(\mathbf{b\cdot c})^2(\mathbf{d\cdot e})+6(\mathbf{b\cdot c})(\mathbf{b\cdot d})(\mathbf{c\cdot e})$$
$$+6(\mathbf{b\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot d})-30(\mathbf{b\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot e})(\mathbf{d\cdot e})-4(\mathbf{b\cdot d})(\mathbf{b\cdot e})+10(\mathbf{b\cdot e})^2(\mathbf{d\cdot e})$$
$$-8(\mathbf{c\cdot d})(\mathbf{c\cdot e})+20(\mathbf{c\cdot e})^2(\mathbf{d\cdot e})-8(\mathbf{d\cdot e})\}. \tag{A19}$$

$$[[[\mathbf{Y}^{[1]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{b}})]^{[1]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[1]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{e}})]^{[1]}]^{[0]}$$
$$=\frac{3\sqrt{3}}{64\sqrt{2}\pi^{5/2}}\{3(\mathbf{a\cdot c})(\mathbf{b\cdot d})(\mathbf{c\cdot e})-3(\mathbf{a\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot d})-3(\mathbf{a\cdot d})(\mathbf{b\cdot c})(\mathbf{c\cdot e})$$
$$+2(\mathbf{a\cdot d})(\mathbf{b\cdot e})+3(\mathbf{a\cdot e})(\mathbf{b\cdot c})(\mathbf{c\cdot d})-2(\mathbf{a\cdot e})(\mathbf{b\cdot d})\}. \tag{A20}$$

$$[[[\mathbf{Y}^{[1]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{b}})]^{[2]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[2]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{e}})]^{[2]}]^{[0]}$$
$$=\frac{3}{64\sqrt{14}\pi^{5/2}}\{-12(\mathbf{a\cdot b})(\mathbf{c\cdot d})(\mathbf{c\cdot e})+8(\mathbf{a\cdot b})(\mathbf{d\cdot e})-12(\mathbf{a\cdot c})(\mathbf{b\cdot c})(\mathbf{d\cdot e})+9(\mathbf{a\cdot c})(\mathbf{b\cdot d})(\mathbf{c\cdot e})$$
$$+9(\mathbf{a\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot d})+9(\mathbf{a\cdot d})(\mathbf{b\cdot c})(\mathbf{c\cdot e})-6(\mathbf{a\cdot d})(\mathbf{b\cdot e})+9(\mathbf{a\cdot e})(\mathbf{b\cdot c})(\mathbf{c\cdot d})-6(\mathbf{a\cdot e})(\mathbf{b\cdot d})\}. \tag{A21}$$

$$[[[\mathbf{Y}^{[1]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{b}})]^{[1]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[2]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[3]}(\hat{\mathbf{e}})]^{[2]}]^{[0]}$$
$$=\frac{3\sqrt{3}}{64\pi^{5/2}}\{-(\mathbf{a\cdot c})(\mathbf{b\cdot d})(\mathbf{c\cdot e})-(\mathbf{a\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot d})+5(\mathbf{a\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot e})(\mathbf{d\cdot e})$$
$$+(\mathbf{a\cdot d})(\mathbf{b\cdot c})(\mathbf{c\cdot e})+(\mathbf{a\cdot e})(\mathbf{b\cdot c})(\mathbf{c\cdot d})-5(\mathbf{a\cdot e})(\mathbf{b\cdot c})(\mathbf{c\cdot e})(\mathbf{d\cdot e})\}. \tag{A22}$$

$$[[[\mathbf{Y}^{[1]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{b}})]^{[2]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[2]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[3]}(\hat{\mathbf{e}})]^{[2]}]^{[0]}$$
$$=\frac{3\sqrt{3}}{64\sqrt{7}\pi^{5/2}}\{4(\mathbf{a\cdot b})(\mathbf{c\cdot d})(\mathbf{c\cdot e})-10(\mathbf{a\cdot b})(\mathbf{c\cdot e})^2(\mathbf{d\cdot e})+4(\mathbf{a\cdot b})(\mathbf{d\cdot e})-6(\mathbf{a\cdot c})(\mathbf{b\cdot c})(\mathbf{d\cdot e})$$
$$-3(\mathbf{a\cdot c})(\mathbf{b\cdot d})(\mathbf{c\cdot e})-3(\mathbf{a\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot d})+15(\mathbf{a\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot e})(\mathbf{d\cdot e})-3(\mathbf{a\cdot d})(\mathbf{b\cdot c})(\mathbf{c\cdot e})+2(\mathbf{a\cdot d})(\mathbf{b\cdot e})$$
$$-3(\mathbf{a\cdot e})(\mathbf{b\cdot c})(\mathbf{c\cdot d})+15(\mathbf{a\cdot e})(\mathbf{b\cdot c})(\mathbf{c\cdot e})(\mathbf{d\cdot e})+2(\mathbf{a\cdot e})(\mathbf{b\cdot d})-10(\mathbf{a\cdot e})(\mathbf{b\cdot e})(\mathbf{d\cdot e})\}. \tag{A23}$$

$$[[[\mathbf{Y}^{[1]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[3]}(\hat{\mathbf{b}})]^{[2]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[1]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{e}})]^{[1]}]^{[0]}$$
$$=\frac{3\sqrt{3}}{64\pi^{5/2}}\{5(\mathbf{a\cdot b})(\mathbf{b\cdot c})(\mathbf{b\cdot d})(\mathbf{c\cdot e})-5(\mathbf{a\cdot b})(\mathbf{b\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot d})-(\mathbf{a\cdot c})(\mathbf{b\cdot d})(\mathbf{c\cdot e})$$
$$+(\mathbf{a\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot d})-(\mathbf{a\cdot d})(\mathbf{b\cdot c})(\mathbf{c\cdot e})+(\mathbf{a\cdot e})(\mathbf{b\cdot c})(\mathbf{c\cdot d})\}. \tag{A24}$$

$$[[[\mathbf{Y}^{[1]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[3]}(\hat{\mathbf{b}})]^{[2]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[2]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[1]}(\hat{\mathbf{e}})]^{[2]}]^{[0]}$$
$$=\frac{3\sqrt{3}}{64\sqrt{7}\pi^{5/2}}\{-10(\mathbf{a\cdot b})(\mathbf{b\cdot c})^2(\mathbf{d\cdot e})+15(\mathbf{a\cdot b})(\mathbf{b\cdot c})(\mathbf{b\cdot d})(\mathbf{c\cdot e})+15(\mathbf{a\cdot b})(\mathbf{b\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot d})$$
$$-10(\mathbf{a\cdot b})(\mathbf{b\cdot d})(\mathbf{b\cdot e})-6(\mathbf{a\cdot b})(\mathbf{c\cdot d})(\mathbf{c\cdot e})+4(\mathbf{a\cdot b})(\mathbf{d\cdot e})+4(\mathbf{a\cdot c})(\mathbf{b\cdot c})(\mathbf{d\cdot e})-3(\mathbf{a\cdot c})(\mathbf{b\cdot d})(\mathbf{c\cdot e})$$
$$-3(\mathbf{a\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot d})-3(\mathbf{a\cdot d})(\mathbf{b\cdot c})(\mathbf{c\cdot e})+2(\mathbf{a\cdot d})(\mathbf{b\cdot e})-3(\mathbf{a\cdot e})(\mathbf{b\cdot c})(\mathbf{c\cdot d})+2(\mathbf{a\cdot e})(\mathbf{b\cdot d})\}. \tag{A25}$$

$$[[[\mathbf{Y}^{[1]}(\hat{\mathbf{a}})\times\mathbf{Y}^{[3]}(\hat{\mathbf{b}})]^{[2]}\times\mathbf{Y}^{[2]}(\hat{\mathbf{c}})]^{[2]}\times[\mathbf{Y}^{[1]}(\hat{\mathbf{d}})\times\mathbf{Y}^{[3]}(\hat{\mathbf{e}})]^{[2]}]^{[0]}$$
$$=\frac{3}{32\sqrt{14}\pi^{5/2}}\{-15(\mathbf{a\cdot b})(\mathbf{b\cdot c})^2(\mathbf{d\cdot e})-15(\mathbf{a\cdot b})(\mathbf{b\cdot c})(\mathbf{b\cdot d})(\mathbf{c\cdot e})-15(\mathbf{a\cdot b})(\mathbf{b\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot d})$$
$$+75(\mathbf{a\cdot b})(\mathbf{b\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot e})(\mathbf{d\cdot e})+10(\mathbf{a\cdot b})(\mathbf{b\cdot d})(\mathbf{b\cdot e})-25(\mathbf{a\cdot b})(\mathbf{b\cdot e})^2(\mathbf{d\cdot e})+6(\mathbf{a\cdot b})(\mathbf{c\cdot d})(\mathbf{c\cdot e})$$
$$-15(\mathbf{a\cdot b})(\mathbf{c\cdot e})^2(\mathbf{d\cdot e})+6(\mathbf{a\cdot b})(\mathbf{d\cdot e})+6(\mathbf{a\cdot c})(\mathbf{b\cdot c})(\mathbf{d\cdot e})+3(\mathbf{a\cdot c})(\mathbf{b\cdot d})(\mathbf{c\cdot e})+3(\mathbf{a\cdot c})(\mathbf{b\cdot e})(\mathbf{c\cdot d})$$

$$- 15(\mathbf{a}\cdot\mathbf{c})(\mathbf{b}\cdot\mathbf{e})(\mathbf{c}\cdot\mathbf{e})(\mathbf{d}\cdot\mathbf{e}) + 3(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{e}) - 2(\mathbf{a}\cdot\mathbf{d})(\mathbf{b}\cdot\mathbf{e}) + 3(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{d})$$
$$- 15(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{c})(\mathbf{c}\cdot\mathbf{e})(\mathbf{d}\cdot\mathbf{e}) - 2(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{d}) + 10(\mathbf{a}\cdot\mathbf{e})(\mathbf{b}\cdot\mathbf{e})(\mathbf{d}\cdot\mathbf{e})\}. \quad \text{(A26)}$$

[1]R. Balian and E. Brezin, Nuovo Cimento B **61**, 403 (1969).

[2]W. Kohn and N. Rostoker, Phys. Rev. **94**, 1111 (1954); see also M. Danos and L. C. Maximon, J. Math. Phys. **6**, 766 (1965) for further references going back to Lord Rayleigh.

[3]This method has been used by J. L. Friar and G. L. Payne in two- and three-body calculations; for details, see J. L. Friar and G. L. Payne, Phys. Rev. C **38**, 1 (1988).

[4]The method was originally developed by one of the authors (DRL) in conjunction with the derivation of the three-body, bound-state equations for ^{6}He and ^{6}Li [A. Ghovanlou and D. R. Lehman, Phys. Rev. C **9**, 1730 (1973); D. R. Lehman, M. Rai, and A. Ghovanlou, Phys. Rev. C **17**, 744 (1978)]. For the work on the $A = 6$ system, the method was worked out for angular-momentum values up to $l = 5$, and used by DRL and his collaborators in numerous applications since that time [for example, D. R. Lehman and M. Rajan, Phys. Rev. C **25**, 2743 (1982); B. F. Gibson and D. R. Lehman, Phys. Rev. C **29**, 1017 (1984)]. Recently, in association with our work on the ^{6}Li quadrupole form factor with A. Eskandarian [A. Eskandarian, D. R. Lehman, and W. C. Parke, Phys. Rev. C **38**, 2341 (1988)], where the method was used to obtain programmable expressions for five spherical harmonics coupled to zero, WCP generalized the method to arbitrary l and derived the irreducible decomposition of a product of two irreducible Cartesian tensors of any rank.

[5]U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic, New York, 1959), pp. 36–38.

[6]M. Danos, Ann. Phys. **63**, 319 (1971); D. R. Lehman and J. S. O'Connell, *Graphical Recoupling of Angular Momenta*, National Bureau of Standards Monograph 136 (1973), p. 12; M. Danos, V. Gillet, and M.Cauvin, *Methods in Relativistic Nuclear Physics* (North-Holland, Amsterdam, 1984), p. 59.

[7]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton UP, Princeton, NJ, 1960), Chaps. 6 and 7.

[8]A. R. Edmonds, Ref. 7, Eq. (2.5.29).

[9]U. Fano and G. Racah, Ref. 5, p. 21.

[10]See Eq. (17b) of Friar and Payne, Ref. 3 above. We used similar relations to derive the equations for the shell structure of the $A = 6$ ground state from three-body dynamics as given in the appendix of D. R. Lehman and W. C. Parke, Phys. Rev. C **28**, 364 (1983).

[11]J. A. R. Coope, R. F. Snider, and F. R. McCourt, J. Chem. Phys. **43**, 2269 (1965); J. A. R. Coope and R. F. Snider, J. Math. Phys. **11**, 1003 (1970); J. A. R. Coope, *ibid.* **11**, 1591 (1970).

Canonical formalism for path-dependent Lagrangians. Coupling constant expansions

X. Jaén, R. Jáuregui,[a] J. Llosa, and A. Molina
Grup de Relativitat, Societat Catalana de Física (I.E.C.) and Departament Física Fonamental, Universitat de Barcelona, Diagonal, 645, 08028-Barcelona, Spain

(Received 10 March 1989; accepted for publication 14 June 1989)

A canonical formalism obtained for path-dependent Lagrangians is applied to Fokker-type Lagrangians. The results are specialized for coupling constant expansions and later on are applied to relativistic systems of particles interacting through symmetric scalar and vector mesodynamics and electrodynamics.

I. INTRODUCTION

Hamiltonian formalism is a desirable feature to demand for a classical system. Indeed, it enables us to define the energy-momentum four-vector and the total angular momentum tensor as generating functions for Poincaré (resp. Galilei) infinitesimal transformations. Furthermore, it permits us to construct a statistical mechanics and a quantum mechanics according to some standard well defined rules.

On the other hand, relativistic dynamics for directly interacting particles, i.e., without an intermediate field, has been developed following a wide variety of approaches.[1] In most of them a canonical formalism has been obtained, either because one *a priori* starts from a Hamiltonian system,[2] or because an invariant symplectic form is obtained.[3]

However, there is an approach to relativistic action-at-a-distance, that based on path-dependent Lagrangian systems,[4–6] which has long refused a Hamiltonian formulation.[7] This is especially striking since the starting point is a Lagrangian system and a variational principle, although of a very particular kind.

Path-dependent Lagrangians were first used by Fokker,[4] who proposed an action principle for symmetric electrodynamics—half-retarded plus half-advanced—of two charges without an intermediate field. This is the reason why Lagrangians of this kind are also called "Fokker-type Lagrangians."

The symmetric electrodynamics of Feynman and Wheeler[5,8] is a generalization of that of Fokker to the case of more than two charges.

Several other relativistic theories of noninstantaneous action-at-a-distance between particles have been set in terms of Fokker-type Lagrangians.[6] This is usually the case of those interactions that are somehow related to a classical field.

Path-dependent Lagrangians exhibit a functional dependence on the trajectories as a whole. That makes these systems more complex than standard ones; but, since they permit us to consider interaction terms depending on noninstantaneous configurations of particles, these Lagrangians are especially useful for describing relativistic systems of directly interacting particles. The basic claim of these theories is that the *field* is nothing but a useful tool to describe forces between particles. The concept of *field*, introduced by Faraday and Maxwell as an intermediate tool to describe the action of some given "source" and a "test charge," must be acknowledged as one of the most fruitful in theoretical physics. However, the self-interaction divergences in classical field theory occur as a result of allowing the field to act on its own source. In order to avoid this "improperty," some authors[5] introduce what they call the "adjunct field." Namely, in Wheeler–Feynman symmetric electrodynamics, each charge is acted on by the "adjunct field" of the others, that is, half the sum of the advanced and retarded solutions of the Maxwell equations for the other charges. This leads to a Fokker-type Lagrangian, which only depends on particle variables.

The most important drawback of Fokker-type systems is that the Euler equations, derived from the Fokker action principle, are of functional–differential type (difference–differential equations in the simplest cases). Therefore, the evolution space (space of initial data) is non-Newtonian; that is, the positions and velocities in a given instant of "time" do not determine uniquely the future evolution of the system. Furthermore, the evolution space is not even well determined.[7] As a consequence, it has not been possible to generalize an algorithm as a Legendre transformation to Fokker-type Lagrangians, nor has an equivalent Hamiltonian formalism been set up yet. We can overcome this problem[9] by changing the point of view that is usual in classical mechanics. Since the Euler equations derived from a path-dependent Lagrangian are of functional–differential type, the initial data space for a Fokker Lagrangian has infinitely many dimensions. In our approach, a whole trajectory of the system is taken as the "initial datum." In doing this, the Euler equations do not rule the evolution anymore (all information about it is already contained in the initial datum), and they are merely considered as constraints on the initial data.

This approach somehow corresponds to a static point of view. The situation is similar to what happens in dealing with a static standard Lagrangian $L(q,t)$ (i.e., one depending on coordinates only): the initial data $(q_{0a})_{a=1,\dots,m}$ can, in principle, be picked out from an m-dimensional continuum, but the physically significant ones are only those satisfying

[a] Permanent address: Instituto de Física, Universidad Nacional Autonoma de México Apdo. Postal 20-364, México, 01000, D.F., Mexico.

 0022-2488/89/122807-08$02.50

$$\left(\frac{\partial L}{\partial q_a}\right)_{q_{01}\cdots q_{0m}}=0.$$

Similarly, in our approach to path-dependent Lagrangian systems, one can, in principle, take any m-tuple of curves $[q_1(t_1),...,q_m(t_m)]$ as initial data, but these will be physically significant only if we have been lucky enough to have chosen a set of curves fulfilling the Euler functional–differential equations, now considered as constraints.

As has been mentioned elsewhere, those functional differential equations admit "too many solutions": contrary to what we are used to in Newtonian mechanics, the knowledge of all positions and velocities at a given time does not determine the future evolution of the system. This is a feature of Newtonian physics that seems worthy to preserve, as many authors[10,11] have considered. Hence several criteria have been put forward in order to get rid of the "physically irrelevant" solutions. Among these criteria, it is worth mentioning that (a) when coupling constants go to zero, motions must become free (namely, uniform and rectilinear)[12–14]; (b) in the limit $1/c^2\to 0$, the solutions must yield the Newtonian ones[11,15,16]; and (c) if one mass is much bigger than than the others, then the external field approximation must be recovered.[17,18]

Any of these three criteria is implemented by requiring the physically relevant solutions to be analytical in the corresponding parameter, namely, the coupling constants, the inverse speed of light, or the mass ratio, respectively. Any of them selects a family of solutions of the functional–differential equations parametrized by either $6N$ (noncovariant formalism) or $8N$ (covariant formalism) parameters. It then happens (apparently as a consequence of the special structure of the Fokker-type Lagrangian) that these solutions also satisfy a second-order differential system that can be obtained by techniques[19] provided by predictive relativistic mechanics. This differential system is called a *second-order reduction*[19a] of the functional–differential system.

So that, in the covariant formalism given in Ref. 9, we consider the map

$$\varphi:\ \mathrm{TM}_4^N\to E,$$
$$(x,\dot x)\to\varphi_a^\mu(\xi,x_b,\dot x_c),$$

where φ_a^μ is the predictive solution of the functional–differential system determined by one of the above-mentioned criteria and the initial datum $(x_b,\dot x_c)$.

Then, using a kind of Ostrogradski transformation, we set up a Hamiltonian formalism for path-dependent Lagrangians (Sec. II). Once this has been done for Fokker-type Lagrangian systems, as a particular case of path-dependent Lagrangians, it can be specialized to any reduction of order of the Fokker system.

The paper is organized as follows: in Sec. II we give the canonical formalism for Fokker-type systems. In Sec. III, we derive a presymplectic form on the infinite-dimensional evolution space E of a Fokker-type system of relativistic particles with two-body interactions. Then (Sec. IV) this presymplectic form is specialized to the Newton-like evolution space TM_4^N, which results from implementing the condition of analytical dependence on the coupling constant. In the latter, we briefly describe perturbation theory and introduce the three-dimensional formalism.

The results obtained are then applied to scalar and vector interactions, up to the first order of approximation in the product g_ag_b of the coupling constants. Special attention is paid to Wheeler–Feynman electrodynamics, for which expansions on $1/c^2$ are carried out and conserved quantities are also calculated.

II. HAMILTONIAN FORMALISM FOR FOKKER LAGRANGIANS

Let us consider a Fokker Lagrangian

$$L=-\sum_{a=1}^N m_a(-\dot x_a^2(t))^{1/2}-\sum_{\substack{a,b=1\\a\neq b}}^N g_ag_b\int_{\mathbf R}d\xi\,\omega^r_{ab}(t,t+\xi),\tag{2.1}$$

where

$$\omega^r_{ab}(t_a,t_b)=G_{ab}((x_a(t_a)-x_b(t_b))^2)F_{ab}(\dot x_a(t_a),\dot x_b(t_b))\tag{2.2}$$

and

$$F_{ab}(\dot x_a,\dot x_b)=(-\dot x_a\dot x_b)^r(-\dot x_a^2)^{(1-r)/2}(-\dot x_b^2)^{(1-r)/2},\tag{2.3}$$

$a^2=a_\mu a^\mu$, for any four-vector a_μ, and $r\in N$ depends on the specific interaction we are considering.

Notice that L is a homogeneous function of first degree of velocities. Therefore, the corresponding action integral is reparametrization invariant. We also notice that

$$\omega^r_{ab}(t_a,t_b)=\omega^r_{ba}(t_b,t_a).\tag{2.4}$$

This Fokker Lagrangian can be put in the form

$$L_t=\int_{\mathbf R^m}d\xi_1,...,d\xi_m\,\mathscr L(q_a(t+\xi_a),\dot q_b(t+\xi_b),\xi_c),$$
$$a,b,c=1,...,m,\tag{2.5}$$

by merely taking

$$\mathscr L=\mathscr L_0+\mathscr L_{\mathrm I}\tag{2.5a}$$

with

$$\mathscr L_0=-\sum_{a=1}^N m_a(-\dot x_a^2\ (t+\xi_a))^{1/2}\prod_{d=1}^N\delta(\xi_d),\tag{2.5b}$$

$$\mathscr L_{\mathrm I}=-\frac12\sum_{\substack{a,b=1\\a\neq b}}^N g_ag_b\omega^r_{ab}(t+\xi_a,t+\xi_b)\prod_{d=1}^N\delta(\xi_d).\tag{2.5c}$$

In Ref. 9 we showed how a canonical formalism can be obtained for this kind of Lagrangian (2.5).

The procedure can be briefly described as follows. We start from the action $S=\int L_t\,dt$ and the variational principle $\delta S=0$, where the variations $\delta q_a(\xi_a)$ are taken so that they have compact support, gives rise to the equations of functional type

$$\int_{\mathbf R}d\xi\left[f_a(\tau-\xi,\xi)-\frac{\partial}{\partial\tau}g_a(\tau-\xi,\xi)\right]=0,\tag{2.6}$$

where

$$f_a(\tau,\xi_a)=\int_{\mathbb{R}^{m-1}}\prod_{b\neq a}d\xi_b\,\frac{\partial\mathscr{L}}{\partial q_a(\tau+\xi_a)} \tag{2.7a}$$

and

$$g_a(\tau,\xi_a)=\int_{\mathbb{R}^{m-1}}\prod_{b\neq a}d\xi_b\,\frac{\partial\mathscr{L}}{\partial\dot q_a(\tau+\xi_a)}. \tag{2.7b}$$

However, because of the functional–differential character of the equations of motion, it is necessary to clarify their meaning. In principle, the initial data that should be given to specify a unique solution are the whole functions $q_a(\lambda)$ themselves. As a consequence, these equations [(2.6)] must be regarded as constraints defining the evolution space E (also called initial data space), rather than as laws of motion.

We then consider the Hamiltonian

$$H=\sum_{a=1}^{m}\int_{\mathbb{R}}d\lambda\,p_a(\lambda)\dot q_a(\lambda)-L_0[q_b(\xi_b)] \tag{2.8}$$

defined on a phase space Γ labeled by

$$q_b(\xi_b),\ p_a(\xi_a),\quad a,b=1,\dots,m,\quad \xi_a,\xi_b\in\mathbb{R}, \tag{2.9}$$

with elementary Poisson brackets

$$\{q_a(\xi),p_b(\xi')\}=\delta_{ab}\delta(\xi-\xi'),$$
$$\{q_a(\xi),q_b(\xi')\}=\{p_a(\xi),p_b(\xi')\}=0. \tag{2.10}$$

We then introduce the primary constraints

$$p_a(\lambda)=g_a(0,\lambda)+\int_{\mathbb{R}}d\xi\,[f_a(\lambda-\xi,\xi)-\partial_\lambda g_a(\lambda-\xi,\xi)]\Theta(\lambda-\xi,\xi), \tag{2.11}$$

with

$$\Theta(u,v)=Y(v)Y(u)-Y(-v)Y(u)=\tfrac{1}{2}[\varepsilon(v)-\varepsilon(u)], \tag{2.12}$$

where $Y(v)$ denotes the Heaviside function and $\varepsilon(v)$ is the "sign function." In Ref. 9, it was found that the secondary constraints that follow from (2.11) and the Hamiltonian (2.8) are the functional equations (2.6). Therefore, the evolution space E can be immersed into the phase space Γ by defining

$$\psi\colon E\to\Gamma,$$
$$z=q_a(\lambda_a)\to\psi(z)=(q_a(\lambda_a),p_b(\lambda_b)), \tag{2.13}$$

where each curve $p_b(\lambda_b)$ is given as a function of $z=(q_a(\lambda_a))$ by Eq. (2.11). This mapping is invariant under "time" t evolution. Then, the symplectic form defined on Γ by the Poisson brackets (2.10) can be pulled back onto E. In this way, an equivalence between the Lagrangian formalism (2.6) and (2.7) and the Hamiltonian one (1.5)–(1.9) is established.

III. PRESYMPLECTIC FORM ON E

If we apply the above results to Fokker Lagrangians we obtain

$$f_{a\mu}(\lambda,\xi)=-\tfrac{1}{2}\sum_{b\neq a}g_ag_b\times\int_{\mathbb{R}}d\eta\,\frac{\partial\omega^r_{ba}(\lambda+\eta,\lambda+\xi)}{\partial x^\mu_a(\lambda+\xi)}[\delta(\xi)+\delta(\eta)] \tag{3.1}$$

and

$$g_{a\mu}(\lambda,\xi)=\delta(\xi)m_a(-(\dot x^2_a(\lambda))^{-1/2}\dot x_{a\mu}(\lambda)-\frac{1}{2}\sum_{\substack{b=1\\b\neq a}}^{N}g_ag_b\int_{\mathbb{R}}d\eta\times\frac{\partial\omega^r_{ba}(\lambda+\eta,\lambda+\xi)}{\partial\dot x^\mu_a(\lambda+\xi)}[\delta(\xi)+\delta(\eta)]. \tag{3.2}$$

As a consequence, the Euler equations take the form

$$\frac{d}{d\lambda}(m_a\dot x_{a\mu}(\lambda)(-\dot x^2_a(\lambda))^{-1/2})=\sum_{\substack{b=1\\b\neq a}}^{N}g_ag_b\int_{\mathbb{R}}d\eta\,\hat{\mathscr{L}}_{a\mu}\omega^r_{ab}(\lambda-\eta,\lambda), \tag{3.3}$$

with the Lagrange operators $\hat{\mathscr{L}}_{a\mu}$ defined by

$$\hat{\mathscr{L}}_{a\mu}=\frac{\partial}{\partial x^\mu_a(\lambda)}-\frac{d}{d\lambda}\frac{\partial}{\partial\dot x^\mu_a(\lambda)}. \tag{3.4}$$

The primary constraints (2.11) can be written as

$$p_{a\mu}(\lambda)=\delta(\lambda)\pi_{a\mu}+\eta_{a\mu}(\lambda), \tag{3.5}$$

where

$$\pi_{a\mu}=m_a(-\dot x^2_a(0))^{-1/2}\dot x_{a\mu}(0)-\sum_{\substack{b=1\\b\neq a}}^{N}g_ag_b\int_{\mathbb{R}}d\eta\,\frac{\partial\omega^r_{ab}(\eta,0)}{\partial\dot x^\mu_a(0)} \tag{3.6a}$$

and

$$\eta_{a\mu}(\lambda)=-\frac{1}{2}\sum_{\substack{b=1\\b\neq a}}^{N}g_ag_b\times\int_{\mathbb{R}}d\eta\,\hat{\mathscr{L}}_{a\mu}\{\omega^r_{ab}(\lambda-\eta,\lambda)\}\Theta(\lambda-\eta,\lambda). \tag{3.6b}$$

Notice that the term $\pi_{a\mu}$ looks like the expression obtained for mechanical momenta in an analogous field description of the system.

The phase space T^*E was endowed with a Liouville form $\Xi\in\Lambda^1(T^*E)$, and the corresponding symplectic structure[9]

$$\Omega=-d\Xi. \tag{3.7}$$

The pullback map ψ^* takes $\Xi\in\Lambda^1(T^*E)$ onto $\psi^*\Xi\in\Lambda^1(E)$; thus yielding

$$\psi^*\Xi=\sum_{a=1}^{N}\int_{\mathbb{R}}d\lambda\,p^\mu_a(\lambda,[x_a(\lambda)])\Delta x_{a\mu}(\lambda). \tag{3.8}$$

Here the symbol Δ denotes the exterior differential in the infinite-dimensional manifold E, and it is used in order to avoid confusion with the symbol $d\lambda$ under the integral sign. Substituting Eq. (2.5) into (2.8) we directly obtain

$$\psi^*\Xi = \sum_{a=1}^{N} \pi_a^\mu \Delta x_{a\mu}(0) + \sum_{a=1}^{N} \int_{\mathbb{R}} d\lambda\, \eta_a^\mu(\lambda) \Delta x_{a\mu}(\lambda). \tag{3.9}$$

Notice the simplicity of the first term compared with the second one, which involves an integral over the region of $\mathbb{R}^2$ having $\Theta(\lambda - \eta,\lambda)$ as its characteristic function.

The expressions for the generating functions of the Poincaré group can be directly obtained, too, as we already discussed in Ref. 9. Thus the total linear momentum is given by the generating functions of space-time translations:

$$P_\mu = \sum_{a=1}^{N} \int_{\mathbb{R}} d\lambda\, p_{a\mu}(\lambda) = \sum_{a=1}^{N} \pi_{a\mu} + \sum_{a=1}^{N} \int_{\mathbb{R}} d\lambda\, \eta_{a\mu}(\lambda). \tag{3.10a}$$

While the generating functions for Lorentz transformations yield the total angular momentum[9]

$$\begin{aligned} J_{\mu\nu} &= \sum_{a=1}^{N} \int_{\mathbb{R}} d\lambda \{x_{a\mu}(\lambda) p_{a\nu}(\lambda) - x_{a\nu}(\lambda) p_{a\mu}(\lambda)\} \\ &= \sum_{a=1}^{N} \{x_{a\mu}(0)\pi_{a\nu} - x_{a\nu}(0)\pi_{a\mu}\} \\ &\quad + \sum_{a=1}^{N} \int_{\mathbb{R}} d\lambda \{x_{a\mu}(\lambda)\eta_{a\nu}(\lambda) - x_{a\nu}(\lambda)\eta_{a\mu}(\lambda)\}. \end{aligned} \tag{3.10b}$$

These results are in agreement with earlier calculations obtained, either by a generalization of the Noether theorem[7] or by the method of Dettman and Schild,[20] provided that

$$\lim_{\lambda\to\infty} \omega^r_{ab}(\lambda,\eta) = 0. \tag{3.11}$$

IV. THE PRESYMPLECTIC STRUCTURE ON TM_4^N

Let us consider a set of solutions of Eqs. (3.3),

$$x_a^\mu(\lambda) = \varphi_a^\mu(\lambda, x_b, \dot{x}_c), \tag{4.1}$$

parametrized with the Newton-like initial data $x_b^\nu(0) = x_b^\nu$, $\dot{x}_c^\rho(0) = \dot{x}_c^\rho$.

When we introduce them in Eq. (3.9) we obtain

$$\begin{aligned} \rho &= j^*\Xi \\ &= \sum_{a=1}^{N} \pi_{a\mu}(x_b,\dot{x}_c)\Delta x_a^\mu \\ &\quad + \sum_{a=1}^{N} \int_{\mathbb{R}} d\lambda\, \eta_{a\mu}(\lambda, x_b, \dot{x}_c) \Delta\varphi_a^\mu(\lambda, x_b, \dot{x}_c), \end{aligned} \tag{4.2}$$

where $j^* = \varphi^* \circ \psi^*$.[9] Moreover, $\pi_{a\mu}(x_b,\dot{x}_c)$ and $\eta_{a\mu}(x_b,\dot{x}_c)$ correspond to $\pi_{a\mu}$ and $\eta_{a\mu}$ evaluated by introducing (4.1) into (3.5), and use has been made of the fact that

$$j^*(\Delta x_a^\mu(\lambda)) = \Delta\varphi_a^\mu(\lambda, x_b, \dot{x}_c) \equiv \sum_{b=1}^{N} \frac{\partial \varphi_a^\mu}{\partial x_b^\nu} dx_b^\nu + \sum_{b=1}^{N} \frac{\partial \varphi_a^\mu}{\partial \dot{x}_b^\nu} d\dot{x}_b^\nu$$

and

$$\varphi_a^\mu(0, x_b, \dot{x}_c) = x_a^\mu. \tag{4.3}$$

In this way, we obtain a formal expression for the Liouville form on TM_4^N. An analogous treatment can be given to the total linear and angular momenta. Now, actual calculations require explicit expressions for the trajectories φ_a^μ, $a = 1,\ldots,N$. As we have already mentioned in the Introduction, these can be dealt with through perturbative techniques, which can be found in the literature.[19] These techniques use the analytical dependence on the physical parameters that we have assumed for the solutions φ_a^μ. Then, the corresponding MacLaurin series on these parameters are worked out and a perturbative expansion is obtained. In the present paper, we mainly work with coupling constant expansions. Let us briefly describe the associated perturbation theory.

Consider the expansion of $x_{a\mu}$ in powers of the coupling constants g_a:

$$x_{a\mu}(\lambda) = \sum_{n_1,\ldots,n_N=0}^{\infty} g_1^{n_1}\cdots g_N^{n_N} \varphi_{a\mu}^{(n_1,\ldots,n_N)}(\lambda; x_b, \dot{x}_c). \tag{4.4}$$

When (4.4) is substituted into the equations of motion (3.3), a hierarchy of second-order differential systems is obtained, such that each step determines the right-hand side of the next one. The condition that in the limit $g_a \to 0$ we have free motion is then introduced by taking

$$\varphi_{a\mu}^{(0)}(\lambda; x_b, \dot{x}_c) = x_{a\mu} + \lambda \dot{x}_{a\mu} \tag{4.5}$$

and the coefficients in the series (4.4) are uniquely determined by the Newton-like set of initial data $(x_b, \dot{x}_c)$ at $\lambda = 0$.

A similar treatment could be used for the problem if the criteria (b) or (c) discussed in Sec. I were chosen, namely, $1/c^2$ (see Refs. 11, 15, and 16) and m/M (see Refs. 17 and 18) expansions, respectively.

Introducing these kinds of expansions in (4.2) we obtain a perturbative expansion for ρ. Furthermore, since the lowest-order term in the expansion for the Lagrange operator $\hat{\mathscr{L}}_{a\mu}$ is given by

$$\hat{\mathscr{L}}_{a\mu}^{(0)} = \frac{\partial}{\partial x^{a\mu}} - \frac{\partial}{\partial \lambda}\left(\frac{\partial}{\partial \dot{x}^{a\mu}} - \lambda \frac{\partial}{\partial x^{a\mu}}\right), \tag{4.6}$$

the order of computing integrals and derivatives in Eq. (2.13) can be reversed.

We also have that

$$\Delta\varphi_{a\mu}^{(0)}(\lambda; x_b, \dot{x}_c) = dx_{a\mu} + \lambda d\dot{x}_{a\mu}. \tag{4.7}$$

And, after some rearrangements, it is found that, to the order $n = 2$,

$$\rho = \sum_{a=1}^{N} (R_{a\mu}^{(2)} dx_a^\mu + Q_{a\mu}^{(2)} d\dot{x}_a^\mu) + O(g^4), \tag{4.8}$$

where the term $O(g^n)$ includes all terms occurring multiplied by at least n "charges" g_b. Moreover,

$$\begin{aligned} R_{a\mu}^{(2)} &= \pi_{a\mu}^{(2)}(x,\dot{x}) \\ &\quad - \frac{1}{2} \sum_{b\neq a} g_a g_b \left[F_{ba}(\dot{x}_b,\dot{x}_a) \frac{\partial}{\partial x_a^\mu}(N_{ba}^0 + M_{ba}^1) \right. \\ &\quad \left. - \frac{\partial}{\partial \dot{x}_a^\mu}(F_{ba}(\dot{x}_b,\dot{x}_a) M_{ba}^0) \right], \end{aligned} \tag{4.9a}$$

$$Q^{(2)}_{a\mu} = -\frac{1}{2}\sum_{b\neq a} g_a g_b \left[F_{ba}(\dot{x}_b,\dot{x}_a)\frac{\partial M^2_{ba}}{\partial x^\mu_a} + \frac{\partial}{\partial \dot{x}^\mu_a}\left[(N^0_{ba} - M^1_{ba})F_{ba}(\dot{x}_b,\dot{x}_a)\right]\right], \tag{4.9b}$$

$F_{ba}(\dot{x}_b,\dot{x}_a)$ is given by (2.3), and

$$N^s_{ba} = \int_{\mathbb{R}^2} d\lambda\, d\eta\, \lambda^s \Theta(\lambda-\eta,\lambda) G^{(0)}_{ba}(\lambda-\eta,\lambda), \tag{4.10a}$$

$$M^s_{ba} = \int_{\mathbb{R}^2} d\lambda\, d\eta\, \partial_\lambda [\lambda^s \Theta(\lambda-\eta,\lambda) G^{(0)}_{ba}(\lambda-\eta,\lambda)], \tag{4.10b}$$

where, according to (4.5),

$$G^{(0)}_{ba}(\lambda_b,\lambda_a) = G_{ba}(x_b - x_a + \dot{x}_b\lambda_b - \dot{x}_a\lambda_a).$$

Similarly, to this order of approximation, the total linear momentum is given by

$$P_\mu = \sum_{a=1}^{N} R^{(2)}_{a\mu} + O(g^4), \tag{4.11}$$

while the total angular momentum is

$$J_{\mu\nu} = \sum_{a=1}^{N} [R^{(2)}_{a\mu} x_{a\nu} + Q^{(2)}_{a\mu}\dot{x}_{a\nu} - R^{(2)}_{a\nu} x_{a\mu} - Q^{(2)}_{a\nu}\dot{x}_{a\mu}] + O(g^4). \tag{4.12}$$

Hitherto, we have been working in the covariant formalism. The price we have had to pay for this is the singular character of the lowest-order term in the Lagrangian. As a consequence, the lowest-order contribution to the presymplectic form on TM^N_4 is not regular. In other words, the zeroth-order terms in the momenta (3.6a), $\pi^{(0)}_{a\mu}(x_b,\dot{x}_c)$, fulfill the N constraints

$$\pi^{(0)\mu}_a \pi^{(0)}_{a\mu}(x_b,\dot{x}_c) = -m^2_a.$$

In order to avoid the problems associated with this singularity, we shall go into the noncovariant formalism by fixing the time coordinates x^0_a and the evolution parameter t according to

$$x^0_a = t. \tag{4.13}$$

Given any quantity A in the covariant formalism, we shall denote by $\tilde{A}$ its noncovariant counterpart, that is, the quantity resulting from introducing the constraints (4.13) into it.]

A direct calculation shows that

$$\tilde{\rho} = H\,dt - \sum_{a=1}^{N} \vec{p}_a d\vec{q}_a, \tag{4.14}$$

where

$$H = -\sum_{a=1}^{N} \tilde{R}^{(2)}_{a0} + O(g^4), \tag{4.15}$$

$$\vec{q}_a = \vec{x}_a - \frac{1}{m_a\gamma_a}[\vec{\tilde{Q}}^{(2)}_a - (\vec{\tilde{Q}}^{(2)}_a\vec{v}_a)\vec{v}_a] + O(g^4), \tag{4.16}$$

and

$$\vec{p}_a = \vec{\tilde{R}}^{(2)}_a + O(g^4). \tag{4.17}$$

The one-form $\tilde{\rho}$ obtained by restriction of (4.8) to the phase space defined by the constraints (4.13) (that is, the phase space of the noncovariant formalism) plays the role of a Poincaré–Cartan integral invariant.[21] According to this H will be the Hamiltonian, i.e., the generating function for time evolution, and $\vec{q}_a$ and $\vec{p}_a$ will be a set of canonical coordinates and momenta. (As usual, an arrow over a symbol indicates the space part of the corresponding four-vector.)

The evaluation of the quantities $\vec{R}^{(2)}_a$ and $\vec{Q}^{(2)}_a$ involves differentiation with respect to x^μ_a and $\dot{x}^\mu_a$. (Some useful formulas are given in the Appendix.)

V. SCALAR AND VECTOR INTERACTIONS. WHEELER–FEYNMAN PREDICTIVE ELECTRODYNAMICS

A. Scalar and vector interactions

The action-at-a-distance counterpart of the field theory for scalar particles interacting through a massive scalar (resp. vector) field is given by the Lagrangian (2.1) with $r=0$ (resp. $r=1$) and

$$G_{ab}((x_a(t_a) - x_b(t_b))^2) = D_{\mathrm{sym}}(\mu,(x_a(t_a) - x_b(t_b))^2), \tag{5.1}$$

where

$$D_{\mathrm{sym}}(\mu,s) = \tfrac{1}{2}[D_{\mathrm{Adv}}(\mu,s) + D_{\mathrm{ret}}(\mu,s)] = 2[\delta(s) - Y(-s)(\mu/2\sqrt{-s})J_1(-\mu\sqrt{-s})] \tag{5.2}$$

is the time symmetric Green's function for the massive Klein–Gordon operator.

In order to obtain a presymplectic structure on TM^N_4, up to the second order in the coupling constants g_a, it is necessary to give explicit expressions for the results (4.10). First of all, we notice that, up to this order,

$$F_{ab}(\dot{x}_a,\dot{x}_b) = (-\dot{x}_a\dot{x}_b)^r(-\dot{x}_a\dot{x}_a)^{(1-r)/2} \times(-\dot{x}_b\dot{x}_b)^{(1-r)/2}, \tag{5.3}$$

and therefore it does not depend on λ.

Substituting (5.1) into Eq. (4.10b), we also have

$$M^s_{ab} = 2\int d\lambda\, d\eta\, \partial_\lambda[\lambda^s\Theta(\lambda-\eta,\lambda)D_{\mathrm{sym}}(\mu,f_{ab}(\lambda,\eta))], \tag{5.4}$$

$$f_{ab}(\lambda,\eta) = [(x_a - x_b) + (\dot{x}_a - \dot{x}_b)\lambda - \dot{x}_b\eta]^2 \equiv [x_{ab} + \dot{x}_{ab}\lambda - \dot{x}_b\eta]^2. \tag{5.5}$$

Then, taking into account that

$$J_1(x) \underset{x\to\infty}{\to} 1/\sqrt{x} \to 0,$$

we obtain

$$M^s_{ab} = 0, \quad s = 0,1,2. \tag{5.6}$$

Consequently, Eqs. (4.10) take the simpler form

$$R^{(2)}_{a\mu} = \pi^{(2)}_{a\mu} - \frac{1}{2}\sum_{b\neq a} g_a g_b F_{ba}(\dot{x}_b,\dot{x}_a)\frac{\partial}{\partial x^\mu_a}N^0_{ba}, \tag{5.7a}$$

$$Q^{(2)}_{a\mu} = -\frac{1}{2}\sum_{b\neq a} g_a g_b \frac{\partial}{\partial \dot{x}^\mu_a}(N^0_{ba}F_{ba}(\dot{x}_b,\dot{x}_a)). \tag{5.7b}$$

A direct calculation yields

$$\pi_{a\mu}(x_c,\dot{x}_d) = m_a(-\dot{x}_a^2)^{-1/2}\dot{x}_{a\mu} - \sum_{b\neq a} g_a g_b \frac{\partial F_{ba}(\dot{x}_b,\dot{x}_a)}{\partial \dot{x}_a^\mu} \times (1/r_{ab}\sqrt{-\dot{x}_b^2})\exp(-\mu r_{ab}), \quad (5.8a)$$

$$r_{ab}^2 = x_{ab}^2 - (\dot{x}_b x_{ab})^2/\dot{x}_b^2. \quad (5.8b)$$

On the other hand, since the presence of the characteristic function $\Theta(\lambda-\eta,\lambda)$ makes the evaluation of N_{ba}^0 rather involved, it is more convenient to work it out in terms of its Fourier representation.

Taking into account that

$$D_{sym}(\mu,x^\mu x_\mu) = 4\pi\,\text{P.V.}\int_{\mathbb{R}^4} \frac{d^4k}{(2\pi)^4}\frac{\exp(ikx)}{k^2+\mu^2}, \quad (5.9)$$

we obtain

$$N_{ab}^0 = I_{ab} - I_{ba}, \quad (5.10a)$$

where

$$I_{ab} = 4\pi i\,\text{P.V.}\int_{\mathbb{R}^4}\frac{d^4k}{(2\pi)^3}\frac{\exp(ikx_{ab})}{k^2+\mu^2}\frac{\delta(k\dot{x}_b)}{k\dot{x}_a}, \quad (5.10b)$$

or, after a little calculation,

$$I_{ab} = \frac{1}{\Lambda_{ab}}\int_{\mu_{ab}}^1 d\eta\,\frac{\exp(-r_{ab}\mu\eta/(-\dot{x}_b^2))}{\sqrt{\eta^2-u_{ab}^2}}, \quad (5.11)$$

where

$$u_{ab} = \left(1-\left(\frac{\Lambda_{ab}z_{ab}}{r_{ab}}\right)^2\frac{1}{(-\dot{x}_{ab}^2)}\right)^{1/2}, \quad (5.12)$$

and the following Poincaré invariants have been introduced:

$$z_{ab} \equiv \Lambda_{ab}^{-2}[-\dot{x}_b^2(x_{ab}\dot{x}_a) + \dot{x}_a\dot{x}_b(x_{ab}\dot{x}_b)] \quad (5.13)$$

and

$$\Lambda_{ab}^2 = k_{ab}^2 - \dot{x}_a^2\dot{x}_b^2, \quad k_{ab} = -\dot{x}_a\dot{x}_b. \quad (5.14)$$

A similar result holds for I_{ba}.

Notice that, since I_{ab} is Poincaré invariant, the calculations leading from (5.10) to (5.11) can be worked out in the rest frame of particle b, followed by the subsequent translation into covariant form.

The Liouville form (4.9), as well as the total linear and angular momenta, are determined to this order by $R_{a\mu}^{(2)}$ and $Q_{a\mu}^{(2)}$. Their explicit expressions can be found by direct substitution of (5.3), (5.8), and (5.10) into (5.7). The three-dimensional formalism can also be developed by merely introducing the constraints (4.13).

B. Wheeler–Feynman electrodynamics

There is no doubt that the most widely known Fokker-type system is Wheeler–Feynman electrodynamics (WFE). First introduced by Tetrode[22] and Fokker[4] for two charges, its generalization to N charges,[5] supplemented by the perfect absorber theory,[8] describes the whole classical electrodynamics, including the radiation reaction effects and the observed retarded interaction.

WFE corresponds to an action-at-a-distance vector interaction with vanishing mass parameter (i.e., $\mu=0$). Hence all the results hitherto obtained still hold.

Taking into account that

$$\int\frac{d\eta}{\sqrt{A\eta^2+C}} = \frac{1}{\sqrt{A}}\ln(\sqrt{A}\eta+\sqrt{A\eta^2+C}), \quad A>0, \quad (5.15)$$

we can express N_{ab}^0 in terms of elementary functions; thus obtaining

$$N_{ab}^0 = \frac{1}{2\Lambda_{ab}}\ln\left(\frac{r_{ab}+\Lambda_{ab}z_{ab}(-\dot{x}_b^2)^{-1/2}}{r_{ba}+\Lambda_{ab}z_{ba}(-\dot{x}_a^2)^{-1/2}}\right),$$

which, when introduced into (5.7), yields

$$R_{a\mu}^{(2)} = m_a\frac{\dot{x}_{a\mu}}{\sqrt{-\dot{x}_a^2}} + \sum_{b\neq a} g_a g_b\left(\frac{\dot{x}_{a\mu}}{r_{ab}\sqrt{-\dot{x}_b^2}} + \frac{1}{2}\frac{k_{ab}}{2\Lambda_{ab}}\frac{\partial}{\partial \dot{x}_a^\mu}\ln\left(\frac{r_{ab}+\Lambda_{ab}z_{ab}(-\dot{x}_b^2)^{-1/2}}{r_{ba}+\Lambda_{ab}z_{ba}(-\dot{x}_a^2)^{-1/2}}\right)\right), \quad (5.16a)$$

$$Q_{a\mu}^{(2)} = \frac{1}{2}\sum_{b\neq a} g_a g_b \times\frac{\partial}{\partial \dot{x}_a^\mu}\left[\frac{k_{ab}}{2\Lambda_{ab}}\ln\left(\frac{r_{ab}+\Lambda_{ab}z_{ab}(-\dot{x}_b^2)^{-1/2}}{r_{ba}+\Lambda_{ab}z_{ba}(-\dot{x}_a^2)^{-1/2}}\right)\right]. \quad (5.16b)$$

By substituting these expressions into (4.15)–(4.17) we easily obtain the Hamiltonian and a set of canonical coordinates and momenta.

Furthermore, in order to compare with other results already known, the post-Newtonian approximation can be carried out. Approximating the canonical coordinates and momenta up to order c^{-2}, we obtain

$$\vec{q}_a = \vec{x}_a - c^{-2}\sum_{b\neq a} g_a g_b\frac{1}{4m_a}\frac{\partial}{\partial \vec{v}_a}\left(\frac{w_{ab}}{|\vec{x}_{ab}|}\right) + O(c^{-4}), \quad (5.17)$$

$$\vec{p}_a = m_a\vec{v}_a + c^{-2}\left(\frac{1}{2}m_a v_a^2\vec{v}_a + \sum_{b\neq a} g_a g_b\frac{\vec{v}_b}{|\vec{x}_{ab}|} - \sum_{b\neq a} g_a g_b\frac{\partial}{\partial \vec{x}_a}\left(\frac{w_{ab}-w_{ba}}{|\vec{x}_{ab}|}\right)\right) + O(c^{-4}), \quad (5.18)$$

where $w_{ab} = \vec{x}_{ab}\vec{v}_a$.

The simple form of the coordinates, up to this order of approximation, suggests the possibility of removing the c^{-2} term by a canonical transformation with generating function,

$$U(q,\tilde{p}) = \sum_a \vec{q}_a\tilde{\vec{p}}_a + U_2(q,\tilde{p}) + O(c^{-4}), \quad (5.19)$$

$$U_2(q,\tilde{p}) = c^{-2}\sum_{\substack{a,b=1\\ a\neq b}}^N \frac{\vec{q}_{ab}\vec{v}_a}{4|\vec{q}_{ab}|} + O(c^{-4}), \quad (5.20)$$

[where $\vec{v}_a = \tilde{\vec{p}}_a/m_a + O(c^{-2})$ must be understood in the U_2 term], which leads to a new set of canonical coordinates and momenta, $\tilde{\vec{q}}_a$ and $\tilde{\vec{p}}_a$, that can be obtained by the well known expressions[21]

$$\tilde{\vec{q}}_a = \frac{\partial U(q,\tilde{p})}{\partial \tilde{\vec{p}}_a}, \quad \vec{p}_a = \frac{\partial U(q,\tilde{p})}{\partial \vec{q}_a}.$$

In this case

$$\tilde{\vec{q}}_a = \vec{x}_a + O(c^{-4}), \tag{5.21a}$$

$$\tilde{\vec{p}}_a = m_a\vec{v}_a + c^{-2}\left(\frac{1}{2} m_a v_a^2 \vec{v}_a + \sum_{b\neq a} g_a g_b \frac{\vec{v}_b}{|\vec{x}_{ab}|} + \sum_{b\neq a} g_a g_b \frac{1}{2}\frac{\partial}{\partial \vec{x}_a}\left(\frac{w_{ba}}{|\vec{x}_{ab}|}\right)\right) + O(c^{-4}). \tag{5.21b}$$

Hence, up to order c^{-2}, physical positions of particles can be taken as canonical coordinates. This result agrees with those given in Refs. 23 and 24.

In order to derive the order c^{-4} terms for the canonical variables $\tilde{\vec{q}}_a$ and $\tilde{\vec{p}}_a$, correction terms must be added to Eqs. (5.17) and (5.18). Naming these correction terms $\vec{q}_{4a}$ and $\vec{p}_{4a}$, respectively, we have

$$\vec{q}_{4a} = c^{-4}\frac{1}{m_a}\left[\sum_{b\neq a} g_b g_a \frac{1}{1bm_b}\frac{\partial}{\partial \vec{v}_a}(\mathscr{M}_{ab} - \mathscr{M}_{ba}) - \frac{v_a^2}{2}\sum_{b\neq a}\frac{g_a g_b}{m_a}\frac{\partial}{\partial \vec{v}_a}\left(\frac{w_{ab}}{|\vec{x}_{ab}|}\right) - \sum_{b\neq a}\frac{g_a g_b}{m_a}\left(\left(\frac{\partial}{\partial \vec{v}_a}\frac{w_{ab}}{|\vec{x}_{ab}|}\right)\vec{v}_a\right)\vec{v}_a\right], \tag{5.22a}$$

$$\vec{p}_{4a} = c^{-4}\left[\frac{3}{8} m_a v_a^4 \vec{v}_a + \sum_{b\neq a} g_a g_b \frac{1}{2|\vec{x}_{ab}|}\left(v_a^2 - \frac{w_{ba}^2}{|\vec{x}_{ab}|^2}\right) - \sum_{b\neq a} g_a g_b \frac{1}{16}\frac{\partial}{\partial \vec{x}_a}(\mathscr{M}_{ab} - \mathscr{M}_{ba})\right], \tag{5.22b}$$

where

$$\mathscr{M}_{ab}(\vec{x},\vec{v}) = \frac{1}{|x_{ab}|^3}[(\vec{x}_{ab}\vec{v}_b)^3 + (\vec{x}_{ab}\vec{v}_a)(\vec{x}_{ab}\vec{v}_b)^2] + \frac{1}{|\vec{x}_{ab}|}[(\vec{x}_{ab}\vec{v}_b)(3\vec{v}_b^2 - 2\vec{v}_a\vec{v}_b) + (\vec{x}_{ab}\vec{v}_a)\vec{v}_b^2]. \tag{5.23}$$

In order to compare these results with those given in Ref. 25, we carry out a canonical transformation with generating function

$$U(q,\tilde{p}) = \sum_a \vec{q}_a\tilde{\vec{p}}_a + U_2(q,\tilde{p}) + U_4(q,\tilde{p}) + O(c^{-4}), \tag{5.24}$$

with

$$U_2(q,\tilde{p}) = c^{-2}\sum_{a,b} g_a g_b \frac{\vec{q}_{ab}\vec{v}_a}{4|\vec{q}_{ab}|} + O(c^{-6}),$$

$$U_4(q,\tilde{p}) = c^{-4}\sum_{a,b} g_a g_b \frac{1}{16 m_a}\mathscr{M}_{ab}(\vec{q},\vec{v}) + O(c^{-6}),$$

and

$$\vec{v}_a = \frac{\tilde{\vec{p}}_a}{m_a} - c^{-2}\left[\frac{\tilde{\vec{p}}_a^2}{2m_a^2}\tilde{\vec{p}}_a + \sum_{b\neq a} g_a g_b \frac{\tilde{\vec{p}}_b}{m_b|\vec{q}_{ab}|} + \sum_{b\neq a}\frac{g_a g_b}{2m_b}\frac{\partial}{\partial \vec{q}_a}\left(\frac{\vec{q}_{ab}\tilde{\vec{p}}_b}{|\vec{q}_{ab}|}\right)\right] + O(c^{-6}).$$

The final expressions for the canonical variables and momenta $\tilde{\vec{q}}_a$ and $\tilde{\vec{p}}_a$ up to order c^{-4} coincide with the above-mentioned results.[25]

VI. CONCLUSIONS AND OUTLOOK

Let us briefly sketch the present state of affairs in those relativistic theories of directly interacting particles that are intended to relate the force acting on one particle to some classical field theory. This will help to understand the role claimed for the present paper and also for Ref. 9.

These theories start from a *Fokker-type Lagrangian* (A), which is nonlocal in time, and then the Euler equations are derived. The latter is a system of *functional differential equations* (B), which is *nonpredictive* in the Newtonian sense, because the particle position and velocities at a given time do not determine their future evolution. This seems to be a strong qualitative difference with regard to what is common in nonrelativistic physics. A way out is to accept that not all the solutions of (B) are "physically significant," but only those satisfying an additional requirement, namely, (C) the *analytical dependence* of particle world lines on some "small parameter" (either the inverse speed of light, the coupling constants, or the mass ratio).

The use of any of these criteria enables one to select a family of "physically significant" solutions of (B). Each of them can be parametrized by the particle positions and velocities at a given time, and hence satisfy a Newton-like set of equations of motion (D).

The path from (B) to (D) via (C) is conceptually simple when the inverse speed of light is taken as a "small parameter" (it can be carried over by merely iterating the substitution of lower orders of approximation into higher ones), and involves the use of more complex perturbative techniques, besides the theoretical framework of predictive relativistic mechanics, in the remaining two cases.

Finally, introducing some additional, although rather general, assumptions concerning the asymptotic conditions in the past and/or future infinity, a Hamiltonian formalism (E) can be set up. Apparently this canonical formalism has a relationship with the Fokker-type variational principle one has started from.

The main contribution of the present paper, and Ref. 9, too, consists of providing a Hamiltonian (presymplectic) formalism for Fokker-type Lagrangian systems. This enables us to introduce the *analyticity condition* directly in the Hamiltonian formalism, so that we can obtain a Hamiltonian formalism on a Newton-like phase space for the system of interacting particles. In this way, we abridge the path from (A) to (E), without losing the connection with the canonical formalism that seemed to underlie the Fokker variational principle.

We have finally specialized these general results to scalar and vector mesodynamics,[6] and to Wheeler–Feynman symmetric electrodynamics, up to the first order in the coupling constants. Although the formalism has here been developed for a specific kind of theory, namely, relativisitc systems of directly interacting particles, its interest goes beyond this topic. Indeed, it could also be applied to nonlocal field theories (presumably with only a few extra technicalities not interfering with the core difficulty of a Lagrangian that is nonlocal in time). The canonical formalism so obtained would then allow one to proceed with a standard canonical quantization of nonlocal field theories and to add some new insight to other quantization procedures.[26,27]

ACKNOWLEDGMENTS

This work has been partially supported by CAICyT under Contract No. 0649-84. One of us (R. J.) wants to acknowledge the hospitality of the Universitat de Barcelona as well as the partial support of CONACYT (Mexico) and Ministerio de Educación y Ciencia, ICI (España).

APPENDIX: FROM COVARIANT TO NONCOVARIANT NOTATION

Here we give some formulas that are useful in translating expressions from covariant to noncovariant formalism, i.e., when the time fixation $x_a^0 = t$ is introduced.

These formulas establish in what cases the operations (i) introducing the time fixations and (ii) differentiation commute or do not commute.

In general, we must deal with a Poincaré-invariant function $f(x_{cd}^2, x_{cd}\dot{x}_e, \dot{x}_c\dot{x}_d)$, $c,d,e = 1,\ldots,N$. Introducing the fixation, we have

$$\tilde{f}(x_{cd}^2, x_{cd}\dot{x}_e, \dot{x}_c\dot{x}_d) = f(|\vec{x}_{cd}|^2, \vec{x}_{cd}\vec{v}_e, -1 + \vec{v}_c\vec{v}_d),$$

where the notation given in Sec. IV has been used.

A short calculation proves that

$$\frac{\partial f}{\partial \vec{x}_a} = \sum_{c\neq a} 2\vec{x}_{ac}\frac{\partial \tilde{f}}{\partial |\vec{x}_{ac}|^2} + \sum_{e,c} \vec{v}_e \frac{\partial \tilde{f}}{\partial(\vec{x}_{ac}\vec{v}_e)} = \frac{\partial \tilde{f}}{\partial \vec{x}_a}, \quad \text{(A1)}$$

$$\frac{\partial f}{\partial x_a^0} = -\sum_{e,c} \frac{\partial \tilde{f}}{\partial(\vec{x}_{ac}\vec{v}_e)}, \quad \text{(A2)}$$

$$\frac{\partial f}{\partial \dot{\vec{x}}_a} = \sum_{c\neq d} \vec{x}_{cd}\frac{\partial \tilde{f}}{\partial(\vec{x}_{cd}\vec{v}_a)} + \sum_{c} \vec{v}_c \frac{\partial \tilde{f}}{\partial(\vec{v}_c\vec{v}_a)} = \frac{\partial \tilde{f}}{\partial \vec{v}_a}, \quad \text{(A3)}$$

$$\frac{\partial f}{\partial \dot{x}_a^0} = -\sum_{c} \frac{\partial \tilde{f}}{\partial(\vec{v}_c\vec{v}_a)}. \quad \text{(A4)}$$

[1] *The Theory of Action at a Distance in Relativistic Dynamics*, edited by E. H. Kerner (Gordon and Breach, New York, 1972); *Relativistic Action at a Distance*, edited by J. Llosa; *Lecture Notes in Physics*, Vol. 162 (Springer, New York, 1982); *Constraint Theory and Relativistic Dynamics*, edited by G. Longhi and L. Lusanna (World Scientific, Singapore, 1987).
[2] P. Droz-Vincent, Rep. Math. Phys. **8**, 79 (1975); F. Rohrlich, Lect. Notes Phys. **162**, 190 (1982); N. Mukunda and E. C. G. Sudarshan, Phys. Rev. D **23**, 2210 (1981); I. T. Todorov, Lect. Notes Phys. **162**, 213 (1982); A. Komar, Phys. Rev. D **18**, 1881 (1978); H. Sazdjian, Ann. Phys. (NY) **136**, 136 (1981); R. Arens, Nuovo Cimento B **21**, 395 (1974).
[3] L. Bel and J. Martin, Ann. Inst. H. Poincaré **22**, 173 (1975); D. Dominici, J. Gomis, J. A. Lobo, and J. M. Pons, Nuovo Cimento B **61**, 306 (1981).
[4] A. D. Fokker, Z. Phys. **58**, 386 (1929).
[5] R. P. Feynman and J. A. Wheeler, Rev. Mod. Phys. **21**, 425 (1949).
[6] P. Havas, Phys. Rev. **74**, 456 (1948); **86**, 974 (1952); H. Van Dam and E. P. Wigner, Phys. Rev. B **138**, 1576 (1965); **142**, 832 (1966); A. Katz, J. Math. Phys. **10**, 1929 (1969): G. J. H. Burgers and H. Van Dam, *ibid.* **28**, 677 (1987).
[7] W. N. Herman, J. Math. Phys. **26**, 2769 (1985).
[8] R. P. Feynman and J. A. Wheeler, Rev. Mod. Phys. **17**, 157 (1945).
[9] X. Jaén, R. Jaúregui, J. Llosa, and A. Molina, Phys. Rev. D **36**, 2385 (1987).
[10] H. J. Babha, Phys. Rev. **70**, 759 (1946).
[11] R. P. Gaida and V. I. Tretyak, Acta Phys. Pol. B **11**, 504 (1980).
[12] F. J. Kennedy, J. Math. Phys. **10**, 1349 (1969).
[13] R. Marnelius, Phys. Rev. D **10**, 2535 (1974).
[14] L. Bel and J. Martin, Ann. Inst. H. Poincaré **22**, 173 (1975).
[15] E. H. Kerner, J. Math. Phys. **3**, 35 (1962).
[16] H. W. Woodcock and P. Havas, Phys. Rev. D **6**, 3422 (1972).
[17] Synge, Proc. R. Soc. London **177**, 118 (1941).
[18] L. Bel and J. Martin, Phys. Rev. D **8**, 4347 (1973).
[19] L. Bel and X. Fustero, Ann. Inst. H. Poincaré **25 (a)**, 411 (1976); (a) L. Bel and Z. H. Sirousse, Phys. Rev. D. **32** 3128 (1985).
[20] J. W. Dettman and A. Schild, Phys. Rev. **95**, 1057 (1954).
[21] F. Gantmacher, *Lectures in Analytical Mechanics* (MIR, Moscow, 1970).
[22] H. Tetrode, Z. Phys. **10**, 317 (1922).
[23] J. Martin and J. L. Sanz, J. Math. Phys. **19**, 780 (1978).
[24] L. Bel and J. Martin, Ann. Inst. H. Poincaré **34**, 231 (1980).
[25] X. Jaén, J. Llosa, and A. Molina, Phys. Rev. D **34**, 2302 (1986).
[26] Ch. Hayashi, Prog. Theor. Phys. **10**, 533 (1953).
[27] H. Hata, Phys. Lett. B **217**, 438, 445 (1989).

Application of perturbation theory to the damped sextic oscillator

Sunita Srivastava and Vishwamittar
Department of Physics, Panjab University, Chandigarh-160014, India

(Received 22 March 1989; accepted for publication 14 June 1989)

Perturbation theory for the anharmonic oscillator with large damping has been used to solve the equation of motion for the damped sextic oscillator. The results so obtained are compared with the values found through numerical integration of the equation of motion.

I. INTRODUCTION

The successes met with by the addition of anharmonic terms to the harmonic potential in understanding many typical characteristics of different classical systems, thermal expansion of a one-dimensional crystal lattice, vibrational spectra of molecules, and the nature of interaction in various physical systems has aroused great interest in the theory of classical and quantum anharmonic oscillators.[1–9] The study of quantum anharmonic oscillators has also been spurred by the problems in quantum field theory.[5,10] Furthermore, since most of the physical systems are dissipative in nature, the study of the damped harmonic oscillator has also been in focus for quite some time.[11] However, the theory of the damped anharmonic oscillator, particularly the quantum theory, has not received the desired attention even though such a model should be closer to reality. Even if the problem is solved for a classical system and the position coordinates as a function of time are found, the results can be employed to carry out quantization using the path integral technique,[12] which has successfully been used to obtain the propagator for a time-dependent harmonic oscillator[13] and a damped harmonic oscillator,[14] in addition to many other systems.[12,15]

The present paper is the outcome of an effort to solve the equation of motion for the damped one-dimensional sextic anharmonic oscillator. The problem of the undamped sextic oscillator has been investigated by many workers[5,16–20] because this potential has been used in describing the vibrational spectra of some molecules,[21] the tricritical phenomena in ^{3}He–^{4}He mixtures, certain types of metamagnets, many multicomponent mixtures, and other systems.[22] The solution has been obtained using the perturbation method developed by Mendelson[23] for anharmonic oscillators subjected to large damping and takes into account the transient response. The equation of motion has also been solved numerically[24] and the two results have been found to be in good accord with each other.

II. SOLUTION TO THE EQUATION OF MOTION FOR THE DAMPED SEXTIC OSCILLATOR

The equation of motion for the linearly damped sextic anharmonic oscillator reads as

$$\ddot{x} + \lambda \dot{x} + \omega_0^2 x + a x^3 + b x^5 = 0\,. \tag{1}$$

For the system to be underdamped, it is assumed that $\lambda < 2\omega_0$. With a view to applying perturbation theory to this problem, we replace a by a small parameter ϵ and b by $\epsilon^2\beta$, so that the first-order contribution from the last term in (1) is of the same order as the second-order correction due to the term preceding it. Accordingly, Eq. (1) takes the form

$$\ddot{x} + \lambda \dot{x} + \omega_0^2 x + \epsilon x^3 + \epsilon^2 \beta x^5 = 0\,. \tag{2}$$

For $\epsilon = 0$, Eq. (2) represents the damped harmonic oscillator, whose solution is given by

$$x_0 = a \cos\phi\,, \tag{3}$$

where

$$a = a_0 \exp(-\lambda t/2) \tag{4}$$

and

$$\phi = \omega t + \phi_0\,, \tag{5}$$

with

$$\omega = (\omega_0^2 - \lambda^2/4)^{1/2}\,. \tag{6}$$

Following Mendelson,[23] we use the following second-order perturbation series expansions:

$$\begin{aligned} x &= x[a(t),\phi(t)] = x_0 + \epsilon x_1 + \epsilon^2 x_2 + \cdots\,, \\ \dot{a} &= \xi(a) = -(\lambda/2)a + \epsilon\xi_1 + \epsilon^2\xi_2 + \cdots\,, \end{aligned} \tag{7}$$

and

$$\dot{\phi} = \Omega(a) = \omega + \epsilon\omega_1 + \epsilon^2\omega_2 + \cdots\,.$$

Substituting expressions (7) in Eq. (2) and retaining terms up to ϵ^2, we obtain a system of parabolic partial differential equations corresponding to ϵ^0, ϵ^1, and ϵ^2. The equation pertaining to ϵ^0 is

$$\omega^2 \frac{\partial^2 x_0}{\partial\phi^2} + \frac{\lambda^2 a^2}{4}\frac{\partial^2 x_0}{\partial a^2} - \omega\lambda a \frac{\partial^2 x_0}{\partial\phi\,\partial a} + \omega\lambda\frac{\partial x_0}{\partial\phi} - \frac{\lambda^2 a}{4}\frac{\partial x_0}{\partial a} + \left(\omega^2 + \frac{\lambda^2}{4}\right)x_0 = 0\,, \tag{8}$$

where the solution is given by Eq. (3).

The terms containing ϵ lead to

$$\begin{aligned} &\omega^2 \frac{\partial^2 x_1}{\partial\phi^2} + \frac{\lambda^2 a^2}{4}\frac{\partial^2 x_1}{\partial a^2} - \omega\lambda a \frac{\partial^2 x_1}{\partial\phi\,\partial a} + \omega\lambda\frac{\partial x_1}{\partial\phi} - \frac{\lambda^2 a}{4}\frac{\partial x_1}{\partial a} + \left(\omega^2 + \frac{\lambda^2}{4}\right)x_1 \\ &= \left(2\omega\omega_1 a - \frac{\lambda\xi_1}{2} + \frac{\lambda a}{2}\frac{d\xi_1}{da} - \frac{3}{4}a^3\right)\cos\phi + \left(2\omega\xi_1 - \frac{\lambda a^2}{2}\frac{d\omega_1}{da}\right)\sin\phi - \frac{a^3}{4}\cos 3\phi\,. \end{aligned} \tag{9}$$

The complementary function for the corresponding homo-

 0022-2488/89/122815-04$02.50

geneous differential equation for x_1 is the same as that for Eq. (8) and the secular terms on the rhs disappear if

$$2\omega\omega_1 a - \frac{\lambda\xi_1}{2} + \frac{\lambda a}{2}\frac{d\xi_1}{da} - \frac{3}{4}a^3 = 0$$

and (10)

$$2\omega\xi_1 - \frac{\lambda a^2}{2}\frac{d\omega_1}{da} = 0 .$$

Writing ξ_1 and ω_1 as a power series in a, i.e.,

$$\xi_1 = \sum_{k=0}^{\infty} C_k^{(1)} a^k, \quad \omega_1 = \sum_{k=0}^{\infty} D_k^{(1)} a^{k-1}, \tag{11}$$

substituting in Eqs. (10), and equating the coefficients of various powers of a equal to zero, we obtain

$$C_3^{(1)} = 3\lambda/16\omega_0^2, \quad D_3^{(1)} = 3\omega/8\omega_0^2 , \tag{12}$$

while all other coefficients vanish. Thus

$$\xi_1 = C_3^{(1)} a^3, \quad \omega_1 = D_3^{(1)} a^2 . \tag{13}$$

The solution of Eq. (9) is found to be

$$x_1 = a^3 [A_3^{(3)} \cos 3\phi + B_3^{(3)} \sin 3\phi] , \tag{14}$$

where

$$A_3^{(3)} = (2\omega^2 - \lambda^2/4)/16\omega_0^2(4\omega^2 + \lambda^2/4)$$

and (15)

$$B_3^{(3)} = 3\omega\lambda/32\omega_0^2(4\omega^2 + \lambda^2/4) .$$

The expressions corresponding to ϵ^2 yield

$$\begin{aligned}
&\omega^2 \frac{\partial^2 x_2}{\partial\phi^2} + \frac{\lambda^2 a^2}{4}\frac{\partial^2 x_2}{\partial a^2} - \omega\lambda a \frac{\partial^2 x_2}{\partial\phi\,\partial a} + \omega\lambda\frac{\partial x_2}{\partial\phi} - \frac{\lambda^2 a}{4}\frac{\partial x_2}{\partial a} + \left(\omega^2 + \frac{\lambda^2}{4}\right)x_2 \\
&\quad = \left[2\omega\omega_2 a - \frac{\lambda\xi_2}{2} + \frac{\lambda a}{2}\frac{d\xi_2}{da} + \left(D_3^{(1)^2} - 3C_3^{(1)^2} - \frac{3}{4}A_3^{(3)} - \frac{5}{8}\beta\right)a^5\right]\cos\phi + \left[2\omega\xi_2 - \frac{\lambda}{2}a^2\frac{d\omega_2}{da}\right. \\
&\quad\quad \left. + \left(4C_3^{(1)}D_3^{(1)} - \frac{3}{4}B_3^{(3)}\right)a^5\right]\sin\phi + a^5\left[\left\{9(2\omega D_3^{(1)} + \lambda C_3^{(1)})A_3^{(3)} - \frac{3}{2}A_3^{(3)}\right.\right. \\
&\quad\quad \left. + 9(\lambda D_3^{(1)} - 2\omega C_3^{(1)})B_3^{(3)} - \frac{5}{16}\beta\right\}\cos 3\phi + \left\{9(2\omega C_3^{(1)} - \lambda D_3^{(1)})A_3^{(3)}\right. \\
&\quad\quad \left.\left. + 9(2\omega D_3^{(1)} + \lambda C_3^{(1)})B_3^{(3)} - \frac{3}{2}B_3^{(3)}\right\}\sin 3\phi - \left(\frac{3}{4}A_3^{(3)} + \frac{\beta}{16}\right)\cos 5\phi - \frac{3}{4}B_3^{(3)}\sin 5\phi\right] .
\end{aligned} \tag{16}$$

The secular terms in Eq. (16) are eliminated if

$$2\omega\omega_2 a - \frac{\lambda\xi_2}{2} + \frac{\lambda a}{2}\frac{d\xi_2}{da} + \left(D_3^{(1)^2} - 3C_3^{(1)^2} - \frac{3}{4}A_3^{(3)} - \frac{5}{8}\beta\right)a^5 = 0$$

and (17)

$$2\omega\xi_2 - \frac{\lambda a^2}{2}\frac{d\omega_2}{da} + \left(4C_3^{(1)}D_3^{(1)} - \frac{3}{4}B_3^{(3)}\right)a^5 = 0 .$$

Expressing ξ_2 and ω_2 as a power series in a, viz.

$$\xi_2 = \sum_{k=0}^{\infty} C_k^{(2)} a^k, \quad \omega_2 = \sum_{k=0}^{\infty} D_k^{(2)} a^{k-1} \tag{18}$$

and making the coefficients of different powers of a zero, we find that

$$\xi_2 = C_5^{(2)} a^5, \quad \omega_2 = D_5^{(2)} a^4 , \tag{19}$$

where

$$C_5^{(2)} = \{\lambda\,[3C_3^{(1)^2} - D^{(1)^2} + \tfrac{3}{4}A_3^{(3)} + \tfrac{5}{8}\beta] - \omega[4C_3^{(1)}D_3^{(1)} - \tfrac{3}{4}B_3^{(3)}]\}[2(\omega^2 + \lambda^2)]^{-1}$$

and (20)

$$D_5^{(2)} = \{\omega[3C_3^{(1)^2} - D_3^{(1)^2} + \tfrac{3}{4}A_3^{(3)} + \tfrac{5}{8}\beta] + \lambda\,[4C_3^{(1)}D_3^{(1)} - \tfrac{3}{4}B_3^{(3)}]\}[2(\omega^2 + \lambda^2)]^{-1} .$$

The solution of Eq. (16), left after the elimination of secular terms, turns out to be

$$x_2 = a^5 [A_5^{(3)} \cos 3\phi + B_5^{(3)} \sin 3\phi + A_5^{(5)} \cos 5\phi + B_5^{(5)} \sin 5\phi] , \tag{21}$$

where

$$A_5^{(3)} = \frac{(-2\omega^2 + \lambda^2)K_1 + 3\omega\lambda K_2}{16\omega_0^2(\omega^2 + \lambda^2)},$$

$$B_5^{(3)} = \frac{-3\omega\lambda K_1 + (-2\omega^2 + \lambda^2)K_2}{16\omega_0^2(\omega^2 + \lambda^2)}, \tag{22}$$

$$A_5^{(5)} = \frac{(6\omega^2 - \lambda^2)(3A_3^{(3)} + \beta/4) - 15\omega\lambda B_3^{(3)}}{64\omega_0^2(9\omega^2 + \lambda^2)},$$

and

$$B_5^{(5)} = \frac{5\omega\lambda(3A_3^{(3)} + \beta/4) + 3(6\omega^2 - \lambda^2)B_3^{(3)}}{64\omega_0^2(9\omega^2 + \lambda^2)}, \tag{23}$$

with

$$K_1 = 9(2\omega D_3^{(1)} + \lambda C_3^{(1)})A_3^{(3)} - \tfrac{3}{2}A_3^{(3)} + 9(\lambda D_3^{(1)} - 2\omega C_3^{(1)})B_3^{(3)} - \tfrac{5}{16}\beta$$

and

$$K_2 = 9(2\omega C_3^{(1)} - \lambda D_3^{(1)})A_3^{(3)} + 9(2\omega D_3^{(1)} + \lambda C_3^{(1)})B_3^{(3)} - \tfrac{3}{2}B_3^{(3)} . \tag{24}$$

Substituting for ξ_1 and ξ_2 from Eqs. (13) and (19) into the expression for $\dot a$ in Eq. (7), we have

$$\dot{a} = -(\lambda/2)a + \epsilon C_3^{(1)}a^3 + \epsilon^2 C_5^{(2)}a^5, \tag{25}$$

which upon integration yields[25]

$$a = \frac{a_0 \exp(-\lambda t/2)}{[1 - \epsilon(2/\lambda)C_3^{(1)}a_0^2\{1 - \exp(-\lambda t)\} - \epsilon^2(C_5^{(2)}/\lambda)a_0^4\{1 - \exp(-2\lambda t)\}]^{1/2}}. \tag{26}$$

Using expressions for ω_1 and ω_2, Eqs. (13) and (19), and that for $\dot{\phi}$, Eq. (7), followed by integration, gives

$$\phi = \omega t + \epsilon D_3^{(1)} I_1 a_0^2 + \epsilon^2 D_5^{(2)} I_2 a_0^4 + \phi_0, \tag{27}$$

where

$$I_1 = \begin{cases} -\dfrac{2}{\epsilon\lambda\sqrt{S}}\left[\tan^{-1}\dfrac{Q + 2\epsilon R \exp(-\lambda t)}{\sqrt{S}} - \tan^{-1}\dfrac{Q + 2\epsilon R}{\sqrt{S}}\right], & \text{for } S > 0, \\ -\dfrac{1}{\epsilon\lambda\sqrt{-S}}\ln\left|\dfrac{S + Q^2 + 2\epsilon R[Q\{1 + \exp(-\lambda t)\} - \sqrt{-S}\{1 - \exp(-\lambda t)\}] + 4\epsilon^2 R^2 \exp(-\lambda t)}{S + Q^2 + 2\epsilon R[Q\{1 + \exp(-\lambda t)\} + \sqrt{-S}\{1 - \exp(-\lambda t)\}] + 4\epsilon^2 R^2 \exp(-\lambda t)}\right|, & \text{for } S < 0 \end{cases} \tag{28}$$

and

$$I_2 = \frac{1}{\epsilon^2 S\lambda}\left[\frac{2P + \epsilon Q \exp(-\lambda t)}{P + \epsilon Q \exp(-\lambda t) + \epsilon^2 R \exp(-2\lambda t)} - \frac{2P + \epsilon Q}{P + \epsilon Q + \epsilon^2 R}\right] - \frac{Q}{\epsilon S} I_1. \tag{29}$$

Here

$$S = (4a_0^4/\lambda^2)[\lambda C_5^{(2)} - C_3^{(1)^2} - 2\epsilon C_3^{(1)} C_5^{(2)} a_0^2 - \epsilon^2 C_5^{(2)^2} a_0^4]. \tag{30}$$

and

$$S = (4a_0^4/\lambda^2)[\lambda C_5^{(2)} - C_3^{(1)^2} - 2\epsilon C_3^{(1)} C_5^{(2)} a_0^2 - \epsilon^2 C_5^{(2)^2} a_0^4].$$

Combining the different expressions obtained above, the solution to the equation of motion of the damped sextic oscillator, up to second order in ϵ, becomes

$$x = a\cos\phi + \epsilon a^3[A_3^{(3)}\cos 3\phi + B_3^{(3)}\sin 3\phi] + \epsilon^2 a^5[A_5^{(3)}\cos 3\phi + B_5^{(3)}\sin 3\phi + A_5^{(5)}\cos 5\phi + B_5^{(5)}\sin 5\phi], \tag{31}$$

where the coefficients $A_3^{(3)}$, $B_3^{(3)}$, $A_5^{(3)}$, $B_5^{(3)}$, $A_5^{(5)}$, and $B_5^{(5)}$ are given by Eqs. (15), (22), and (23); a is given by Eq. (26); and ϕ is given by Eq. (27).

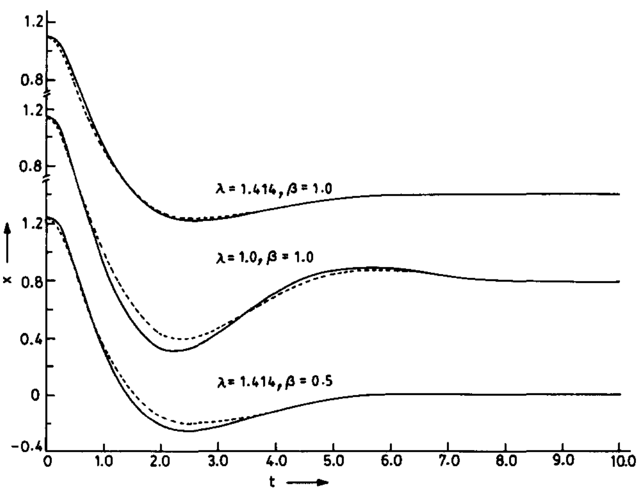


FIG. 1. Plot exhibiting dependence of x on t for the damped sextic oscillator, as obtained by the perturbation method (solid line) and the Runge–Kutta method (dashed line) for different values of λ and β.

III. RESULTS

The dependence of x on t has been determined by performing computations for different values of λ and β, taking $\omega_0 = 1$ and $\epsilon = 1$. The typical results so obtained are shown in Fig. 1. In order to check these findings, the solution to Eq. (2) has also been found by numerical integration using the fourth-order Runge–Kutta method.[24] The initial conditions used for the numerical calculations have been defined by $\phi_0 = 0$ and $\dot{x} = 0$ at $t = 0$; these are, in turn, employed to find a_0 and x at $t = 0$. The time interval used for the numerical integration is $\Delta t = 0.01$ sec. The results based on these computations are also included in Fig. 1 and are found to be in good agreement with the plots obtained from the solution attained from the perturbation theory. Therefore, Eq. (31) can be taken as the solution to the equation of motion of the linearly damped sextic oscillator.

ACKNOWLEDGMENT

Author SS is thankful to the UGC, New Delhi for financial assistance in the form of a senior research fellowship.

[1]C. M. Bender and T. T. Wu, Phys. Rev. **184**, 1231 (1969).
[2]T. C. Bradbury and A. Brintzenhoff, J. Math. Phys. **12**, 1269 (1971).
[3]S. N. Biswas, K. Datta, R. P. Saxena, P. K. Srivastava, and V. S. Varma, J. Math. Phys. **14**, 1190 (1973).

[4]R. Dutt and M. Lakshmanan, J. Math. Phys. **17**, 482 (1976).
[5]F. T. Hioe, D. Mac Millen, and E. W. Montroll, J. Math. Phys. **17**, 1320 (1976).
[6]F. T. Hioe, D. Mac Millen, and E. W. Montroll, Phys. Rep. **43**, 305 (1978).
[7]A. Nayfeh and D. T. Mook, *Nonlinear Oscillations* (Wiley, New York, 1979).
[8]C. Hayashi, *Nonlinear Oscillations in Physical Systems* (Princeton U.P., Princeton, 1985).
[9]E. R. Vrscay and J. Cizek, J. Math. Phys. **27**, 185 (1986).
[10]C. Itzykson and J. B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1985).
[11]H. Dekker, Phys. Rep. **80**, 1 (1981).
[12]R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[13]D. C. Khandekar and S. V. Lawande, J. Math. Phys. **16**, 384 (1975).
[14]K. H. Yeon, C. I. Um, W. H. Kahng, and Y. D. Kim, New Phys. (Korea) **25**, 109, 117 (1985).
[15]D. C. Khandekar and S. V. Lawande, Phys. Rep. **137**, 115 (1986).
[16]C. A. Aragão de Carvalho, Nucl. Phys. B **119**, 401 (1977).
[17]G. E. Sobelman, Phys. Rev. D **19**, 3754 (1979).
[18]M. Znojil, Phys. Lett. A **116**, 207 (1986).
[19]M. Tater, J. Phys. A: Math. Gen. **20**, 2483 (1987).
[20]A. K. Dutta and R. S. Willey, J. Math. Phys. **29**, 892 (1988).
[21]D. G. Lister, J. N. MacDonald, and N. L. Owen, *Internal Rotation and Inversion* (Academic, New York, 1978).
[22]J. M. Kincaid and E. G. D. Cohen, Phys. Rep. **22**, 57 (1975).
[23]K. S. Mendelson, J. Math. Phys. **11**, 3413 (1970).
[24]M. L. James, G. M. Smith, and J. C. Wolford, *Applied Numerical Methods for Digital Computation* (Harper and Row, New York, 1977).
[25]I. S. Gradshteyn and I. M. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1980).

Application of spectral deformation to the Vlasov–Poisson system. II. Mathematical results

Peter D. Hislop[a)]
Department of Mathematics, University of Toronto, Toronto, Ontario M5S 1A1, Canada
John David Crawford[b)]
Institute for Nonlinear Science, University of California, San Diego, La Jolla, California 92093

(Received 14 December 1987; accepted for publication 14 June 1989)

This paper presents a mathematical description of the linearized Vlasov–Poisson operator L_k acting on a family of Banach spaces X_p, related to $L^p(\mathbb{R})$, and the application of the method of spectral deformation to this model. It is shown that a type-A analytic family of operators $L_k(\theta)$, $\theta \in \mathbb{C}$, $L_k(0) = L_k$ can be associated with L_k. By means of this family, the Landau damped modes of the plasma are identified as the spectral resonances of L_k. Existence and uniqueness of solutions to the initial-value problem for the evolution equation $\partial_v g = L_k(\theta)g$ is proven. An expansion of any solution to the initial-value problem (with sufficiently smooth initial data) is obtained in terms of the eigenfunctions of $L_k(\theta)$ and a spectral integral over the essential spectrum. This is applied to derive an expansion for solutions to the Vlasov equation in which the Landau damped portions of the distribution function are manifestly exhibited. A self-contained discussion of the spectral deformation method and an extension of it to certain closed operators on Banach spaces is also given.

I. INTRODUCTION

This paper gives a technical description of the linearized Vlasov–Poisson model and of the application of spectral deformation techniques to this model. It provides the mathematical descriptions and proofs of the results discussed in our previous paper.[1] We refer the reader to Ref. 1 for a detailed discussion of the physical motivation for this work: We use the same notation and terminology.

In Ref. 1 we applied the method of spectral deformation[2,3] to Vlasov theory to analyze certain phenomena which occur in the linearized Vlasov equation and extended the method to study the nonlinear equation. This technique was originally developed to study self-adjoint Schrödinger operators H on a Hilbert space. In that context, it provides a powerful method for studying the meromorphic continuation of matrix elements of the resolvent of H onto the second Riemann sheet, and hence the singular continuous spectrum, and the perturbation of eigenvalues embedded in the continuous spectrum.

Many problems of this type also arise in the study of models in plasma physics and fluid dynamics. In many ways, the one-dimensional Vlasov–Poisson model is the simplest example. Because we choose to analyze the linearized Vlasov equation for spatially homogeneous equilibria, it is possible to explicitly calculate the resolvent for the linear operator. Consequently, the application of the spectral deformation technique is considerably simplified. However, one cannot realistically expect to calculate the resolvent except in the simplest models. For example, if the Vlasov–Poisson system is linearized about an inhomogeneous equilibrium, this will not be possible in general. Thus for more complicated (and more realistic) models an extension of the approach developed in quantum mechanics of studying matrix elements of the resolvent between vectors from a dense set is likely to be more useful. Because of this and the scarcity of a concise discussion of these matters in the literature, we include in this paper a general discussion of the spectral deformation method as applied to quantum mechanical models and an extension of this to Banach spaces in certain situations.

Although our extension in Ref. 1 of spectral deformation to include nonlinear effects involves some subtle and not yet fully understood mathematical issues (as described in Ref. 1), the treatment of the linearized Vlasov–Poisson equation is amenable to a rigorous formulation, which we present in this paper. Our analysis differs from existing mathematical treatments of the linear equation[4] in that we discuss, among other things, the initial-value problem for the analytically continued equation and apply a simplified method introduced in Ref. 5 for obtaining the eigenfunction expansion associated with the analytically continued linear Vlasov operator.

The Vlasov–Poisson equation describes the dynamics of a collisionless plasma in the electrostatic approximation. We treat this equation in its simplest setting: a neutral, unmagnetized plasma in one dimension with periodic boundary conditions. The ions form a fixed background of positive charge. We let $F_0(v)$ denote the spatially homogeneous equilibrium distribution of the electrons and write the electron distribution function as $F(x,v,t) = F_0(v) + f(x,v,t)$. Since F_0 is spatially homogeneous, we Fourier analyze $f(x,v,t)$ and write $f(x,v,t) = \Sigma_k e^{ikx} f_k(v,t)$. The linearized Vlasov–Poisson equation for $f_k(v,t)$ is

$$\partial_t f_k(v,t) = L_k f_k(v,t)\,, \tag{1.1}$$

where

$$L_k f_k(v,t) = \begin{cases} 0, & k = 0, \\ -ik\,[v f_k(v,t) + \eta(k,v) \\ \quad \times \int_{-\infty}^{\infty} dv' f_k(v',t)\,], & k \neq 0, \end{cases} \tag{1.2}$$

and

[a)] Present address: Mathematics Department, University of Kentucky, Lexington, Kentucky 40506-0027.

[b)] Address after January 1, 1990: Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260.

$$\eta(k,v) \equiv -(\omega_p k^{-1})^2 \left(\frac{\partial F_0}{\partial v}\right). \tag{1.3}$$

In (1.3), $\omega_p = [4\pi n_0 e^2/m]^{1/2}$ is the plasma frequency.

The linear stability of the equilibrium F_0 depends on the spectrum of the linear operator L_k. Of special importance is an understanding of the behavior of the eigenvalues of L_k as the physical characteristics of the equilibrium are varied. For example, for certain critical distributions, the spectrum of L_k, denoted $\sigma(L_k)$, has eigenvalues embedded in the essential spectrum $\sigma_{\text{ess}}(L_k) = i\mathbb{R}$. As F_0 is altered slightly, these eigenvalues may move into the right half-plane, in which case they describe (linearly) unstable modes of the plasma. A thorough understanding of the onset of such an instability requires a detailed understanding of the embedded eigenvalues at criticality. One result of the spectral deformation method is that the continuum can be separated from the critical, embedded eigenvalues so that they become isolated and hence amenable to study by standard techniques. Another result of the spectral deformation method is that the critical embedded eigenvalues are shown to be generically simple.

On the other hand, it may happen that as F_0 is varied the embedded eigenvalues of L_k simply disappear from $\sigma(L_k)$ altogether. For example, if F_0 is a Maxwellian distribution, then $\sigma(L_k) = i\mathbb{R}$ and L_k has no eigenvalues (for $k \neq 0$). However, in this case it is well known that there are physically observable, collective modes which in linear theory decay in time because of Landau damping.[6] The existence of such modes is difficult to discern from a direct study of $\sigma(L_k)$. By using the spectral deformation method, we show that these modes are *spectral resonances* of L_k. We also obtain a new expansion for solutions to the initial-value problem associated with (1.1), in which the Landau damped portions of the distribution function are manifestly exhibited.

In Ref. 1, we analyze these problems from a physical point of view and refer the reader there for further details. We also discuss in Ref. 1 the extension of the spectral deformation method to the full nonlinear problem and derive, among other things, exact wave amplitude equations.

The contents of this paper are as follows. In Sec. II, we provide a general discussion of the spectral deformation method. This method possesses several intrinsic properties which we state and prove. The power and flexibility of the method is primarily due to these properties. We also discuss extensions of the technique to the more general Banach space setting, as required for the Vlasov model.

In Sec. III we apply the method to the linearized Vlasov operator L_k. We show that L_k is densely defined and closed on its natural domain in $L^p(\mathbb{R})$, $1 \leqslant p < \infty$. However, the dynamics is not easily formulated on $L^p(\mathbb{R})$ so we introduce related Banach spaces X_p, $1 \leqslant p < \infty$. On these spaces, we introduce the one-parameter group of velocity translations $U(\theta)$, $\theta \in \mathbb{R}$, and define a family of operators $L_k(\theta) \equiv U(\theta) L U(\theta)^{-1}$. It is shown that these form a type-A analytic family of operators on some strip $S_0 \equiv \{\theta \in \mathbb{C} \mid |\text{Im } \theta| < \theta_0\}$ in the complex θ plane.

The spectral properties of the family $L_k(\theta)$, $\theta \in S_0$ are studied in Sec. IV. We construct the resolvent operator $R_k(z,\theta) \equiv (z - L_k(\theta))^{-1}$, identify the spectrum of $L_k(\theta)$, and analyze its dependence on θ. It is shown that the spectral resonances of L_k arise from the Landau poles. Elementary semigroup theory is used to prove that each $L_k(\theta)$, $\theta \in S_0$ generates a one-parameter group W_t which provides a unique solution to the initial-value problem for the evolution equation $\partial_t g = L_k(\theta) g$.

In Sec. V, an inverse Laplace transform integral for $R_k(z,\theta)$ is constructed and proved to converge to the evolution group W_t for $t \geqslant 0$ in the strong sense. From this we derive a convergent expansion for $g(v,t) \equiv W_t g(v,0)$ for the suitably regular initial conditions $g(v,0) \in X_p$ in terms of the eigenfunctions of $L_k(\theta)$ and a spectral integral over the essential spectrum of $L_k(\theta)$. When $\text{Im } \theta = 0$, this reduces to the familiar Vlasov expansion.[4,7,8] An interesting feature of this expansion for $\text{Im } \theta < 0$ is that the influence of the Landau poles on the long-time behavior of $g(v,t)$ emerges naturally from the spectral theory. By a result in Ref. 1, $U(\theta)^{-1} g(v,t)$ provides a solution to the original evolution equation (1.1). When the eigenfunction expansion of $g(v,t)$ is combined with this representation of the Vlasov solution, we obtain an expression for $f_k(v,t)$ which exhibits simultaneously the time-asymptotic signficance of the Landau poles and their role as spectral resonances for L_k.

The three dielectric functions that arise in the analysis of the linearized problem are studied in Appendix A. The behavior of the roots of these functions is analyzed under suitable analyticity conditions on the equilibrium distribution function F_0. In particular, sufficient conditions are given for the existence of at most a finite number of zeros. This analysis is used to prove that the multiplicity of the embedded eigenvalues of the Vlasov operator at criticality is the same as the resulting unstable mode, i.e., generically simple.

The spectral deformation group $U(\theta)$ described in Sec. III is suitable only when F_0 is the restriction to $\mathbb{R}$ of a function analytic in a strip about $\mathbb{R}$. In Appendix B, we outline a more general method of spectral deformation applicable to a wide class of equilibrium distribution functions, including those with compact support.

II. SPECTRAL DEFORMATION METHODS AND RESONANCES

The spectral deformation method is a tool for the study of phenomena associated with the continuous spectrum of linear operators. Heretofore, applications of the method have been restricted to self-adjoint operators defined on Hilbert spaces.[2,3,9–11] This reflects the historical fact that the methods were invented by mathematical physicists studying quantum mechanical models. In this section, we first review the major characteristics of this method as applied to a self-adjoint operator H. We then indicate how the method can be extended to certain classes of closed operators on Banach spaces. The results for self-adjoint operators are known to specialists, but are not readily accessible; thus we have included them for completeness.

Information about the spectrum and resonances of H can be obtained from studying the resolvent $R_H(z) \equiv (z - H)^{-1}$ as a function of $z \in \rho(H)$, the resolvent set of H, and the meromorphic continuation of $R_H(z)$ onto the second Riemann sheet. As mentioned in Sec. I, an explicit construction of $R_H(z)$ is in general impossible. One of the primary virtues of the spectral deformation method is that it allows the explicit construction of the meromorphic continuation of matrix elements of $R_H(z)$ between a suitable dense set of states onto the second Riemann sheet. This suffices to allow one to prove, for example, the absence of a singular continuous spectrum for H.

More important for our purposes is the fact that the poles of the meromorphic continuation of (matrix elements of) $R_H(z)$ are in one-to-one correspondence with the complex eigenvalues of an analytic family of non-self-adjoint operators $H(\theta)$, $\theta\in\mathbb{C}$, which is explicitly constructed from H. The poles of the continuation of $R_H(z)$ onto $\mathbb{C}^- \equiv\{z\in\mathbb{C}|\operatorname{Im} z<0\}$ are called *spectral resonances* of H and physically represent the metastable states of H. This connection between the poles of $R_H(z)$ and the complex eigenvalues of a family of operators $H(\theta)$ was generalized in Ref. 1 to identify the Landau damped plasma waves with the resonances of the Vlasov operator L_k.

The basic idea of the spectral deformation method is to associate with H an analytic family of operators $H(\theta)$ in the following way. Let $\{U(\theta)|\theta\in\mathbb{R}\}$ be a strongly continuous unitary group. We define the family of operators $H(\theta)$ by

$$H(\theta)\equiv U(\theta)HU(\theta)^{-1}, \quad \theta\in\mathbb{R}. \tag{2.1}$$

The group $U(\theta)$ is chosen to satisfy two criteria:

(i) The operator-valued function $R(z,\theta)\equiv(z-H(\theta))$ is jointly analytic in z and θ on some open set in $\mathbb{C}^2$ (typically, θ belongs to some strip $S_0\equiv\{\theta\,|\,|\operatorname{Im}\theta|<\theta_0\}$).

(ii) The essential spectrum of $H(\theta)$, denoted $\sigma_{\rm ess}(H(\theta))$, $\theta\in S_0$, can be calculated (at least locally) and is deformed from $\sigma_{\rm ess}(H)$ (which is typically a half-line).

Here, $\sigma_{\rm ess}(A)\equiv\sigma(A)\setminus\sigma_d(A)$ for a closed operator A and $\sigma_d(A)$ is the set of all isolated eigenvalues with finite multiplicity. It is criterion (ii) that gives the technique its name. Usually, the family $H(\theta)$ can be shown to be an "analytic family of type A" on the strip S_0. This then implies the analyticity stated in criterion (i).[12] (Type-A analyticity is defined in Sec. III.) In the case that $H=-\nabla^2+V$, criterion (ii) is often replaced by a nontrapping condition[13,14] on the potential V and a vector field f and the group $U(\theta)$ is generated by $D\equiv(i/2)(\nabla\cdot f+f\cdot\nabla)$.

The first example to be extensively studied was the case where $U(\theta)$ implements the dilation group on $L^2(\mathbb{R}^n)$:

$$(U(\theta)f)(x)=e^{n\theta/2}f(e^\theta x), \quad \theta\in\mathbb{R}. \tag{2.2}$$

Suppose that $H=p^2+V$, where $p^2\equiv-\nabla^2$ is the kinetic energy and V is a multiplication operator describing the potential. Then

$$H(\theta)=e^{2\theta}p^2+V_\theta, \quad \theta\in\mathbb{R}, \tag{2.3}$$

where $(V_\theta f)(x)=V(e^\theta x)f(x)$. Now if V is relatively p^2 compact and has an analytic continuation onto some sector $\{z\in\mathbb{C}|\,|\arg z|<\beta\}$, then V_θ is an analytic operator-valued function on $S_0\equiv\{\theta\,|\,|\operatorname{Im}\theta|<\beta\}$. It then follows that $H(\theta)$ is an analytic family of type A on S_0. By the Weyl theorem, $\sigma_{\rm ess}(H(\theta))=\mathbb{R}_+$ for $\theta\in\mathbb{R}$, but for $\theta\in S_0$, $\sigma_{\rm ess}(H(\theta))=e^{2\theta}\mathbb{R}_+$, i.e., the effect of continuing H in θ is to rotate the essential spectrum off the real axis (see Fig. 1). The resonances of H are the complex eigenvalues of $H(\theta)$ located between $\sigma_{\rm ess}(H)=\mathbb{R}_+$ and the half-line $e^{2\theta}\mathbb{R}_+$, $\operatorname{Im}\theta<0$.

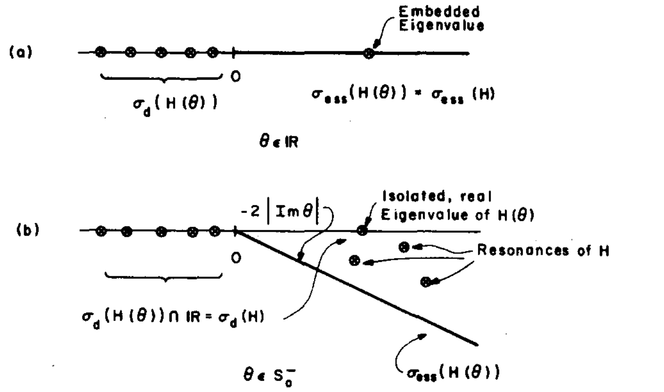


FIG. 1. Typical spectrum of a dilation–analytic Schrödinger operator $H(\theta)=e^{2\theta}p^2+V_\theta$. (a) $\theta\in\mathbb{R}$; thus $H(\theta)$ is self-adjoint and $\sigma(H(\theta))=\sigma(H)$, (b) $\theta\in S_0$ with $\operatorname{Im}\theta<0$; thus $\sigma_{\rm ess}(H(\theta))$ is rotated into the lower-half complex plane.

Let us suppose that a group $U(\theta)$ has been found such that criteria (i) and (ii) are satisfied. Furthermore, near some point λ, let us suppose that the spectrum of $H(\theta)$, $\theta\in\mathbb{R}$ appears as in Fig. 2(a). When $\theta\in S_0$, the spectrum of $H(\theta)$ will typically deform, as illustrated schematically in Fig. 2(b). Note that there may be eigenvalues of $H(\theta)$ lying between the real axis and $\sigma_{\rm ess}(H(\theta))$ which are "uncovered" by taking θ complex.

An important application of this construction is to study the behavior of $R_H(z)$ in a neighborhood of λ using $H(\theta)$. This can be done as follows. Let $A(U)$ denote a dense set of analytic vectors for the group $U(\theta)$, i.e., if $g\in A(U)$, then $\theta\in\mathbb{R}\to U(\theta)g$ has an analytic continuation in θ off the real axis and into $\mathbb{C}$. For $f,g\in A(U)$ the matrix element

$$F_{gf}(z)\equiv\langle g,R_H(z)f\rangle \tag{2.4}$$

for $\operatorname{Im} z\neq0$ satisfies the identity

$$\begin{aligned}F_{gf}(z)&=\langle U(\theta)g,U(\theta)R_H(z)U(\theta)^{-1}U(\theta)f\rangle\\&=\langle g_\theta,R_H(z,\theta)f_\theta\rangle=\langle g_{\theta^*},R_H(z,\theta)f_\theta\rangle\end{aligned} \tag{2.5}$$

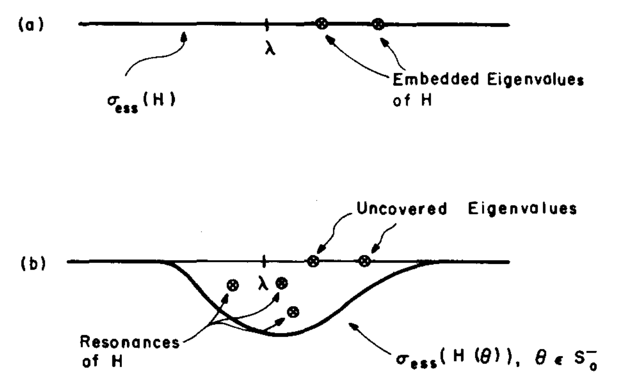


FIG. 2. The method of local spectral deformation provides one with a means of studying eigenvalues of H embedded in the continuous spectrum, as shown in (a), by distorting the continuous spectrum away from these eigenvalues, as shown in (b). The real eigenvalues become isolated eigenvalues of $H(\theta)$ and new complex eigenvalues may appear.

for $\theta \in \mathbb{R}$. Here, $g_\theta \equiv U(\theta)g$ and $R_H(z,\theta) = (z - H(\theta))^{-1}$. The identity (2.5) follows from the unitarity of $U(\theta)$. Since H is self-adjoint, $\sigma(H) \subset \mathbb{R}$. For fixed $z \in \mathbb{C}^+ \equiv \{z \in \mathbb{C} | \text{Im}\, z > 0\}$, we can analytically continue the function

$$F_{gf}(z,\theta) \equiv \langle g_{\theta^*}, R_H(z,\theta) f_\theta \rangle \tag{2.6}$$

in θ onto the strip $S_0^- \equiv \{\theta \in S_0 | -\theta_0 < \text{Im}\, \theta < 0\}$ so that $\sigma(H(\theta))$ swings into the lower half-plane, as shown in Fig. 2(b). This continuation is possible since each component in (2.6) is analytic.

Since $\sigma_{\text{ess}}(H(\theta))$, $\theta \in S_0^-$ has been deformed off the real axis in a neighborhood of λ, we can, for fixed $\theta \in S_0^-$ in (2.6), meromorphically continue $F_{gf}(z,\theta)$ in z from $\text{Im}\, z > 0$ onto this small neighborhood of λ; see Fig. 3. This continuation is meromorphic (rather than analytic) since there may exist poles of $F_{gf}(z,\theta)$ in the local region about λ. We will discuss this below. To apply this continuation analysis to the original matrix element, we recall relation (2.5) and apply the identity principle for meromorphic functions: For $\theta \in S_0^-$ and $z \in \mathbb{C}^+$,

$$F_{gf}(z) = F_{gf}(z,\theta) \tag{2.7}$$

and, by the above discussion, $F_{gf}(z,\theta)$ is meromorphic in a full complex neighborhood of λ. Hence, $F_{gf}(z,\theta)$ provides a meromorphic continuation, denoted by $F_{gf}^{\text{II}}(z)$, of $F_{gf}(z)$ onto the second Riemann sheet near λ, including a real neighborhood of λ. [This continuation is *independent of* θ provided that $\sigma_{\text{ess}}(H(\theta))$ stays away from a neighborhood of λ.] Moreover, according to relation (2.7), the poles of $F_{gf}^{\text{II}}(z)$ in a neighborhood of λ are in one-to-one correspondence with the eigenvalues of $H(\theta)$. These poles/eigenvalues are the spectral resonances of H.

The construction outlined above yields several results. (i) We obtain explicit control over $\langle g, R_H(\mu + i\epsilon) f \rangle$ as $\epsilon \to 0$, with $g, f \in A(U)$, and μ near λ; in particular, the only singularity occurs when μ is an eigenvalue. (ii) As we will show below, if H has real eigenvalues near λ, they remain eigenvalues of $H(\theta)$, $\theta \in S_0^-$ (with the same multiplicity), but now these eigenvalues are isolated eigenvalues of $H(\theta)$ since the $\sigma_{\text{ess}}(H(\theta))$ has been moved away. (iii) The poles of the meromorphic continuation of matrix elements of $R_H(z)$ are identified with the eigenvalues of $H(\theta)$. Results (ii) and (iii) are of primary interest here. Result (ii) provides one with a technique for studying embedded eigenvalues of H (for example, their behavior under perturbations) by viewing them as isolated eigenvalues of $H(\theta)$. Result (iii) reexpresses the problem of locating the poles of $F_{gf}^{\text{II}}(z)$ as an eigenvalue problem, although at the cost of introducing non-self-adjoint operators.

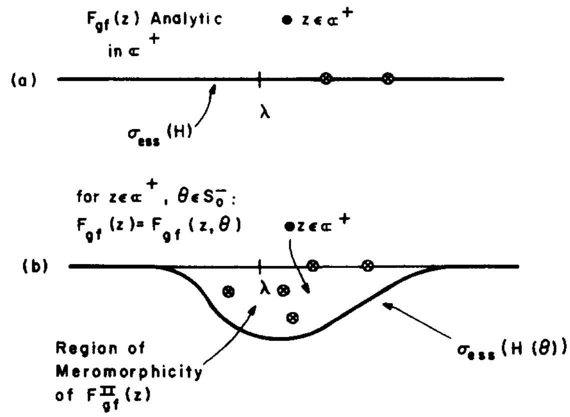


FIG. 3. Meromorphic continuation of the matrix elements $\langle g, R_H(z) f \rangle \equiv F_{gf}(z)$ from $z \in \mathbb{C}^+$, shown in (a), to $z \in \mathbb{C}^-$ with Re z near λ, can be obtained by deforming $\sigma_{\text{ess}}(H(\theta))$ near λ away from the real axis, as shown in (b), and using the identity $F_{gf}(z) = F_{gf}(z,\theta)$, which holds for $z \in C^+$ and $\theta \in S_0$, Im $\theta < 0$, and then continuing in z.

There is a degree of arbitrariness in the above construction, but we show that it satisfies two intrinsic properties:

(i) If $z(\theta)$ is a complex eigenvalue of $H(\theta)$, then it is *independent of* θ provided that θ varies in such a way that $z(\theta)$ remains isolated from $\sigma_{\text{ess}}(H(\theta))$.

(ii) If the spectral deformation construction is made for two different groups $U_1(\theta)$ and $U_2(\theta)$ that both satisfy criteria (i) and (ii) *and* that share a common dense domain of analytic vectors, then on a common neighborhood of λ both deformations will reveal the same complex eigenvalues on the second Riemann sheet.

To understand property (i), note that for $\phi \in \mathbb{R}$, $U(\phi)H(\theta)U(\phi)^{-1} = H(\theta + \phi)$; thus $\sigma_d(H(\theta)) = \sigma_d(H(\theta + \phi))$, where σ_d is the discrete spectrum. Hence, if $z(\theta) \in \sigma_d(H)$, it is independent of Re θ. However, since $H(\theta)$ is an analytic family, $z(\theta)$ is analytic and hence a constant as long as θ varies in such a way that $z(\theta)$ remains isolated from $\sigma_{\text{ess}}(H(\theta))$.

Property (ii) can be seen as follows. Let $U_i(\theta)$, $i = 1,2$ be two groups satisfying criteria (i) and (ii) for $\theta \in S_0$ (a common strip) and such that $A_I \equiv A(U_1) \cap A(U_2)$ is dense. Then for $f, g \in A_I$ we have equality of the matrix elements for $z \in \mathbb{C}^+$ and $\theta \in \mathbb{R}$:

$$F_{gf}(z) = F_{gf}^{(1)}(z,\theta) = F_{gf}^{(2)}(z,\theta), \quad \theta \in \mathbb{R}, \tag{2.8}$$

where

$$F_{gf}^{(i)}(z,\theta) \equiv \langle U_i(\theta)g, (z - H_i(\theta))^{-1} U_i(\theta) f \rangle,$$
$$H_i(\theta) \equiv U_i(\theta) H U_i(\theta)^{-1}$$

for $i = 1,2$. We now continue the latter two functions of (2.8) in θ for $z \in \mathbb{C}^+$ fixed and then in z, as in the previous discussion. Hence, we obtain two meromorphic continuations of $F_{gf}(z)$ into $\mathbb{C}^-$; therefore, they must be the same meromorphic function. The poles of $F_{gf}^{(i)}(z,\theta)$, $i = 1,2$ are therefore eigenvalues of both $H_1(\theta)$ and $H_2(\theta)$. This proves that the eigenvalues of $H(\theta)$ are independent of the choice of group.

Properties (i) and (ii) give the spectral deformation method its power and flexibility. Provided that one can prove the analyticity of $H(\theta)$, then one knows that the results are independent of the group used and are "locally" independent of $\theta \in S_0$. Hence, the group chosen can be tailored to the problem at hand. It should be noted that the poles of $F_{gf}^{\text{II}}(z)$ are also independent of the pair of vectors $g, f \in A(U)$ since they are eigenvalues of $H(\theta)$: They may, however, depend on the dense set $A(U)$.[15]

We next consider the manner in which spectral deformation affects a separation of embedded eigenvalues from the essential spectrum. Let λ be a real, embedded eigenvalue of H (with finite multiplicity) and let P_λ denote the projection onto the spectral subspace of H for λ. Then by the spectral theorem, $P_\lambda = s - \lim_{\epsilon\to 0^+} i\epsilon(\lambda + i\epsilon - H)^{-1}$. We show that $\lambda \in \sigma_d(H(\theta))$ for $\theta \in S_0^-$. By (2.5) and (2.6), we have, for $\theta \in \mathbb{R}$ and $g, f \in A(U)$,

$$\lim_{\epsilon\to 0^+} i\epsilon F_{gf}(\lambda + i\epsilon) = \langle g, P_\lambda f\rangle = \lim_{\epsilon\to 0^+} i\epsilon F_{gf}(\lambda + i\epsilon, \theta) . \tag{2.9}$$

The last term in (2.9) can be analytically continued in θ before the limit is taken. If $H(\theta)$ has no eigenvalue at λ, $F_{gf}(z,\theta)$ is analytic in a neighborhood of λ and consequently, the limit on the rhs of (2.9) vanishes for all f and g. This is a contradiction since the density of $A(U)$ implies $\langle g, P_\lambda f\rangle \neq 0$ for some $f, g \in A(U)$. Thus $H(\theta)$ must have an eigenvalue at λ. A similar argument establishes the converse: If $\lambda \in \sigma_d(H(\theta)) \cap (0,\infty)$, then λ is also an eigenvalue of H. This completes our synopsis of the self-adjoint theory.

The linear operators that arise in plasma and fluid models are frequently non-self-adjoint. It is fortunate that this does not preclude the application of spectral deformation. We now outline a generalization of the method described above to the study of a closed operator A on a Banach space X. Certain additional assumptions will be required to compensate for the loss of self-adjointness. For a closed linear operator A, $\sigma(A)$ is in general a closed subset of $\mathbb{C}$ with a nonempty interior. If $\lambda \in \sigma(A)$ is an embedded eigenvalue lying on the boundary of the spectrum, the spectral deformation method allows local distortion of $\sigma_{\text{ess}}(A)$, so that λ becomes isolated and hence amenable to standard methods (see Fig. 4). In order to discuss the meromorphic continuation of $R_A(z)$ in a neighborhood of a point λ, we must further suppose that A locally has a spectrum with an empty interior, i.e., it is a curve Γ (see Fig. 5).

In problems in fluid mechanics and plasma physics, $\sigma_{\text{ess}}(A)$ is typically a curve and represents the "continuum

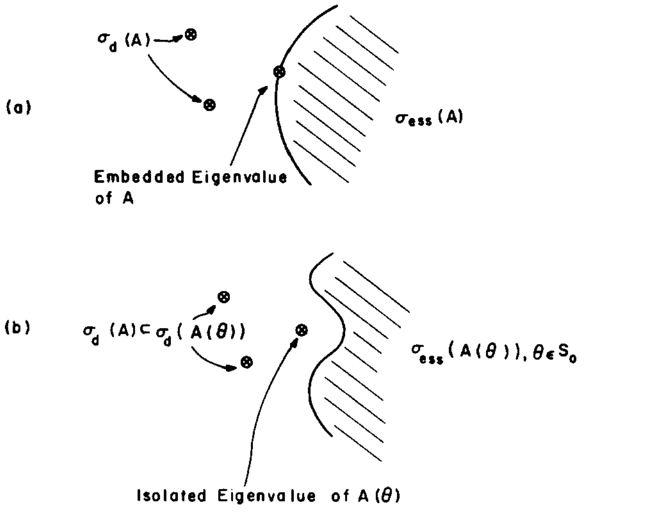


FIG. 4. In the general case of a closed operator A on a Banach space X, $\sigma(A)$ may have a nonempty interior and eigenvalues embedded on the boundary of its spectrum, as in (a). Local spectral deformation methods can be used to move $\sigma(A)$, so that the eigenvalue becomes isolated, as in (b).

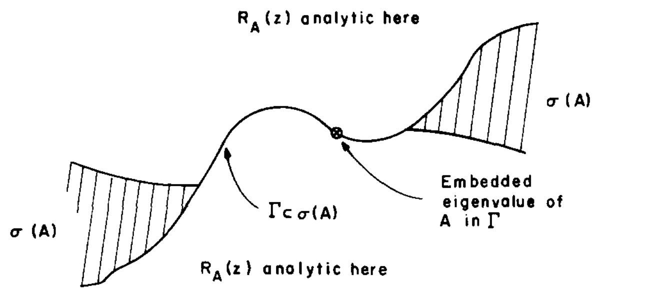


FIG. 5. In many cases. $\sigma(A)$ may consist locally of a curve Γ with, perhaps, an embedded eigenvalue. Typically, the resolvent is defined on either side of Γ and discontinuous across it.

modes" of the system. The notions of eigenvalues embedded in $\sigma(A)$ and of resonances carry the same dynamical meaning as in the Hilbert space case, in the context of the evolution equation $\partial_t u = Au$. Instead of exhibiting the oscillatory behavior characteristic of continuum modes, the resonance states are typically damped modes of the system; their presence is not easily discernable from $\sigma(A)$.

To formulate the notion of resonance, let A be a closed operator as above and assume that $\sigma_{\text{ess}}(A)$ in a neighborhood of a point $\lambda \in \mathbb{C}$ consists of a curve Γ, as in Fig. 5. Let X^* denote the dual Banach space to X which consists of all continuous linear functionals on X. We consider "matrix elements" of $R_A(z)$ formed from the pairs $(l,f) \in X^* \times X$ and set $F_{lf}(z) \equiv l(R_A(z)f)$. For each pair $(l,f) \in X^* \times X$, $F_{lf}(z)$ is analytic on $\rho(A)$. We define *spectral resonances of A near* Γ as poles in the meromorphic continuation of matrix elements of $R_A(z)$ across Γ relative to a dense set $N^* \times N \subset X^* \times X$.

To study $R_A(z)$ in a neighborhood of Γ we can introduce the method of spectral deformation as follows. Let $\theta \in \mathbb{R} \to U(\theta)$ be a strongly continuous one-parameter group of bounded operators on X. As before, we require that $U(\theta)$ be such that $A(\theta) \equiv U(\theta) A U(\theta)^{-1}$ extends to an analytic family of operators on some strip S_0, at least in the sense that $R_A(z,\theta)$ is jointly analytic. Note that for $\theta \in \mathbb{R}$, $\sigma(A) = \sigma(A(\theta))$. We also require the $U(\theta)$ be such that we can compute $\sigma(A(\theta))$ near Γ and that for the Im $\theta \neq 0$, it is deformed away from Γ to some curve Γ'; see Fig. 6. Of course, eigenvalues of $A(\theta)$ may be uncovered lying between Γ and Γ'.

Given a group $U(\theta)$, we assume the existence of a dense

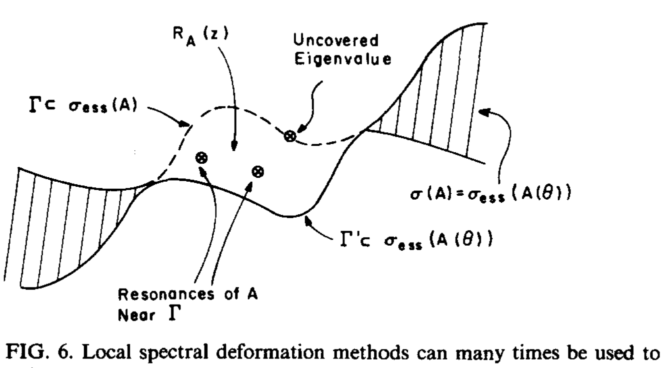


FIG. 6. Local spectral deformation methods can many times be used to deform $\Gamma \subset \sigma(A)$ to a new curve $\Gamma' \subset \sigma(A(\theta))$ in a way such that the eigenvalues of A embedded in Γ become isolated and spectral resonances may be uncovered.

set N of analytic vectors for the group in X. (Unlike the self-adjoint case, the existence of such a set is not guaranteed by a general theorem.) In an obvious way, $U(\theta)$ induces an action $\tilde{U}(\theta)$ on X^*. For any $l \in X^*$ and $f \in X$, we have

$$l(U(\theta)f) \equiv l_\theta(f), \quad \theta \in \mathbb{R},$$

where

$$l_\theta \equiv \tilde{U}(\theta) l.$$

It follows that $\theta \in \mathbb{R} \to \tilde{U}(\theta)$ is a strongly continuous one-parameter group of bounded operators on X^*. Let $N^* \subset X^*$ be a dense set of analytic vectors for $\tilde{U}(\theta)$, whose existence we assume.

The major notions discussed above extend to A, $A(\theta)$, and $N^* \times N \subset X^* \times X$ relative to $\Gamma \subset \sigma_{\text{ess}}(A)$. In particular, for $(l,f) \in N^* \times N$, the jointly analytic function

$$F_{lf}(z,\theta) \equiv l(R_A(z,\theta) f)$$

provides a meromorphic continuation of $F_{lf}(z)$ through $\sigma_{\text{ess}}(A)$ near Γ. The poles of this continuation in the region between Γ and Γ' are in one-to-one correspondence with the eigenvalues of $A(\theta)$, $\theta \in S_0$. These are the resonances of A near Γ. Moreover, it is immediate to check that the two intrinsic properties of the construction described above continue to hold in the setting described here.

The question of the stability of the eigenvalues of A embedded in Γ under the continuation is more subtle. This is due to the fact that in general, an explicit expression for the projection P_λ onto the eigenspace of A for an embedded eigenvalue λ in terms of $R_A(z)$ is not available, as it was in the self-adjoint case. However, we can envision the following situation (which includes the Vlasov equation). Suppose that λ is an embedded eigenvalue of A and $A\psi = \lambda\psi$. Furthermore, suppose that $U(\theta)$ is a deformation group for A [in particular, $A(\theta)$ is type A] and ψ is an analytic vector for $U(\theta)$ on some strip S_0, i.e., $\theta \in S_0 \to U(\theta)\psi$ is an analytic function. Then $U(\theta)\psi \equiv \psi_\theta$ is an eigenvector for $A(\theta)$, for $\theta \in S_0$, with eigenvalue λ: $A(\theta)\psi_\theta = \lambda\psi_\theta$. To check this, note that for $\theta \in \mathbb{R}$,

$$A(\theta)\psi_\theta = \lambda\psi_\theta, \quad \theta \in \mathbb{R}$$

and the rhs is analytic on S_0; thus the lhs is also and it suffices to show that $\psi_\theta \in D(A)$. As A is closed, $A^\dagger$ is densely defined on X^* (provided that X is reflexive). Then for all pairs (l,ψ) with $l \in D(A^\dagger)$,

$$l(A\psi_\theta) = (A^\dagger l)(\psi_\theta), \quad \theta \in \mathbb{R}$$

and the rhs is analytic in θ on S_0. Since A is closed, this implies that $\psi_\theta \in D(A)$ for $\theta \in S_0$. In summary, a sufficient condition for an embedded eigenvalue of A to persist as an eigenvalue of $A(\theta)$ is that the corresponding eigenfunction be an analytic vector.

III. LINEAR VLASOV OPERATOR AND COMPLEX TRANSLATIONS

In this and the following sections, we are concerned with the linear Vlasov operator L_k defined by

$$(L_k f)(v) = \begin{cases} -ik\,[vf(v) + \eta(k,v) \int_{-\infty}^{\infty} f(v')dv'], & k \neq 0, \\ 0, & k = 0, \end{cases} \tag{3.1}$$

where k indicates the wavenumber. The operator L_k depends on the initial equilibrium velocity distribution F_0 through $\eta(k,v)$, as defined in (1.3).

We assume that L_k acts on a function space X which we now consider. In order to analyze L_k in a reasonable way, it must be densely defined and closable on a domain in X. We shall restrict our preliminary discussion to the classical Banach spaces $L^p(\mathbb{R})$, $1 \leqslant p < \infty$, although we will find that the model is more readily described on different, but related Banach spaces. (We note that since L_k is not defined on constant functions, it is not densely defined on L^∞.)

The theory assumes various properties for F_0; we state these properties in two assumptions. Assumption 1 pertains to the decay of F_0 as $|v| \to \infty$. Assumption 2 concerns the analyticity of F_0 and is stated and used later in this section.

Assumption 1: The function $F_0 \in S(\mathbb{R})$, the set of all infinitely differentiable functions which decay faster than any polynomial as $|v| \to \infty$, and F_0 is real.

Actually, weaker decay conditions on F_0 are adequate; we mention these below.

For each p, $1 \leqslant p < \infty$, L_k is defined on a natural domain $D_p(L_k) \equiv \{f \in L^p | L_k f \in L^p\}$. To describe this domain, let M_v denote the operator of multiplication by v, i.e., $M_v f = vf$. Let $D_p(M_v)$ be the natural domain of M_v in L^p. It is clear from (3.1) that if $f \in D_p(L_k)$ $(k \neq 0)$, then $\int_{-\infty}^{\infty} f < \infty$ and hence, from Assumption 1 on η, $f \in D_p(M_v)$. Conversely, if $f \in D_p(M_v)$, we show that $f \in D_p(L_k)$ and $\int_{-\infty}^{\infty} f < \infty$. For $p = 1$, this is obvious. For $p > 1$, it follows by the Hölder inequality that for any $\alpha \in \mathbb{R} \setminus \{0\}$,

$$\left| \int_{-\infty}^{\infty} f(v)dv \right| \leqslant \left[\int_{-\infty}^{\infty} |v + i\alpha|^{-q} dv \right]^{1/q} \|(M_v + i\alpha) f\|_p, \tag{3.2}$$

where $q > 1$ satisfies $p^{-1} + q^{-1} = 1$. Since both factors on the rhs in (3.2) are finite, it follows that $\int_{-\infty}^{\infty} f < \infty$ and $f \in D_p(L_k)$. Consequently,

$$D_p(L_k) = D_p(M_v); \tag{3.3}$$

this domain is dense in L^p since it includes $S(\mathbb{R})$. We remark (with respect to Assumption 1) that it suffices to assume simply that $\eta \in L^p$ for (3.3) to hold.

Our first result concerns the closability of L_k on its natural domain $D_p(L_k) \subset L^p(\mathbb{R})$. It is convenient to write $L_k = A + B$, where A and B are defined by

$$(Ag)(v) \equiv -ik(M_v g)(v) \tag{3.4}$$

and

$$(Bg)(v) \equiv -ik\eta(k,v) \int_{-\infty}^{\infty} g(v')dv' \tag{3.5}$$

for $g \in D_p(M_v)$. We first consider M_v.

Lemma 3.1: The operator M_v is closed on its natural domain $D_p(M_v)$, $1 \leqslant p < \infty$.

Proof: Let $\{h_\alpha\}$ be a Cauchy sequence in $D_p(M_v)$ with $h = \lim_\alpha h_\alpha$ and such that $\{M_v h_\alpha\}$ is Cauchy with k

$\equiv \lim_\alpha M_v h_\alpha$. Let χ_R be the characteristic function for $[-R,R]$. Then $\chi_R M_v$ is bounded and $\|\chi_R M_v\| = R$. By the dominated convergence theorem, $\lim_{R\to\infty} \chi_R\, g = g$ for any $g\in L^p$ and $\lim_{R\to\infty} M_v\, \chi_R\, g = M_v\, g$ for any $g\in D_p(M_v)$. Now we have

$$\lim_{R\to\infty} M_v\chi_R h = \lim_{R\to\infty}\lim_{\alpha\to\infty} M_v\chi_R h_\alpha = k$$

(all limits are L^p limits). Consequently, there exists a subsequence $\{M_v\chi_{R_n} h\}$ which converges to k pointwise a.e. as $n\to\infty$. However, $M_v\chi_R h = \chi_R M_v h$ converges to $M_v h$ pointwise a.e. so that $M_v h = k$ a.e. Since $k\in L^p$ this implies that $h\in D_p(M_v)$. Hence, M_v is closed on $D_p(M_v)$. □

We note that in linear Vlasov theory there is a fundamental difference between $p=1$ and $p>1$. For $p=1$, Assumption 1 implies tht B is a bounded operator, whereas for $p>1$, B is unbounded. However, we have the following lemma.

Lemma 3.2: The operator B defined in (3.5) is bounded relative to A, with relative bound zero.

Proof: For $p=1$, B is bounded. For $p>1$, it follows from (3.2) that $D_p(B)\supset D_p(A)$. Let $g\in D_p(A)$; then for any $\alpha>0$,

$$\|Bg\|_p = |k|\,\|\eta\|_p \left|\int_{-\infty}^{\infty} \frac{dv}{v+i\alpha}(v+i\alpha)g(v)\right|$$
$$\leqslant |k|\,\|\eta\|_p\,\|(M_v+i\alpha)g\|_p\left[\int_{-\infty}^{\infty}\frac{dv}{(v^2+\alpha^2)^{q/2}}\right]^{1/q}, \tag{3.6}$$

where $q^{-1}+p^{-1}=1$. Let $Q(\alpha)^q$ be the integral on the rhs in (3.6). Then

$$Q(\alpha) = \alpha^{1/q-1}\beta \tag{3.7}$$

for some $\beta>0$ independent of α. As for $q>1$, (3.7) shows that $\lim_{\alpha\to\infty} Q(\alpha)=0$ and it follows from (3.6) that the A bound of B is zero. □

Corollary 3.3: The operator L_k is closed on its natural domain.

Proof: By Lemma 3.2, B is relatively A bounded with a relative bound of less than 1 and, by Lemma 3.1, A is closed on $D_p(M_v)$; thus by a stability theorem[16] for closed operators, $L_k = A+B$ is closed on $D_p(M_v)=D_p(L_k)$ by (3.3). □

When $k=0$, $L_{k=0}=0$ and for convenience we take $L^p(\mathbb{R})$ to be the function space in this case as well (this is the set of spatially homogeneous perturbations). We note that from a physical perspective the assumption that $f\in D_p(L_k)$ imposes very mild restrictions on f. For instance, if f is piecewise continuous and decays to zero faster than $|v|^{-1}$ as $|v|\to\infty$, then $f\in D_p(L_k)$. Moreover, if we consider $D_p(L_k)$ as a set of initial conditions for the evolution equation (1.1) (as we will in Sec. IV), we see that the initial electrostatic energy, which is proportional to $\int_{-\infty}^{\infty} f(v)dv$, is finite by (3.3), but the initial kinetic energy, which is proportional to $\int_{-\infty}^{\infty} v^2 f(v)dv$, need not be finite. Hence, $D_p(L_k)$ admits infinitely many states of infinite energy. There are additional problems which arise when we study the dynamics generated by L_k on L^p; these are discussed in Sec. IV. For these reasons, we find it more convenient to work with another family of Banach spaces, which we now define.

We define a new norm on $D_p(M_v)$ called the graph norm for M_v and denoted by $\|\cdot\|_{M_v,p}$:

$$\|g\|_{M_v,p} \equiv \|g\|_p + \|M_v\, g\|_p \tag{3.8}$$

for $g\in D_p(M_v)$. By Lemma 3.1, $D_p(M_v)$ with the norm $\|\cdot\|_{M_v,p}$ is a Banach space. We call this Banach space X_p, as sets $X_p\subset L^p(\mathbb{R})$. We note that by (3.3), we could equally well consider the graph norm on $D_p(M_v)$ relative to L_k:

$$\|g\|_{L_k,p} \equiv \|g\|_p + \|L_k\, g\|_p \tag{3.9}$$

for $g\in D_p(M_v)$. However, it is a consequence of Lemma 3.2 that the two norms (3.8) and (3.9) are equivalent. For simplicity, we will consider X_p with the norm (3.8) and denote it by $\|\cdot\|_{\Gamma,p}$.

Henceforth, we will consider L_k acting on its natural domain in X_p, which we denote by $D(L_k)$; similarly, M_v acts on X_p with domain $D(M_v)$. We will also denote the operator norm on X_p by $\|\cdot\|_{\Gamma,p}$. The main technical advantage to using the spaces X_p is given in Lemma 3.4: This will also be used in Sec. IV.

Lemma 3.4: The operator B defined in (3.5) on X_p is a bounded operator and $\|B\|_{\Gamma,p}\leqslant |k|\,\|\eta\|_{\Gamma,p}$.

Proof: Let $\alpha>0$ and set $Q(\alpha)^q$ equal to the integral on the last line in (3.6). For any $g\in X_p$,

$$\|Bg\|_{\Gamma,p} = \|Bg\|_p + \|M_v Bg\|_p$$
$$\leqslant |k|\,\|\eta\|_{\Gamma,p}(Q(\alpha)\tilde\alpha)\|\,g\|_{\Gamma,p},$$

where $\tilde\alpha\equiv\max(\alpha,1)$; the result follows from this and (3.7) upon taking $\alpha=\beta^{-q}$. □

Lemma 3.5: The operators M_v and hence L_k are closed on $D(M_v)=D(L_k)$.

Proof: The equality $D(M_v)=D(L_k)$ in X_p follows just as in the discussion around (3.2). In light of Lemma 3.4 and the stability of closed operators under (relatively) bounded perturbations,[16] L_k is closed on $D(M_v)$ if and only if M_v is closed. Let $\{h_\alpha\}$ be a Cauchy sequence in $D(M_v)$ with $X_p-\lim_{\alpha\to\infty} h_\alpha = h$ and $X_p-\lim_{\alpha\to\infty} M_v\, h_\alpha = g$. By (3.8) and the closedness of M_v, this implies that $L^p-\lim_\alpha M_v h_\alpha = M_v h$. Let $k_\alpha\equiv M_v h_\alpha$. Then $L^p-\lim_\alpha k_\alpha = M_v h$ and $L^p-\lim_\alpha M_v k_\alpha = g$. Since M_v is closed in L^p, these imply that $L^p-\lim_\alpha M_v k_\alpha = M_v^2 h = g$, so that M_v is closed on $D(M_v)\subset X_p$. □

We remark that if $f\in D(L_k)$, then f has finite energy.

We now apply the method of spectral deformation described in Sec. II to the linear Vlasov operator L_k given in (3.1) on X_p. Because of the form of L_k, we find it convenient to work with the *velocity translation group* on X_p defined as follows. For each k and $\theta\in\mathbb{R}$, we define a transformation $U(\theta)$ on X_p by

$$(U(\theta)f)(v) = f(v+\theta_k), \tag{3.10}$$

where $\theta_k = \mathrm{sgn}(k)\theta$. We note that $U(\theta)$ is the restriction to the k th Fourier subspace of the group introduced in Ref. 1. It is standard to demonstrate that for each k, $\{U(\theta)|\theta\in\mathbb{R}\}$ is a strongly continuous, one-parameter group of bounded operators on X_p with $U(\theta)^{-1}=U(-\theta)$ and for any $f\in X_p$,

$$\|U(\theta)f\|_{\Gamma,p}\leqslant(1+|\theta|)\|f\|_{\Gamma,p}\,. \tag{3.11}$$

It is important to note that for $\theta\in\mathbb{R}$,

$$U(\theta)D(L_k)=D(L_k)\,, \tag{3.12}$$

as follows from (3.3), the invertibility of $U(\theta)$, and the inequality

$$\|M_v U(\theta)f\|_{\Gamma,p}\leqslant\|M_v f\|_{\Gamma,p}+(2|\theta|+|\theta|^2)\|f\|_{\Gamma,p}$$

for any $f\in D(L_k)$. Consequently, we construct a family of operators on $D(L_k)$ for $\theta\in\mathbb{R}$ by:

$$L_k(\theta)\equiv U(\theta)L_k U(\theta)^{-1}\,. \tag{3.13}$$

For any $f\in D(L_k)$, we have

$$(L_k(\theta)f)(v)=\begin{cases}-ik\Big[(v+\theta_k)f(v)+\eta(k,v+\theta_k)\\ \qquad\times\displaystyle\int_{-\infty}^{\infty}f(v')dv'\Big], & k\neq 0,\\ 0, \quad k=0\,.\end{cases} \tag{3.14}$$

Lemma 3.6: For each $\theta\in\mathbb{R}$, $L_k(\theta)$ is closed on $D(L_k)$ and $\sigma(L_k(\theta))=\sigma(L_k)$.

Proof: It is easy to check that $L_k(\theta)$ is closed on $D(L_k)$ using Lemma 3.5, (3.12), and the invertibility of $U(\theta)$. As for the spectrum, let $R_k(z,\theta)\equiv(z-L_k(\theta))^{-1}$. Then from the facts that $R_k(z,\theta)=U(\theta)R_k(z,0)U(\theta)^{-1}$ and $U(\theta)$ is invertible, it follows easily that

$$\|R_k(z,\theta)\|=\|R_k(z,0)\|\,,$$

so that $\rho(L_k)=\rho(L_k(\theta))$ and hence the spectra are equal. □

We now consider the analytic continuation of $L_k(\theta)$ onto a strip in the complex θ plane. It is clear from (3.14) that this is possible only if $\eta(k,v)$, and hence F_0, has some analyticity properties which we now indicate.

Assumption 2: F_0 is the restriction to $\mathbb{R}$ of a function (also denoted by F_0) analytic on the strip $S_0\equiv\{z|\ |\text{Im}\, z|<\theta_0\}$ for some $\theta_0>0$ and analytic on the boundary of S_0. Furthermore, for any σ, $\tau\in\mathbb{R}$ such that $|\tau|<\theta_0$, $\eta(k,z)\equiv-(\omega_p k^{-1})^2\ (\partial F_0/\partial z)$ satisfies (i) $\eta(k,\sigma+i\tau)$ and $\eta'(k,\sigma+i\tau)\in X_p$ as functions of σ [here, $\eta'(k,z)\equiv d\eta(k,z)/dz$], (ii) $\lim_{|\sigma|\to\infty}\eta(k,\sigma+i\tau)=0$ for any $|\tau|\leqslant\theta_0$, and (iii) $\sup_{|\tau|<\theta_0}\int_{-\infty}^{\infty}|\eta(k,\sigma+i\tau)|^2\,d\sigma<\infty$.

Assumption 2 will be satisfied for many model equilibria F_0, e.g., Gaussians, Lorentzians, sums of Gaussians, etc. However, equilibria F_0 that have compact support or vanish on open subsets of $\mathbb{R}$ (as discussed, for example, by Weitzner[17]) obviously fail to satisfy such an analyticity assumption. In these cases spectral deformation can still be applied using a different choice for $U(\theta)$ which requires a weaker analyticity assumption. We discuss these cases briefly in Appendix B. Conditions (ii) and (iii) of Assumption 2 will be used in Appendix A when we discuss the zeros of the dielectric functions. Note that Assumption 2 implies that $\eta(k,z)^*=\eta(k,z^*)$.

There are several notions of analyticity for operator-valued functions. We refer the reader to Kato[16] or Reed and Simon[12] for the technical details. As discussed in Sec. II, we want sufficient conditions on $L_k(\theta)$ such that $R_k(z,\theta)$ is jointly analytic on some open subset of $\mathbb{C}^2$. We prove that $L_k(\theta)$ for $\theta\in S_0$ of Assumption 2 is an analytic family of type A. This type of analytic family is the easiest with which to work.

Definition 3.7: A family of closed operators $A(\theta)$, $\theta\in N\subset\mathbb{C}$, N open and connected is an analytic family of type A on N if and only if (i) $A(\theta)$, $\theta\in N$ have a common domain of closure D independent of θ and (ii) for any $\psi\in D$, $A(\theta)\psi$ is an analytic vector-valued function on N.

Theorem 3.8: Let F_0 satisfy Assumptions 1 and 2. The family of operators $L_k(\theta)$, $\theta\in\mathbb{R}$ defined in (3.14) on $D(L_k)\subset X_p$ extends to an analytic family of type A on the strip S_0 with domain $D(L_k)$.

Proof: We check that conditions (i) and (ii) in Definition 3.7 are satisfied. By Assumption 2, $\eta(k,v)$ is analytic on S_0 and $\eta(k,v+i\theta_2)\in X_p$ as a function of v for $|\theta_2|<\theta_0$. With this assumption and (3.14), it follows that the natural domains of $L_k(\theta)$ and L_k coincide: $D(L_k(\theta))=D(L_k)$. An argument similar to the proof of Lemma 3.5 shows that $L_k(\theta)$ is closable on $D(L_k)$ and hence is closed on $D(L_k(\theta))=D(L_k)$ independent of $\theta\in S_0$. As for condition (ii), it suffices by (3.8) and (3.14) to check that $\theta\in S_0\to\eta(k,v+\theta_k)$ and $\theta\in S_0\to v\eta(k,v+\theta_k)$ are weakly L^p analytic (because weak and strong analyticity are equivalent[16]). Let $\tilde\eta(k,v)$ denote $\eta(k,v)$ or $v\eta(k,v)$; $\tilde\eta\in S(\mathbb{R})$. Let $\psi\in S_0$; by Assumption 2 and Taylor's theorem for $|\theta|$ small such that $\psi+\theta\in S_0$,

$$\tilde\eta(k,z_k+\theta_k)=\tilde\eta(k,z_k)+\theta_k\tilde\eta'(k,z_k)+R_\eta(\theta,z_k)\,,$$

where $z_k\equiv\psi_k+v$, the remainder is given by

$$R_\eta(\theta,z_k)=\frac{\theta_k^2}{2\pi i}\int_{\gamma_\sigma}\frac{\tilde\eta(k,\xi)}{(\xi-z_k)^2(\xi-z_k-\theta_k)}\,d\xi\,,$$

and γ_σ is the contour $\gamma_\sigma=\{z_k+\sigma e^{i\phi}|\ |\theta|<\sigma<\theta_0-|\text{Im}\,\psi|,\ \phi\in[0,2\pi]\}$. Now $(L^p)^*=L^q$, where $q=\infty$ if $p=1$ and satisfies $q^{-1}+p^{-1}=1$ for $p>1$. Denoting the dual pairing $f\in L^p$, $l\in L^q$ by $\int_{-\infty}^{\infty}l(v)^*f(v)dv$, we have, for any $l\in L^q$,

$$\begin{aligned}&\theta^{-1}l(\tilde\eta(k,\cdot+\psi_k+\theta_k)-\tilde\eta(k,\cdot+\psi_k))\\&=\text{sgn}(k)\int_{-\infty}^{\infty}l(v)^*\tilde\eta'(k,v+\psi_k)dv\\&\quad+\frac{\theta}{2\pi i}\int_{-\infty}^{\infty}dv\,l(v)^*\int_{\gamma_\sigma}\frac{\tilde\eta(k,\xi)d\xi}{(\xi-z_k)^2(\xi-z_k-\theta_k)}\end{aligned} \tag{3.15}$$

and by condition (i) of Assumption 2 and Hölder's inequality, it follows that each integral in (3.15) is bounded. Since the limit as $\theta\to 0$ of the integral on the rhs in (3.15) exists and is zero, it follows that the map $\theta\in S_0\to\tilde\eta(k,v+\theta_k)$ is weakly and hence strongly L^p analytic. This implies that $L_k(\theta)$ is strongly X_p analytic on S_0. □

IV. THE SPECTRUM OF $L_k(\theta)$ AND ITS DYNAMICS

We study the analytic family of operators $L_k(\theta)$, $\theta\in S_0$ acting on X_p, with domain $D(L_k)=D(M_v)$. In this section, we construct the resolvent $R_k(z,\theta)\equiv(z-L_k(\theta))^{-1}$, determine the spectrum of $L_k(\theta)$, and prove an existence and uniqueness theorem for the family of evolution equations generated by $L_k(\theta)$: $\partial_t g = L_k(\theta)g$. When $\theta = 0$, the results presented here apply to L_k and we write $R_k(z) \equiv R_k(z,0)$ for the resolvent of the Vlasov operator. We discuss $k \neq 0$ in this section, the case $k = 0$ refers to the $\lambda = 0$ eigenspace of the Vlasov operator L and consists of spatially homogeneous perturbations; we refer the reader to Ref. 1 for a discussion of the significance of this eigenspace.

By a simple calculation from (3.14), one finds that for any $f \in X_p$ and $\theta \in S_0$, the inverse of $(z - L_k(\theta))$ is given formally by $\widetilde{R}_k(z,\theta)$, where

$$(\widetilde{R}_k(z,\theta)f)(v) = (z + ik(v+\theta_k))^{-1}\left[f(v) - \frac{\eta(k,v+\theta_k)}{\Omega_k(iz/k)} \times \int_{-\infty}^{\infty} \frac{f(v')}{v'+\theta_k - iz/k}\,dv'\right] \tag{4.1}$$

for all z for which the rhs is in X_p. In (4.1), the function Ω_k is defined by

$$\Omega_k(z) \equiv 1 + \int_{\Gamma_k} (v-z)^{-1}\eta(k,v)dv\,, \tag{4.2}$$

where Γ_k is the contour $\{s + i\,\mathrm{Im}\,\theta_k \,|\, s\in\mathbb{R}\}$. This function Ω_k is called the *hybrid dielectric function*. It is discussed in detail in Appendix A, along with related functions in Vlasov theory. We shall use the following notation for the integral in (4.1):

$$H(f,\theta)(z) \equiv \int_{-\infty}^{\infty} (v'+\theta-z)^{-1}f(v')dv' \tag{4.3}$$

and $H(f) \equiv H(f,0)$.

We first show that the operator $\widetilde{R}_k(z,\theta)$ given in (4.1) is the resolvent of $L_k(\theta)$. We need the following preliminary result on the functions Ω_k and H defined in (4.2) and (4.3).

Lemma 4.1: (i) The function $\Omega_k(z)$ is analytic on $\mathbb{C}\setminus\Gamma_k$ for any θ_k with $|\mathrm{Im}\,\theta| < \theta_0$. (ii) For any $f\in L^p$ and any $\theta\in\mathbb{C}$, $H(f,\theta)(z)$ is analytic on $\mathbb{C}\setminus\{z|\mathrm{Im}\,z = \mathrm{Im}\,\theta\}$.

Proof: (i) By condition (i) of Assumption 2, the integral in (4.2) converges absolutely for $z\in\mathbb{C}\setminus\Gamma_k$ and is continuous on this set. Analyticity follows by an application of Morera's theorem: Let γ be any closed curve (of finite length) in $\mathbb{C}\setminus\Gamma_k$. Then

$$\int_\gamma \Omega_k(z)dz = \int_\gamma dz \int_{\Gamma_k} (v-z)^{-1}\eta(k,v)dv$$

and since the integrals in either order are absolutely convergent the order can be interchanged and the resulting integral vanishes by Cauchy's theorem.

(ii) To prove the second part of the Lemma, we require some simple estimates for H. Note that $|v+\theta-z| > |\mathrm{Im}(\theta-z)|$; thus for $p=1$,

$$|H(f,\theta)(z)| \leqslant |\mathrm{Im}(\theta-z)|^{-1}\|f\|_1\,. \tag{4.4}$$

For $p>1$, we have $q = p(p-1)^{-1} > 1$; thus by Hölder's inequality:

$$|H(f,\theta)(z)| \leqslant 2\|f\|_p\left[|\mathrm{Im}(\theta-z)|^{-1}R^{1/q} + \left(\int_R^\infty |v+\mathrm{Re}(\theta-z)|^{-q}dv\right)^{1/q}\right], \tag{4.5}$$

where $R > |\mathrm{Re}(\theta-z)|$ and the integral is convergent. Now we show that

$$\lim_{\epsilon\to 0}\left|\epsilon^{-1}[H(f,\theta)(z+\epsilon) - H(f,\theta)(z)] - \int_{-\infty}^{\infty}(v+\theta-z)^{-2}f(v)dv\right| = 0 \tag{4.6}$$

for any z and $|\epsilon| \leqslant 1$ such that $\mathrm{Im}\,z \neq \mathrm{Im}\,\theta$, $\mathrm{Im}(z+\epsilon) \neq \mathrm{Im}\,\theta$. This establishes the analyticity of $H(f,\theta)$. Let $\Delta H(\theta,z+\epsilon)$ denote the quotient in (4.6). Then

$$\left|\Delta H(\theta,z+\epsilon) - \int_{-\infty}^{\infty}(v+\theta-z)^{-2}f(v)dv\right| \leqslant |\epsilon|\left|\int_{-\infty}^{\infty} f(v)(v+\theta-z)^{-2}(v+\theta-z-\epsilon)^{-1}\right| \tag{4.7}$$

For $p>1$ one applies the Hölder inequality on the rhs of (4.7) (after using a partition of unity relative to the interval $[-R,R]$, $R > |\mathrm{Re}(\theta-z)|$, as in (4.5)) and shows that the resulting integral is finite as $\epsilon\to 0$. Hence (4.6) holds. For $p=1$, one simply estimates the denominators. □

Let σ_0 denote the subset of $\mathbb{C}$ defined by

$$\sigma_0(L_k(\theta)) \equiv \{\lambda \equiv -ikz|\Omega_k(z) = 0\} \cup \{\lambda\,|\mathrm{Re}\,\lambda = |k|\mathrm{Im}\,\theta\} \tag{4.8}$$

for any $\theta\in S_0$ and define ρ_0 by

$$\rho_0(L_k(\theta)) \equiv \mathbb{C}\setminus\sigma_0(L_k(\theta))\,. \tag{4.9}$$

Lemma 4.2: For any $z\in\rho_0(L_k(\theta))$ and $\theta\in S_0$, $\widetilde{R}_k(z,\theta)$ defined in (4.1) is a bounded operator on X_p and for $f\in X_p$ satisfies the bound

$$\|\widetilde{R}_k(z,\theta)f\|_{\Gamma,p} \leqslant |\mathrm{Re}\,z - |k|\mathrm{Im}\,\theta|^{-1} \times\{1+\epsilon(\theta,z)\}\|f\|_{\Gamma,p}\,, \tag{4.10a}$$

where

$$\epsilon(\theta,z) \equiv \|\eta_\theta\|_{\Gamma,p}Q(z)|\Omega_k(iz/k)|^{-1}\,,$$
$$Q(z)^q \equiv \int_{-\infty}^{\infty}\frac{ds}{|s+\theta_k - iz/k|^q}\,, \tag{4.10b}$$

and

$$\|\eta_\theta\|_p^p = \int_{-\infty}^{\infty} dv|\eta(k,v+\theta_k)|^p \tag{4.10c}$$

and similarly for $\|\eta_\theta\|_{\Gamma,p}$ (when $p=1$, $Q(z)$ is replaced by $|\mathrm{Im}\,\theta_k - \mathrm{Re}\,z|k|^{-1}|$).

Proof: This follows by simple calculation using estimate (4.4) and (4.5) and Assumption 2. Note that the estimate in (4.5) shows that $Q(z)$ is finite. □

It follows from Appendix A that $\sigma_0(L_k(\theta))$ is a closed set, so that $\rho_0(L_k(\theta))$ is an open set.

Proposition 4.3: The bounded operator $\widetilde{R}_k(z,\theta)$ defined in (4.1) is the resolvent operator $R_k(z,\theta)$ for $L_k(\theta)$, $\theta\in S_0$ on X_p and hence $\rho(L_k(\theta)) = \rho_0(L_k(\theta))$ and $\sigma_0(L_k(\theta)) = \sigma(L_k(\theta))$. In particular,

$$\sigma_{\text{ess}}(L_k(\theta)) = \{\lambda \in \mathbb{C} | \text{Re}\, z = |k| \text{Im}\, \theta\}$$

and the set of eigenvalues of $L_k(\theta)$, $\sigma_p(L_k(\theta))$, is

$$\sigma_p(L_k(\theta)) = \{\lambda \in \mathbb{C} | \Omega_k(i\lambda/k) = 0\}.$$

Proof: We first note that for any $f \in X_p$ the function $F(v,z) \equiv (v-z)^{-1}\tilde{f}(v)$, where $\tilde{f}$ denotes f or $M_v f$, is strongly L^p analytic for Im $z \neq 0$ since one easily finds for $z \in \mathbb{C}\setminus\mathbb{R}$ and all ϵ, $|\epsilon|$ small, such that $\text{Im}(z+\epsilon) \neq 0$,

$$\left\{\int_{-\infty}^{\infty} |\epsilon^{-1}(F(v,z+\epsilon) - F(v,z)) - (v-z)^{-2}\tilde{f}(v)|^p\, dv\right\}^{1/p} \leqslant |\epsilon|\,|\text{Im}\, z|^{-2}|\text{Im}(z+\epsilon)|^{-1}\|\tilde{f}\|_p.$$

Hence, $(v-z)^{-1}f(v)$, $f \in X_p$ is strongly X_p analytic on $\mathbb{C}\setminus\mathbb{R}$. It now follows by Assumption 2 on η, Lemma 4.1, and this observation that for any $\theta \in S_0$, the map $z \in \rho_0 \mapsto \tilde{R}_k(z,\theta) f$, $f \in X_p$ is strongly analytic. Next, we note that by Hölder's inequality (as in the proof of Lemma 4.2) and Assumption 2, $|\int_{-\infty}^{\infty}(\tilde{R}_k(z,\theta)f)(v)dv| < \infty$ for any $f \in X_p$ and since $v(z + ik(v+\theta_k))^{-1}$ is bounded, $\tilde{R}_k(z,\theta) f \in D(L_k)$. By a direct calculation one shows that for all $f \in X_p$, $\theta \in S_0$, and $z \in \rho_0$,

$$(z - L_k(\theta))\tilde{R}_k(z,\theta) f = f. \tag{4.11}$$

Similarly, for any $f \in D(L_k)$,

$$\tilde{R}_k(z,\theta)(z - L_k(\theta)) f = f. \tag{4.12}$$

By the analyticity of $\tilde{R}_k(z,\theta)$, the identity principle for analytic functions, and relations (4.11) and (4.12), $\tilde{R}_k(z,\theta)$ is equal to the resolvent of $L_k(\theta)$ restricted to ρ_0; thus $\rho_0 \subset \rho(L_k(\theta))$ and $(z - L_k(\theta))^{-1}$ is an analytic continuation of $\tilde{R}_k(z,\theta)$. However, it is obvious from (4.1) that $\tilde{R}_k(z,\theta)$ cannot be continued beyond ρ_0; thus $\rho_0 = \rho$ and $\tilde{R}_k(z,\theta)$ is the resolvent of $L_k(\theta)$. The eigenvalues of $L_k(\theta)$ (poles of the resolvent) come from the zeros of Ω_k and $\sigma_{\text{ess}}(L_k(\theta))$ is the complement of the set of isolated zeros of Ω_k in $\sigma(L_k(\theta))$. (It is shown in Appendix A that Ω_k has finitely many isolated zeros of finite multiplicity off the contour Γ_k.) □

The $\sigma(L_k)$ and $\sigma(L_k(\theta))$ for $\theta \in S_0^-$ are shown in Fig. 7.

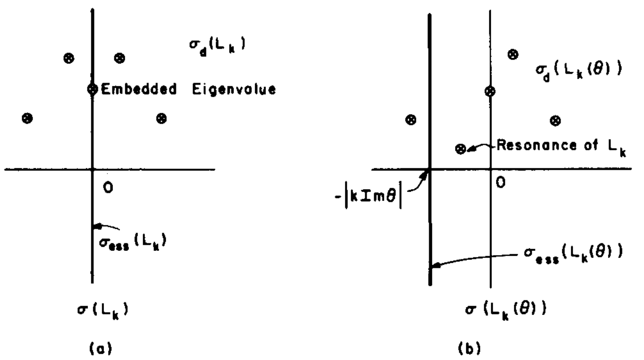


FIG. 7. The spectrum of the linearized Vlasov operator L_k, $k \neq 0$ consists of (a) the essential spectrum equal to the imaginary axis $i\mathbb{R}$ and discrete eigenvalues which occur in symmetric pairs or embedded in $i\mathbb{R}$. After a complex velocity translation, $\sigma(L_k(\theta))$, shown in (b), consists of the translated line $-|k\,\text{Im}\,\theta| + i\mathbb{R}$ and discrete eigenvalues.

Remarks 4.4: (i) The results obtained thus far in this section and in particular, Proposition 4.3, apply to the analytic family $L_k(\theta)$, $\theta \in S_0$ on the domain $D_p(M_v) \subset L^p(\mathbb{R})$, with the appropriate modification of the norms.

(ii) In light of the Lemma 3.4 and (3.14), $L_k(\theta)$ is obtained from the multiplication operator $M_{-ik(v+\theta_k)}$ by a compact (rank 1) perturbation $U(\theta)BU(\theta)^{-1}$. By a generalized Weyl theorem,[18] it then follows that $\sigma_{\text{ess}}(L_k(\theta)) = \sigma_{\text{ess}}(M_{-ik(v+\theta_k)}) = i\mathbb{R} + |k|\text{Im}\,\theta$. Hence, the effect of the complex velocity translation is to shift the essential spectrum of L_k into the left half-plane for Im $\theta < 0$ and into the right half-plane for Im $\theta > 0$.

(iii) An interesting distinction between the spectral resonances of L_k and the resonances one finds for self-adjoint operators is the symmetry of the L_k resonances with respect to $\text{Im}(\theta)$. For self-adjoint operators, resonances are only found for one sign of $\text{Im}(\theta)$; however, for the Vlasov operator a resonance for Im $\theta < 0$ implies the existence of a resonance for Im $\theta > 0$. Consider first the case $\text{Im}(\theta) < 0$; then the zeros of $\Omega_k(z)$ that lie in the strip $\{z | 0 > \text{sgn}(k)\text{Im}(z) > \text{Im}(\theta)\}$ correspond to eigenvalues of $L_k(\theta)$ which have been "uncovered" by the deformation: As discussed in Sec. II, they are spectral resonances for L_k. For $\text{Im}(\theta) > 0$, it is the roots of $\Omega_k(z)$ in the conjugate strip $\{z | 0 < \text{sgn}(k)\text{Im}(z) < \text{Im}(\theta)\}$ that locate the resonances. If we explicitly indicate the dependence of the hybrid function (4.2) on θ by writing $\Omega_k(z;\theta)$, then the symmetry of this function with respect to $\theta \to -\theta$ may be expressed as

$$\Omega_k(z;\theta)^* = \Omega_k(z^*;-\theta),$$

which follows as $\eta(k,z)^* = \eta(k,z^*)$. Thus a resonance for Im $\theta < 0$ implies a conjugate resonance for Im $\theta > 0$. In Ref. 1 particular attention was paid to the Im $\theta < 0$ resonances since these correspond to the Landau damped electrostatic modes and are significant for the initial-value problem.

The final topic in this section is the existence and uniqueness of solutions to the initial-value problem associated with $L_k(\theta)$:

$$\partial_t g = L_k(\theta) g, \quad g(t=0) \in D(M_v). \tag{4.13}$$

Following Sec. III, it is convenient to write $L_k(\theta) = A + B(\theta)$, where, for $g \in D(M_v)$,

$$(Ag)(v) \equiv -ik(M_v g)(v) \tag{4.14a}$$

and

$$(B(\theta)g)(v) \equiv -ik\left[\theta_k g(v) + \eta(k,v+\theta_k)\int_{-\infty}^{\infty} g(v')dv'\right]. \tag{4.14b}$$

We show below that A generates a one-parameter contractive semigroup on $L^p(\mathbb{R})$ or X_p. When $p=1$, $B(\theta)$ is a bounded operator on $L^1(\mathbb{R})$ and it follows from analytic perturbation theory for contractive semigroups[16,19] that $L_k(\theta)$ generates a quasibounded semigroup. When $p > 1$, however, $B(\theta)$ is not bounded on $L^p(\mathbb{R})$; by Lemma 3.2 $B(\theta)$ is only relatively A bounded. Moreover, $B(\theta)$ is not closable on $L^p(\mathbb{R})$ and, consequently, perturbation theory cannot be used. Another possibility is to use the Hille–Yosida–Phillips theorem,[19] which characterizes generators of quasibounded semigroups. Although $L_k(\theta)$ satisfies the spectra criterion, it is not clear how to compute the necessary uniform bounds on the nth power of the resolvent.

To circumvent this difficulty we solve (4.13) on the Banach space X_p. It follows from Assumption 2 and Lemma 3.4 that $B(\theta)$, $\theta\in S_0$ is bounded on X_p. To apply the perturbation theory to $L_k(\theta)$, we first note the following lemma.

Lemma 4.5: The closed operator A on $D(M_v)\subset X_p$ generates a strongly continuous, one-parameter group of isometries $U_t^{(0)}$, $t\in\mathbb{R}$, with $\|U_t^{(0)}\|_{\Gamma,p}=1$ and $U_{-t}^{(0)}=(U_t^{(0)})^{-1}$.

Proof: This proof follows from the Hille–Phillips theorem.[16] For $t\geqslant 0$, it suffices to check that (i) $(-\infty,0)\subset\rho(A)$ and (ii) $\|(\lambda+A)^{-1}\|_{\Gamma,p}\leqslant\lambda^{-1}$, $\lambda>0$. However, it is obvious that $\sigma(A)=i\mathbb{R}$; thus (i) is satisfied and as $\|(\lambda-ikM_v)^{-1}\|_{\Gamma,p}\leqslant\lambda^{-1}$ so is (ii). By symmetry, the argument extends to $t\leqslant 0$. These operators clearly form a group as $(U_t^{(0)}g)(v)=e^{ikvt}g(v)$, $t\in\mathbb{R}$, $g\in X_p$. □

To discuss $L_k(\theta)$, we use the following abstract theorem to treat the perturbation $B(\theta)$ of A.

Theorem 4.6[16]: Let T be the generator of a contractive semigroup on a Banach space X and let S be a bounded operator on X. Then $T+S$ generates a quasibounded semigroup W_t, $t\geqslant 0$ and

$$\|W_t\|\leqslant e^{t\|S\|},\quad t\geqslant 0. \tag{4.15}$$

In view of the above discussion we can apply this theorem to $A+B(\theta)$ and obtain the following proposition.

Proposition 4.7: The closed operator $L_k(\theta)$, $\theta\in S_0$ generates a strongly continuous one-parameter group of bounded operators $\{W_t\,|\,t\in\mathbb{R}\}$ on X_p satisfying

$$\|W_t\|_{\Gamma,p}\leqslant\exp\{|t|\,[\|\eta_\theta\|_{\Gamma,p}+|\theta|\,]\,|k|\}, \tag{4.16}$$

where $\|\eta_\theta\|_{\Gamma,p}=\|\eta_\theta\|_p+\|M_v\eta_\theta\|_p$ and $\|\eta_\theta\|_p$ is defined in (4.10c).

Proof: For $t\geqslant 0$, it follows immediately from Lemmas 3.4 and 4.5 and Theorem 4.6 that $A+B(\theta)$ generates a strongly continuous quasibounded semigroup $W_t^{(+)}$ (formally, $W_t^{(+)}=e^{tL_k(\theta)}$) and

$$\|W_t^{(+)}\|_{\Gamma,p}\leqslant e^{t\|B(\theta)\|_{\Gamma,p}},\quad t\geqslant 0. \tag{4.17}$$

By (4.14) and Lemma 3.4,

$$\|B(\theta)\|_{\Gamma,p}\leqslant|k|\,[\,|\theta|+\|\eta_\theta\|_{\Gamma,p}]\,. \tag{4.18}$$

For $t\leqslant 0$, we note that this is equivalent to proving that $-L_k(\theta)$ generates a semigroup for $t\geqslant 0$, which is obvious from Lemma 4.5 and Theorem 4.6; we call this semigroup $W_t^{(-)}$, $t\leqslant 0$. This satisfies a bound similar to (4.17) and (4.18), with $|t|$ instead of t. We next construct the group W_t, $t\in\mathbb{R}$. For any $t\in\mathbb{R}$, define

$$W_t\equiv\begin{cases}1, & t=0,\\ W_t^{(+)}, & t>0,\\ W_t^{(-)}, & t<0.\end{cases} \tag{4.19}$$

Then to show $\{W_t\,|\,t\in\mathbb{R}\}$ is a group, it suffices to prove that $W_t^{(+)}W_{-t}^{(-)}=W_{-t}^{(-)}W_t^{(+)}=1$ for all $t>0$. Let $L_r\equiv\pm r^{-1}(W_{\pm r}^{(\pm)}-1)$, $r>0$. Then for any $g\in D(L_k)$,

$$s-\lim_{r\to 0}L_rg=L_k(\theta)g\,.$$

Since $L_rW_{\pm t}^{(\pm)}g=W_{\pm t}^{(\pm)}L_rg$, $g\in D(L_k)$, and $W_{\pm t}^{(\pm)}$ is bounded for $t>0$, it follows that $W_{\pm t}^{(\pm)}: D(L_k)\to D(L_k)$. Consequently, for any $g\in D(L_k)$,

$$\begin{aligned}\partial_t\,W_{\pm t}^{(\pm)}g&=s-\lim_{r\to 0}W_{\pm t}^{(\pm)}L_rg=\pm L_k(\theta)W_{\pm t}^{(\pm)}g\\&=\pm W_{\pm t}^{(\pm)}L_k(\theta)g\,.\end{aligned} \tag{4.20}$$

Now we compute by (4.20) that

$$\partial_t\,W_t^{(+)}W_{-t}^{(-)}g=0\,,$$

so that $W_t^{(+)}W_{-t}^{(-)}g=g$. Since $D(L_k)$ is dense and $W_t^{(\pm)}$ are bounded, $W_t^{(+)}W_{-t}^{(-)}=1$ and similarly for $W_{-t}^{(-)}W_t^{(+)}$. Hence, (4.19) defines a group which satisfies (4.16) and is strongly continuous. □

This result allows us to construct solutions to the initial-value problem (4.13).

Theorem 4.8: For each $\theta\in S_0$ and $g\in D(L_k)\subset X_p$ there exists a unique, strongly continuous map $t\in\mathbb{R}\to g(t)\in D(L_k)$ such that for all $t\in\mathbb{R}$,

$$\partial_t g(t)=L_k(\theta)g(t) \tag{4.21}$$

(in the strong sense) and $g(t=0)=g$. Moreover, $g(t)=W_tg$, where $t\in\mathbb{R}\to W_t$ is the strongly continuous group generated by $L_k(\theta)$ and

$$\|g(t)\|_{\Gamma,p}\leqslant\exp\{|t|\,[\,\|\eta_\theta\|_{\Gamma,p}+|\theta|\,]\,|k|\}\|g\|_{\Gamma,p}\,. \tag{4.22}$$

Proof: For any $g\in D(L_k)$, set $g(t)\equiv W_tg$, where W_t is constructed in Proposition 4.7. Then $t\in\mathbb{R}\to g(t)$ is strongly continuous, $g(t=0)=g$, and (4.22) follows from (4.16). That (4.21) is satisfied is proven in (4.20). To prove uniqueness of the solution suppose that $h(s)$ is another solution of (4.21), with $h(s)\in D(L_k)$ for all $s\in\mathbb{R}$ and $h(0)=g$. Then for any t,

$$\begin{aligned}\partial_s\,W_{t-s}h(s)=&-W_{t-s}L_k(\theta)h(s)\\&+W_{t-s}L_k(\theta)h(s)=0\end{aligned}$$

by (4.20) and (4.13) for $h(s)$. Hence, $W_{t-s}h(s)=g(t)$. Applying W_{s-t} to both sides of the above equation we obtain $h(s)=W_sg$. □

V. EIGENFUNCTION EXPANSION FOR $L_k(\theta)$

We derive an expansion for any $g\in D(L_k^2)\subset X_p$ in terms of the eigenfunctions of $L_k(\theta)$ and a convergent spectral integral involving generalized eigenfunctions of $L_k(\theta)$. The derivation is based upon the inverse Laplace transform of $R_k(z,\theta)$, $z\in\rho(L_k(\theta))$, $\theta\in S_0$:

$$G(t)\equiv s-\lim_{\omega\to\infty}(2\pi i)^{-1}\int_{\gamma-i\omega}^{\gamma+i\omega}e^{zt}R_k(z,\theta)g\,dz\,, \tag{5.1}$$

which is proved to be strongly convergent for all $g\in D(L_k^2)$ and $t>0$. The map $t\in\mathbb{R}^+\to G(t)$ extends to a strongly continuous map on $[0,\infty)$ and is shown to be a solution of the evolution equation (4.21) with the initial condition g. Hence, by the uniqueness part of Theorem 4.8, $G(t)=W_tg$, where W_t is the evolution group constructed in Proposition 4.7.

We first consider the integral in (5.1). We will work with $\theta\in S_0^-\equiv\{\theta\,|\,-\theta_0<\operatorname{Im}\theta<0\}$, so that $\sigma_{\rm ess}(L_k(\theta))$ lies in the left half-plane. By Proposition 4.3 and Corollary A4 of Appendix A, $\sigma(L_k(\theta))$ lies to the left of the contour

$\Gamma = \{\gamma + ir | r \in \mathbb{R}\}$ for any $\gamma > |k| \, \|\eta\|_1$; see Fig. 8. Let $\Gamma_\omega \equiv \{\gamma + ir| \, |r| \leqslant \omega\}$.

Lemma 5.1: For any $g \in D(L_k^2)$, $t > 0$, $\theta \in S_0^-$ and for γ such that $|k| \, \|\eta\|_1 < \gamma < \infty$, the limit on the rhs of (5.1) exists. Furthermore, the map $t \in \mathbb{R}^+ \to G(t)$ extends to a strongly continuous map on $[0, \infty)$ and $G(0) = g$.

Proof: From the identity

$$1 = (z - L_k(\theta))R_k(z,\theta) = zR_k(z,\theta) - L_k(\theta)R_k(z,\theta)$$

we obtain

$$R_k(z,\theta) = z^{-1} + z^{-1}L_k(\theta)R_k(z,\theta) .$$

Iterating this identity once more yields

$$R_k(z,\theta) = z^{-1} + z^{-2}L_k(\theta) + z^{-2}L_k(\theta)^2 R^k(z,\theta) . \quad (5.2)$$

Let $\gamma > |k| \, \|\eta\|_1$, $\omega > 0$, $t > 0$ and consider the integral in (5.1). Using (5.2), we see that there are three integrals over Γ_ω to evaluate. By contour deformation, the first integral is

$$\lim_{\omega \to \infty} (2\pi i)^{-1} \left(\int_{\Gamma_\omega} e^{zt} z^{-1} \, dz \right) g = g . \quad (5.3)$$

Similarly, for the second integral we obtain

$$\lim_{\omega \to \infty} (2\pi i^{-1}) \left(\int_{\Gamma_\omega} e^{zt} z^{-2} \, dz \right) L_k(\theta) g = tL_k(\theta) g . \quad (5.4)$$

Let $I(\omega)$ denote the third integral:

$$I(\omega) \equiv (2\pi i)^{-1} \int_{\gamma - i\omega}^{\gamma + i\omega} e^{zt} z^{-2} R_k(z,\theta) L_k(\theta)^2 g \, dz .$$

We show that $\{I(\omega) | \omega \in \mathbb{R}^+\}$ forms a Cauchy sequence. We use the estimate on the resolvent from (4.10):

$$\|R_k(z,\theta)\|_{\Gamma,p} \leqslant (\gamma + |k \operatorname{Im} \theta|)^{-1} \{1 + \epsilon(\theta,z)\}$$

since $\theta \in S_0^-$ and $z \in \Gamma_\omega$. Using the bound

$$|\Omega_k(iz/k)| \geqslant 1 - \gamma^{-1}|k| \, \|\eta_\theta\|_1$$

for $z \in \Gamma_\omega$ and evaluating (4.10b) for $Q(z)$, $z \in \Gamma_\omega$, it follows that $\epsilon(\theta,z)$ depends only on $\gamma = \operatorname{Re} z$. Since $L_k(\theta)^2 R_k(z,\theta) g = R_k(z,\theta) L_k(\theta)^2 g$, we obtain, for $\omega > \omega' > 0$,

$$\begin{aligned} &\|I(\omega) - I(\omega')\| \\ &\leqslant (2\pi)^{-1} \left\| \left(\int_{\Gamma_\omega} - \int_{\Gamma_{\omega'}} \right) e^{zt} z^{-2} R_j(z,\theta) L_k(\theta)^2 g \, dz \right\| \\ &\leqslant K e^{\gamma t} (\gamma + |k \operatorname{Im} \theta|)^{-1} \|L_k(\theta)^2 g\|_{\Gamma,p} \left(\int_{\omega'}^{\omega} \frac{ds}{\gamma^2 + s^2} \right) \end{aligned} \quad (5.5)$$

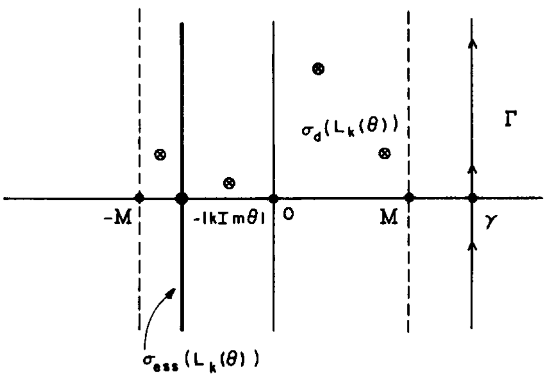


FIG. 8. The contour Γ for the inverse Laplace transform (5.1) is chosen to lie to the right of $\sigma(L_k(\theta))$, $\operatorname{Im} \theta < 0$.

for some $K > 0$ depending only on η, γ, and $\operatorname{Im} \theta_k$. As

$$\int_{\omega'}^{\omega} \frac{ds}{\gamma^2 + s^2} = \gamma^{-1} \left[\tan^{-1}\left(\frac{\omega}{\gamma}\right) - \tan^{-1}\left(\frac{\omega'}{\gamma}\right) \right]$$

we see that the sequence is Cauchy by the properties of $\tan^{-1} u$. Hence, $s - \lim_{\omega \to \infty} I(\omega)$ exists and we have

$$\lim_{\omega \to \infty} \|I(\omega)\| \leqslant \gamma^{-1} e^{t\gamma} M_1 \|L_k(\theta)^2 g\|_{\Gamma,p} .$$

Hence, it follows from (5.3)–(5.5) that the strong limit of the integral in (5.1) exists and is bounded above by

$$\begin{aligned} \|G(t)\|_{\Gamma,p} \leqslant \|g\|_{\Gamma,p} &+ |t| \, \|L_k(\theta) g\|_{\Gamma,p} \\ &+ \gamma^{-1} e^{t\gamma} M_1 \|L_k(\theta)^2 g\|_{\Gamma,p} . \end{aligned} \quad (5.6)$$

It is clear from (5.3)–(5.5) that $G(t)$ can be extended to $t = 0$. For $t = 0$, the integral (5.5) can be made arbitrarily small by taking γ large [by (4.10b), K is a decreasing function of γ], so that $G(0) = g$. The strong continuity of $t \in [0, \infty) \to G(t)$ follows from the identity

$$\begin{aligned} G(t) = g + tL_k(\theta) g + (2\pi i)^{-1} \int_\Gamma e^{zt} z^{-2} R_k(z,\theta) \\ \times L_k(\theta)^2 g \, dz \end{aligned} \quad (5.7)$$

and the continuity of each term [the continuity of the integral follows from the majorization in (5.6)]. □

Lemma 5.2: For any $g \in D(L_k^3)$, the map $t \in [0, \infty) \to G(t)$ defined in (5.1) is strongly differentiable and satisfies

$$\partial_t G(t) = L_k(\theta) G(t), \quad G(t = 0) = g$$

and, consequently, $G(t) = W_t g$.

Proof: Let $g \in D(L_k^3)$. From identity (5.7) we have

$$\begin{aligned} \partial_t G(t) = L_k(\theta) g + (2\pi i)^{-1} \int_\Gamma e^{zt} z^{-1} \\ \times R_k(z,\theta) L_k(\theta)^2 g \, dz \end{aligned} \quad (5.8)$$

since the integral in (5.7) is strongly differentiable in t. This can be seen by computing the strong derivative directly. [The integral in (5.8) converges by the same argument given in the proof of Lemma 5.1.] Next, note the identity for $t > 0$ used in (5.3):

$$\lim_{\omega \to \infty} (2\pi i)^{-1} \int_{\Gamma_\omega} z^{-1} e^{zt} = 1 ,$$

so that the rhs of (5.8) can be written as

$$\begin{aligned} &(2\pi i)^{-1} \int_\Gamma [z^{-1} + z^{-1} R_k(z,\theta) L_k(\theta)] e^{zt} L_k(\theta) g \, dz \\ &= (2\pi i)^{-1} \int_\Gamma R_k(z,\theta) e^{zt} L_k(\theta) g \, dz \\ &= L_k(\theta) G(t) . \end{aligned}$$

The last equality follows by the fact that $L_k(\theta)$ is closed. That $G(t)$ can be extended to $t = 0$ and $G(0) = g$ was proven in Lemma 5.1. Finally, by the uniqueness part of Theorem 4.8, $G(t) = W_t g$. □

Remark 5.3: Although the contour integral (5.1) provides a solution to the initial-value problem (4.13) associated with $L_k(\theta)$, it is difficult to establish from (5.1) that

$L_k(\theta)$ generates a one-parameter group and that $G(t)$ satisfies the exponential bound (4.22).

We are now in a position to construct the eigenfunction expansion for $G(t)$, $g \in D(L_k^2)$ by deforming the contour $\Gamma \equiv \{\gamma + ir | r \in \mathbb{R}\}$ in the integral (5.1). We will assume that F_0 satisfies Assumptions 1 and 2. Moreover, without additional information on the behavior of $\eta(k,z)$ on the boundary of its domain of analyticity $\text{Im}\, z = +\theta_0$, we must limit the maximum complex translation such that $|\text{Im}\, \theta| < \theta_0 - \delta$ for some small $\delta > 0$. Then by the results of Appendix A, $\sigma(L_k(\theta))$ consists of finitely many isolated eigenvalues of finite multiplicity and the line $\{z | \text{Re}\, z = -|k| \text{Im}\, \theta\}$ (on which there may be at most finitely many embedded eigenvalues). Since we are interested in the contribution from the damped modes, we will take $\theta \in S_0^-$.

The eigenfunction expansion that we derive here expresses $W_t g$ for $t > 0$ in terms of a finite sum over the eigenfunctions of $L_k(\theta)$ (including those corresponding to the embedded eigenvalues) and an integral over the essential spectrum of $L_k(\theta)$. This integral is shown to converge absolutely in X_p and can be interpreted in terms of distributional solutions to the eigenvalue equation for $L_k(\theta)$, i.e., generalized eigenfunctions. The weights for the eigenfunctions are given by the dual pairing between $X_p^* \supset L^q(\mathbb{R})$ and X_p, which we write as $(f,l) \in X_p \times X_p^* \mapsto \langle l,f \rangle$ and with $\lambda \in \mathbb{C}$, $\langle \lambda l, f\rangle = \lambda^* \langle l,f\rangle$ [i.e., an extension of the dual pairing for L^p and $L^q = (L^p)^*$ introduced above].

The eigenfunctions and generalized eigenfunctions of $L_k(\theta)$, $k \neq 0$ are derived in Ref. 1 and we simply list them here. Let $z \equiv i\lambda k^{-1}$, where $L_k(\theta)\psi = \lambda\psi$. Let $\{\lambda_j\}_{i=1}^N$ denote the eigenvalues of $L_k(\theta)$ and let $z_j = i\lambda_j k^{-1}$. The isolated eigenvalues of $L_k(\theta)$ are given by the roots $\Omega_k(z_j) = 0$ which satisfy $\text{Im}(\theta_k - z_j) \neq 0$ and the corresponding eigenfunctions are

$$\psi_{k,j}(v) = -\eta(k,v+\theta_k)/(v+\theta_k - z_j) . \tag{5.9}$$

The embedded eigenvalues satisfy $\Omega_k^{(\pm)}(z_j) = 0$ and $\text{Im}(z_j - \theta_k) = 0$, where $\Omega_k^{(\pm)}(z) \equiv \lim_{\epsilon \to 0^+} \Omega_k(z \pm \epsilon)$. The eigenfunctions have the same form as in (5.9):

$$\psi_{k,j}(v) = -\eta(k,v+\theta_k)/(v+\theta_k - z_j) . \tag{5.10}$$

Note that $\Omega_k^{(\pm)}(z_j) = 0$ implies $\eta(k,z_j + \theta_k) = 0$; thus $\psi_{k,j}$ is not singular. Note that by Assumption 2, $\psi_{k,j} \in X_p$. The generalized eigenfunctions corresponding to points with $\text{Im}\, z = \text{Im}\, \theta_k$, $\Omega_k^{(+)}(z)$ or $\Omega_k^{(-)}(z)$, or both nonvanishing are given by

$$\psi_{k,z}(v) = P[-\eta(k,v+\theta_k)/(v+\theta_k - z)] + \omega_k(z)\delta(v+\theta_k - z) , \tag{5.11}$$

where $\lim_{\epsilon\to 0^+} \Omega^{(\pm)}(z \pm \epsilon) = \omega_k(z) \pm i\pi\eta(k,z)$ for $\text{Im}\, z = \text{Im}\, \theta_k$ and P denotes the Cauchy principal value.

The action of $L_k(\theta)$ on X_p induces an adjoint operator $L_k(\theta)^+$ acting on the dual space X_p^*: It is densely defined and closable.[16] By a simple calculation from (3.14), we find

$$(L_k(\theta)^+ g)(v) = ik\left[(v+\theta_k^*)g(v) + \int_{-\infty}^{\infty} \eta(k,v'+\theta_k^*)g(v')dv'\right] \tag{5.12}$$

for $g \in D(M_v) \subset X_p^*$. As above, we list the eigenfunctions for $L_k(\theta)^+$ using the scaled eigenvalues $z^* = i\lambda^* k^{-1}$. As in Ref. 1, we write the eigenvalue equation as $L_k(\theta)^+\zeta_k = \lambda^*\zeta_k$. The isolated eigenvalues $\{\lambda_j^*\}_{j=1}^{N_1}$ satisfy $\text{Im}(\theta_k - z_j) \neq 0$, $\Omega_k(z_j^*) = 0$ and the eigenfunctions are

$$\zeta_{k,j}(v) = (1/\Omega_k'(z_j)^*)[-1/(v+\theta_k^* - z_j^*)] . \tag{5.13}$$

The embedded eigenvalues $\{\lambda_j^*\}$ satisfy $\Omega_k^{(\pm)}(z_j^*) = 0$, $\text{Im}(\theta_k - z_j) = 0$. There are two linearly independent eigenfunctions

$$\zeta_{k,j}^{(1)}(v) = \delta(v+\theta_k^* - z_j^*) , \tag{5.14a}$$

$$\zeta_{k,j}^{(2)}(v) = P[-c^{(2)}(k,\theta)/(v+\theta_k^* - z^*)] , \tag{5.14b}$$

where the normalization constant $c^{(2)} \neq 0$. The generalized eigenfunctions corresponding to $\lambda^* \equiv iz^*k^{-1} \in \sigma_{\text{ess}}(L_k(\theta))$, $\text{Im}\, z = \text{Im}\, \theta_k$ are

$$\zeta_{k,z}(v) = \left(\frac{1}{\Omega_k^{(+)}(z)\Omega_k^{(-)}(z)}\right)^* \times \{P[-\eta(k,z)^*/(v+\theta_k^* - z^*)] + \omega_k(z)^*\delta(v+\theta_k^* - z^*)\} , \tag{5.15}$$

where ω_k is defined after (5.11). We can now state the main theorem.

Theorem 5.3: Let F_0 satisfy Assumptions 1 and 2. Let $g \in D(L_k^3)$ and $\theta \in S_0^-$, with $\text{Im}\, \theta > -\theta_0 + \delta$ for some $\delta > 0$ small. Let $g(t) \equiv W_t g$ be the solution to the evolution equation $\partial_t g(t) = L_k(\theta) g(t)$ as described in Theorem 4.8, with the initial condition g. Let $\{\lambda_i\}_{i=1}^N$ be the eigenvalues of $L_k(\theta)$ (which we assume for convenience are simple). Then there exist constants $c_j(g,\theta)$ given by

$$c_j(g,\theta) \equiv \langle \zeta_{k,j}, g\rangle \tag{5.16}$$

and a distribution $A_z(g,\theta)$ with $\text{Im}\, z = \text{Im}\, \theta$ and

$$A_z(g,\theta) \equiv (i|k|)^{-1}\langle \zeta_{k,z}, g\rangle \tag{5.17}$$

such that for all $t \geqslant 0$,

$$(W_t g)(v) = \sum_{j=1}^N e^{\lambda_j t} c_j(g,\theta)\psi_{k,j}(v) + \int_{-i\infty - |k \,\text{Im}\, \theta|}^{i\infty - |k\, \text{Im}\, \theta|} e^{zt} A_{iz/k}(g,\theta)\psi_{k,iz/k}(v)dz . \tag{5.18}$$

The integral is absolutely convergent and defines an element of X_p.

Proof: (i) By the hypotheses, it is shown in Appendix A that there exists $\gamma > 0$ such that $\sigma(L_k(\theta)) \subset \{z | \text{Re}\, z < \gamma\}$. Hence, $W_t g \equiv g(t)$ is given by the contour integral in (5.1) for $t > 0$. Choose ϵ, $0 < \epsilon \ll 1$ such that $\sigma(L_k(\theta)) \cap \{z | -|k\, \text{Im}\, \theta| < \text{Re}\, z < -|k\, \text{Im}\, \theta| + \epsilon\} = \phi$; see Fig. 9. For any $R > 0$, let $\Gamma_{R,\epsilon}$ be the closed rectangular contour, with the sides $\Gamma_{1,R} \equiv \{\gamma + iy | -R \leqslant y \leqslant R\}$ and $\Gamma_{2,R} \equiv \{-|k\, \text{Im}\, \theta| + \epsilon + iy | -R \leqslant y \leqslant R\}$ and the ends $E_R^\pm \equiv \{x \pm iR \mid -|k\, \text{Im}\, \theta| + \epsilon \leqslant x \leqslant \gamma\}$. We can take R so large

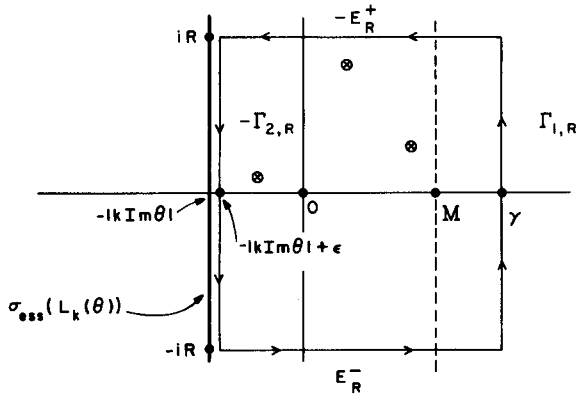


FIG. 9. The contour $\Gamma_{R,\epsilon} = \Gamma_{1,R} \cup (-E_R^+) \cup (-\Gamma_{2,R}) \cup E_R^-$ appearing in the contour integral (5.19); ϵ and R are chosen such that all eigenvalues of $L_k(\theta)$ with real parts greater than $-|k \operatorname{Im} \theta|$ lie within the contour.

that all the eigenvalues of $L_k(\theta)$, $\{\lambda_j\}_{j=1}^{N_1}$, with $\operatorname{Re}\lambda_j > -|k \operatorname{Im}\theta|$, lie within $\Gamma_{R,\epsilon}$. Let

$$I(z) \equiv (2\pi i)^{-1} e^{zt} R_k(z,\theta) g .$$

Then by the residue theorem,

$$\int_{\Gamma_{R,\epsilon}} I(z)dz = \sum_{j=1}^{N_1} e^{\lambda_j t} \operatorname{Res}(I(z), i\lambda_j k^{-1})$$
$$= \left(\int_{\Gamma_{1,R}} - \int_{\Gamma_{2,R}} - \int_{E_R^+} + \int_{E_R^-} \right) I(z)dz , \quad (5.19)$$

where $\operatorname{Res}(f,z_0)$ is the residue of f at z_0. By Lemma 5.1,

$$\lim_{R\to\infty} \int_{\Gamma_{1,R}} I(z)dz = g(t) . \quad (5.20)$$

We show that

$$\lim_{R\to\infty} \int_{E_R^\pm} I(z)dz = 0 . \quad (5.21)$$

Since $g \in D(L_k^3)$ we can use identity (5.2) in (5.21) and show that each integral vanishes as $R\to\infty$. For the first integral, we have

$$\left|\left| \int_{E_R^\pm} z^{-1} e^{zt} g \, dz \right|\right|_{\Gamma,p}$$
$$\leqslant R^{-1} e^{\gamma t} (\gamma + |k \operatorname{Im}\theta| - \epsilon) \|g\|_{\Gamma,p} \quad (5.22a)$$

and for the second,

$$\left|\left| \int_{E_R^\pm} z^{-2} e^{zt} L_k(\theta) g \right|\right|_{\Gamma,p}$$
$$\leqslant R^{-2} e^{\gamma t} (\gamma + |k \operatorname{Im}\theta| - \epsilon) \|L_k(\theta) g\|_{\Gamma,p} . \quad (5.22b)$$

Equations (5.22a) and (5.22b) vanish in the limit $R\to\infty$. For the third integral, we must estimate $\|R_k(z,\theta)\|$, $z \in E_R^\pm$. From (4.10), it follows that for all R sufficiently large,

$$\|R_k(z,\theta)\|_{\Gamma,p} \leqslant \epsilon^{-1} M(\eta,\theta,\epsilon)$$

for some constant M depending on η, $\operatorname{Im}\theta$, and ϵ, but independent of R. Hence, we have

$$\left|\left| \int_{E_R^\pm} z^{-2} e^{zt} R_k(z,\theta) L_k(\theta)^2 g \, dz \right|\right|_{\Gamma,p}$$
$$\leqslant R^{-2} \epsilon^{-1} e^{\gamma t} M (\gamma + |k \operatorname{Im}\theta| - \epsilon) \|L_k(\theta)^2 g\|_{\Gamma,p} . \quad (5.22c)$$

From (5.22a)–(5.22c), we see that (5.21) holds. Consequently, (5.21) and (5.20) imply, by virtue of (5.19), that

$$W_t g = \lim_{R\to\infty} \frac{1}{2\pi i} \int_{\Gamma_{2,R}} e^{zt} R_k(z,\theta) g \, dz$$
$$+ \sum_{j=1}^{N_1} e^{\lambda_j t} \operatorname{Res}(I(z), i\lambda_j k^{-1}) . \quad (5.23)$$

(ii) With reference to Fig. 10, we define the following contours. Let Γ_ϵ^+ denote $\lim_{R\to\infty} \Gamma_{2,R}$ and on the lhs of $\{z | \operatorname{Re} z = -|k \operatorname{Im}\theta|\}$, we define $\Gamma_{\epsilon_1}^- \equiv \{-|k \operatorname{Im}\theta| - \epsilon_1 + iy | y \in \mathbb{R}\}$ and $\Gamma_{3,R} \equiv \{-|k \operatorname{Im}\theta| - \epsilon_1 + iy | -R \leqslant y \leqslant R\}$; for $\omega > 0$ let $\Gamma_{\omega,R} \equiv \{-\omega + iy | -R \leqslant y \leqslant R\}$ and $E_{\omega,R}^\pm \equiv \{x \pm iR \mid -\omega \leqslant x \leqslant -|k \operatorname{Im}\theta| - \epsilon_1\}$. Let $\{\lambda_j\}_{j=N_1+1}^{N_2}$ denote the eigenvalues of $L_k(\theta)$ lying in $\{z | \operatorname{Re} z < -|k \operatorname{Im}\theta|\}$; we assume $\epsilon_1 > 0$ is chosen small enough such that $\sigma(L_k(\theta)) \cap \{z \mid -|k \operatorname{Im}\theta| - \epsilon_1 \leqslant \operatorname{Re} z \leqslant -|k \operatorname{Im}\theta|\} = \phi$. We now denote by ϵ the minimum of ϵ_1 and ϵ from part (i). By taking ω and R sufficiently large we have, by the residue theorem,

$$\sum_{j=N_1+1}^{N_2} \operatorname{Res}(I(z), i\lambda_j k^{-1})$$
$$= \left(\int_{\Gamma_{3,R}} - \int_{\Gamma_{\omega,R}} + \int_{E_{\omega,R}^-} - \int_{E_{\omega,R}^+} \right) I(z)dz . \quad (5.24)$$

We show that for fixed R,

$$\lim_{\omega\to\infty} \int_{\Gamma_{\omega,R}} I(z)dz = 0 \quad (5.25)$$

and then that

$$\lim_{R\to\infty} \lim_{\omega\to\infty} \int_{E_{\omega,R}^\pm} I(z)dz = 0 . \quad (5.26)$$

To prove (5.25), we have

$$\lim_{\omega\to\infty} \left|\left| \int_{\Gamma_{\omega,R}} I(z)dz \right|\right|_{\Gamma,p}$$
$$\leqslant \lim_{\omega\to\infty} 2\,\mathrm{R}e^{-\omega t} \left[\sup_{z \in [\omega - iR, \omega + iR]} \|R_k(z,\theta)\|_{\Gamma,p} \right]$$
$$\times \|g\|_{\Gamma,p} = 0 \quad (5.27)$$

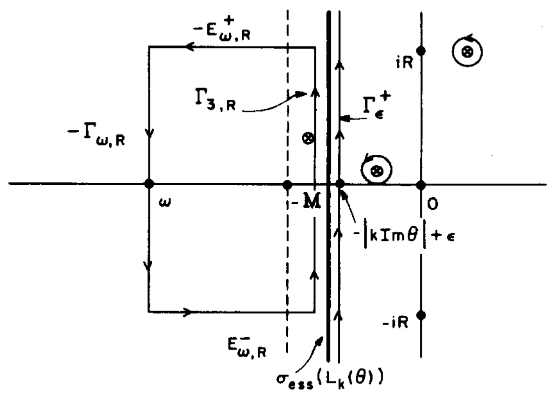


FIG. 10. The contours lying on the lhs of $\sigma_{ess}(L_k(\theta))$ used in identity (5.24).

since $(\sup\|R_k(z,\theta)\|_{\Gamma,p}) = 0(\omega^{-1})$ by (4.10). As for (5.26), we use identity (5.2) and show that each integral vanishes. For the first integral we have

$$\left\|\int_{E^{\pm}_{\omega,R}} z^{-1} e^{zt} g\, dz\right\|_{\Gamma,p}$$

$$\leqslant R^{-1}\int_{|k\,\mathrm{Im}\,\theta|-\epsilon}^{-\omega} e^{st}\, ds \|g\|_{\Gamma,p}$$

$$\leqslant (Rt)^{-1}[e^{-(|k\,\mathrm{Im}\,\theta|+\epsilon)t} - e^{-\omega t}]\|g\|_{\Gamma,p} \qquad (5.28)$$

and a similar estimate holds for the second integral. For the third integral, we again use the estimate on the resolvent (4.10) and obtain $\|R_k(z,\theta)\|_{\Gamma,p} = O(\epsilon^{-1})$ for $z \in E^{\pm}_{\omega,R}$ [as after (5.22b)]. Hence, (5.26) holds.

(iii) We now combine (5.24)–(5.28) with (5.23) to obtain the following representation of $W_t g$:

$$W_t g = \sum_{j=1}^{N_2} \mathrm{Res}(e^{zt}R_k(z,\theta)g, i\lambda_j k^{-1})$$

$$+\left(\int_{\Gamma_\epsilon^+} - \int_{\Gamma_\epsilon^-}\right) e^{zt}R_k(z,\theta)g\, dz\,, \qquad (5.29)$$

where both integrals converge in the strong X_p sense and the sum is over all isolated eigenvalues of $L_k(\theta)$, i.e., $\mathrm{Re}(z_j) \neq -|k\,\mathrm{Im}\,\theta|$. That the coefficients $c_j(g,\theta)$ for these eigenvalues are given as in (5.16) (assuming that they are simple) can be easily seen by evaluating the residues and using formulas (5.9) and (5.13). The contour integrals in (5.29) can be evaluated by first taking the limit as $\epsilon \to 0^+$ for R sufficiently large and then the limit as $R \to \infty$. This argument is presented in Ref. 1 for the case of no embedded eigenvalues and can easily be extended to include embedded eigenvalues. As shown in Ref. 1, the $\epsilon \to 0^+$ limit requires the existence of boundary values of the hybrid dielectric function $\Omega_k(r + \theta_k \pm i\epsilon)$, $r \in \mathbb{R}$. By Assumption 2 and the condition that $\mathrm{Im}\,\theta > -\theta_0 + \delta$, the boundary values exist and are continuous on $\Gamma_k \equiv \{-|k\,\mathrm{Im}\,\theta| + iy | y \in \mathbb{R}\}$.[20] □

Remark 5.4: (i) Let f be a $U(\theta)$ analytic vector and define $g \equiv U(\theta) f$. Let $g(v,t)$ be the solution to the evolution equation (4.13), with the initial condition g. In Ref. 1, it was shown that $U(\theta)^{-1} g(v,t)$ is equal to $f(v,t)$, the solution to the evolution equation for L_k with the initial condition f. One can obtain another proof of this result using the eigenfunction expansion (5.18) directly. If $U(\theta)^{-1}$ is applied to the finite sum in (5.18), one finds that the contribution from the zeros of Ω_k with non-negative real parts is identical to that obtained in the van Kampen–Case expansion for f (unstable and neutral modes). The contributions from those zeros of Ω_k with $-|k\,\mathrm{Im}\,\theta| < \mathrm{Re}\, z < 0$ correspond to the Landau resonances and do not appear in the van Kampen–Case expansion. The amplitude of these modes are the same (independent of θ) as in the usual Landau theory. What is interesting is that evaluation of

$$U(\theta)^{-1}\int_{-i\infty - |k\,\mathrm{Im}\,\theta|}^{i\infty - |k\,\mathrm{Im}\,\theta|} e^{zt} A_{iz/k}(g,\theta)\psi_{k,iz/k}(v)$$

by contour deformation yields three contributions. One contribution precisely cancels the finite sum over the Landau resonances. Another contribution provides the contribution to $f(v,t)$ from the stable van Kampen–Case eigenfunctions ($\mathrm{Re}\, z < 0$) and one contribution is the van Kampen–Case spectral integral for L_k over the real axis. Hence, one recovers the van Kampen–Case eigenfunction expansion for $f(v,t)$.

(ii) By applying $U(\theta)^{-1}$ to $g(v,t)$ as described in (i), we obtain an expansion for $f(v,t)$ in which the physically observed Landau damped waves appear explicitly in the sum over resonances: This has the form

$$f(v,t) = \sum_{\substack{\Lambda_k(z_j)=0 \\ \mathrm{Re}\, z_j > 0}} c_j(f) e^{-ikz_j t}\psi_{k,j}$$

$$+ \sum_{\substack{\epsilon_k(z_1)=0 \\ -|k\,\mathrm{Im}\,\theta| < \mathrm{Re}\, z_1 < 0}} d_1(f) e^{-ikz_j t}\psi_{k,1}$$

$$+ \int_{-i\infty - |k\,\mathrm{Im}\,\theta|}^{i\infty - |k\,\mathrm{Im}\,\theta|} A_{iz/k}(U(\theta) f,\theta)$$

$$\times U(\theta)^{-1}\psi_{k,iz/k}(v)\,, \qquad (5.30)$$

where

$$\psi_{k,j}(v) = \eta(k,v)/(v - z_j)\,.$$

Note that the rate of decay of the integral in (5.30) is faster than that for the Landau resonances. Hence, in the absence of zeros of Λ_k with positive real part, the Landau resonances determine the time-asymptotic behavior of $f(v,t)$. Comparison of the results discussed here with the van Kampen–Case expansion described in Ref. 1 shows that the spectral deformation method provides for a decomposition of the van Kampen–Case spectral integral in such a way that some of the resonances are exhibited. The number of resonances appearing depends upon the size of the complex velocity translation (and hence the analyticity of η and f).

(iii) Weitzner[17] has observed that the time-asymptotic behavior of the electrostatic potential $\phi_k(t)$ or electric field $E_k(t)$, both of which are proportional to $\int f(v,t)dv$, is dominated by terms $O(t^{-N})$, rather than an exponential, if the initial perturbation f is not analytic in a strip around the real axis. We note that such vectors in X_p are not $U(\theta)$ analytic vectors. Indeed, for such nonanalytic vectors, (5.30) is not valid. Recalling the normalization of eigenfunctions, (5.30) predicts exponential decay of the electric field $E_k(t)$ for stable equilibria [all $c_j(f) = 0$].

(iv) The electric field amplitude $E_k(t)$ computed from (5.30) agrees exactly with that computed using the one-sided Laplace transform method as utilized by Landau.[6] This can be seen by evaluating the coefficients $c_j(f)$ and $d_k(f)$ as in Theorem 5.3, using the normalization $\int \psi_{k,j}(v)dv = 1$, and evaluating the integral over v of the integral in (5.30). Hence, we easily obtain the usual results using this new expansion for f.

(v) We mention related results obtained by Trocheris[21,22] and Degond.[23] Trocheris introduces a modification of the linearized Vlasov operator L such that the damped modes appear as true eigenfunctions of this modified operator. This modification is obtained by redefining the velocity integral in Poisson's equation. Degond considers the inverse Laplace transform formula (5.1) on $L^1(R \times R)$ and deforms the contour of integration into the left-half plane after defining a meromorphic continuation of the resolvent of L_k into that region: Hence, he obtains an expansion for f similar to (5.30).

ACKNOWLEDGMENTS

This work was begun when P. H. was at the Mathematics Department, University of California, Irvine, which provided some travel support for this collaboration. The work was completed during two visits by P. H. to the UCSD Institute for Nonlinear Science which were supported by DARPA University Research Initiative Grant No. N00014-86-K-0758 and NSERC Research Grant No. A7901.

P. H. thanks I. M. Sigal and M. Krishna for useful conversations and H. Abarbanel for the hospitality of the Institute for Nonlinear Science.

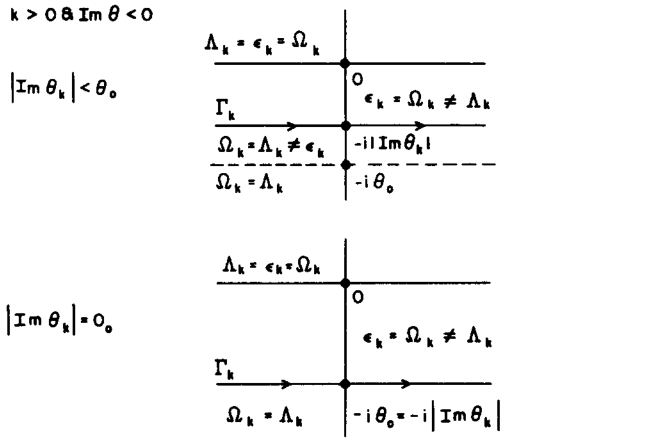


FIG. 12. An illustration of the various relations between the functions Ω_k, ϵ_k, and Λ_k, depending on Im z, for the case of sgn(k) > 0 and Im $\theta<0$.

APPENDIX A: DIELECTRIC FUNCTIONS

Three related analytic functions appear in our discussion of the linearized Vlasov operator (see, also, Ref. 1) and we present some properties of these functions in this Appendix. Let $\eta(k,v)$ be as defined in (1.3). We assume that the equilibrium distribution F_0 and η satisfy Assumptions 1 and 2. We assume $k\neq 0$ and let $p\equiv\mathrm{sgn}(k)$ and $\theta_k\equiv p\theta$, as above.

Definition A1: (i) The Case–van Kampen spectral function $\Lambda_k(z)$ is defined for any $z\in\mathbb{C}\setminus\mathbb{R}$ by

$$\Lambda_k(z)\equiv 1+\int_{-\infty}^{\infty}(v-z)^{-1}\eta(k,v)dv\,.$$

(ii) The Landau dielectric function $\epsilon_k(z)$ is defined for any z with $p\,\mathrm{Im}\,z>-\theta_0$, where θ_0 is given in Assumption 2 by

$$\epsilon_k(z)\equiv 1+\int_L (v-z)^{-1}\eta(k,v)dv\,,$$

where L is the Landau contour as shown in Fig. 11.

(iii) The hybrid dielectric function $\Omega_k(z)$ is defined for any z with $\mathrm{Im}\,z\neq\mathrm{Im}\,\theta_k$ by

$$\Omega_k(z)\equiv 1+\int_{\Gamma_k}(v-z)^{-1}\eta(k,v)dv\,,$$

where $\Gamma_k\equiv\{v\in\mathbb{C}|\mathrm{Im}\,v=\mathrm{Im}\,\theta_k\}$ (see Fig. 11) and $|\mathrm{Im}\,\theta_k|\leqslant\theta_0$.

It follows from Lemma 4.1 that functions (i)–(iii) are analytic (also in the strong X_p sense) on the regions where they are defined. As is well known, $\epsilon_k(z)$ can also be defined as the unique analytic continuation onto $\mathbb{R}$ of the function $\tilde{\epsilon}_k(z)$ given by

$$\tilde{\epsilon}_k(z)\equiv\begin{cases}\Lambda_k(z), & k\,\mathrm{Im}\,z>0\,,\\ \Lambda_k(z)+2\pi i\eta(k,z), & 0>k\,\mathrm{Im}\,z>-\theta_0\,.\end{cases}$$

Functions (i)–(iii) are simply related: To describe the relationships, we must consider the boundary values of these functions on the boundaries of their regions of analyticity.

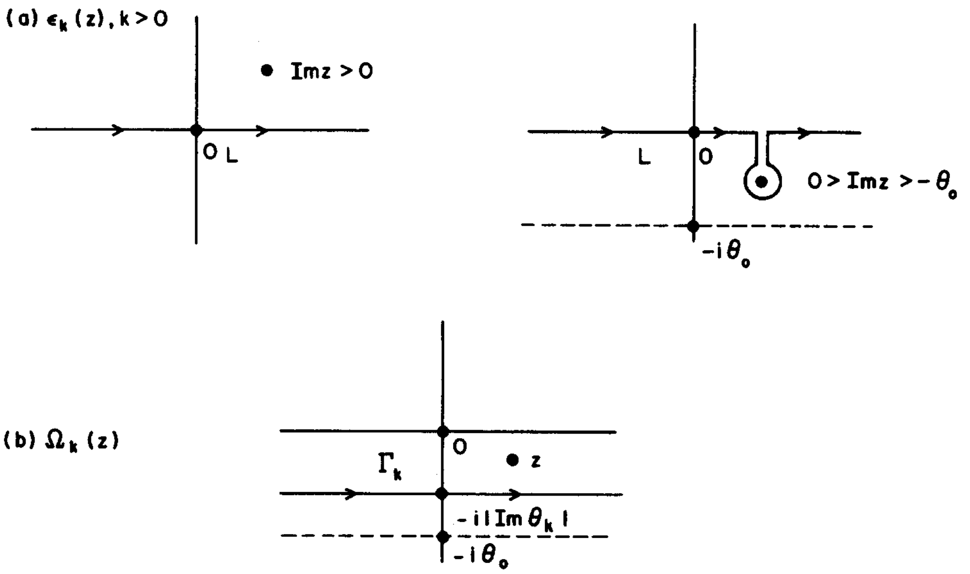


FIG. 11. Contours used in the definition of (a) $\epsilon_k(z)$ and (b) $\Omega_k(z)$. The location of the contours depends on sgn(k) and sgn(Im θ).

We define

$$\Lambda^{(\pm)} \equiv \lim_{\epsilon\to 0^+} \Lambda(x \pm i\epsilon), \quad x\in\mathbb{R},$$

$$\epsilon_k^{(+)}(s - i\theta_0) \equiv \epsilon_k(s + i(\epsilon - \theta_0)), \quad s\in\mathbb{R},$$

$$\Omega_k^{(\pm)}(r + i\,\mathrm{Im}\,\theta_k) \equiv \lim_{\epsilon\to 0^+} \Omega_k(r + i\,\mathrm{Im}\,\theta_k \pm i\epsilon), \quad r\in\mathbb{R}.$$

It follows from Assumption 2 that these boundary value functions are continuous and, in fact, Hölder continuous with the Hölder index $1-\epsilon$ for any $\epsilon>0$.[20] As above, we assume $|\mathrm{Im}\,\theta_k| < \theta_0$. Then for $\mathrm{Im}\,\theta<0$ and depending upon the location of z, we have the following relations (see Fig. 12):

$$\epsilon_k(z) = \Omega_k(z) = \Lambda_k(z), \quad p\,\mathrm{Im}\,z>0,$$

$$\epsilon_k(z) = \Omega_k(z) = \Lambda_k^{(p)}(z), \quad p\,\mathrm{Im}\,z=0,$$

$$\epsilon_k(z) = \Omega_k(z) \neq \Lambda_k(z), \quad 0>p\,\mathrm{Im}\,z>\mathrm{Im}\,\theta.$$

If $|\mathrm{Im}\,\theta_k|<\theta_0$,

$$\epsilon_k(z) \neq \Omega_k(z) = \Lambda_k(z), \quad \mathrm{Im}\,\theta>p\,\mathrm{Im}\,z>-\theta_0,$$

$$\epsilon_k(z) \neq \Omega_k(z) = \Lambda_k(z), \quad p\,\mathrm{Im}\,z=-\theta_0.$$

If $|\mathrm{Im}\,\theta_k|=\theta_0$,

$$\epsilon_k^{(p)}(z) = \Omega_k^{(p)}(z) \neq \Lambda_k(z) = \Omega_k^{(-p)}(z), \quad p\,\mathrm{Im}\,z=-\theta_0$$

and for $p\,\mathrm{Im}\,z<-\theta_0$, $\Omega_k(z)=\Lambda_k(z)$; however, this function does not equal $\epsilon_k(z)$ whenever it can be extended into this region.

Finally, we turn to the location and number of zeros of functions (i)–(iii). For the Landau dielectric function, this has been discussed by many authors.[24–26] Our first result is similar to that obtained by Saenz.[26]

Theorem A2: Let F_0 and η satisfy Assumptions 1 and 2.

(i) All the zeros of $\epsilon_k(z)$ lie in the strip $-\theta_0 \leqslant p\,\mathrm{Im}\,z \leqslant M$, where $M \equiv \|\eta_k\|_1 = \int_{-\infty}^{\infty} |\eta(k,v)|dv$.

(ii) For any b such that $pb>-\theta_0$, $\epsilon_k(z)$ has finitely many zeros in the strip $-b\leqslant p\,\mathrm{Im}\,z\leqslant M$.

Proof: Let $z=x+iy$. Statement (i) of Theorem A2 follows from the fact that $|v-z|\geqslant|y|$, so that

$$\left|\int_{-\infty}^{\infty} (v-z)^{-1}\eta(k,v)dv\right| \leqslant |y|^{-1}\|\eta_k\|_1.$$

Hence, if $|y|>\|\eta_k\|_1$, $\epsilon_k(z)$ cannot vanish. In fact, $\lim_{|y|\to\infty}|\epsilon_k(z)|=1$. Statement (ii) of Theorem A2 follows from the Paley–Wiener theorem[27] for functions analytic in strips. Condition (iii) of Assumption 2 and the analyticity part of Assumption 2 imply that $\hat{\eta}$, the Fourier transform of η, is a function of exponential type, i.e., for any σ such that $|\sigma|\leqslant\theta_0$:

$$e^{\sigma|s|}\hat{\eta}(s)\in L^2(\mathbb{R}). \tag{A1}$$

Given any δ, $0<\delta<2\theta_0$, let $\sigma=-\theta_0+\delta$. Then by the Plancherel theorem for $z=\tau+i\sigma$,

$$\int_{-\infty}^{\infty} (s-z)^{-1}\eta(k,s)ds = \int_0^{\infty} \hat{\eta}(s)e^{s(\sigma-i\tau)}\,ds. \tag{A2}$$

Since $\sigma-\theta_0<0$, it follows by the Schwarz inequality and (A1) that the integral on the rhs of (A2) converges absolutely. Hence, since $\tau\in\mathbb{R}$, $\eta(s)e^{s\sigma}\in L^1(\mathbb{R}^+)$. By the Riemann–Lebesgue theorem,

$$\lim_{|\tau|\to\infty}\int_0^{\infty} \hat{\eta}(s)e^{s(\sigma-i\tau)} = \lim_{|\mathrm{Re}\,z|\to\infty}\int_{-\infty}^{\infty} (s-z)^{-1}\eta(k,s)=0,$$

so that along any line $z=\tau+i\sigma$ with $p\sigma>0$, $\epsilon_k(z)\to 0$ as $|\mathrm{Re}\,z|\to\infty$. Furthermore, for $-\theta_0<p\sigma<0$, it follows from the Paley–Wiener theorems and part (ii) of Assumption 2 that

$$\lim_{|\mathrm{Re}\,z|\to\infty}\left[\int (s-z)^{-1}\eta(k,s) + 2\pi i\eta(k,z)\right]=0.$$

Hence, for each $\sigma=\mathrm{Im}\,z$ with $p\sigma>-\theta_0$ there exists $R(\sigma)>0$ such that $\tau=\mathrm{Re}\,z>R(\sigma)$ implies that $\epsilon_k(z)$ is bounded below away from zero. It now follows from the analyticity of $\epsilon_k(z)$ and the assumed continuity of $\eta(k,s-ip\theta_0)$ that $R(\sigma)$ is a continuous function of σ on $(p\infty,-p\theta_0]$, in particular, on $[pM,-p\theta_0]$. Consequently, for any $\delta>0$, $|\epsilon_k(z)|<1$ if and only if $|\mathrm{Re}\,z| > \max_{\sigma\in[pM,-p(\theta_0-\delta)]} R(\sigma)$ and $-\theta_0+\delta\leqslant p\,\mathrm{Im}\,z\leqslant M$. Since $\epsilon_k(z)$ is analytic on an open set containing this rectangle, it can have at most a finite number of zeros. (Note that we cannot take $\delta=0$ unless we specify further the behavior of η on the boundary $s-ip\theta_0$, $s\in\mathbb{R}$.) □

Corollary A3: The Case–van Kampen spectral function $\Lambda_k(z)$ has at most finitely many zeros and they all lie in the strip $-M\leqslant\mathrm{Im}\,z\leqslant M$, where $M=\|\eta_k\|_1$.

Proof: For $p\,\mathrm{Im}\,z>0$, $\epsilon_k(z)=\Lambda_k(z)$ and $\Lambda_k^{(p)}(x)=\epsilon_k(z)$. Hence, by Theorem A2, Λ_k has finitely many zeros in $0\leqslant p\,\mathrm{Im}\,z\leqslant M$. Since $\Lambda_k(z)^*=\Lambda_k(z^*)$, Λ_k has only finitely many zeros in $-M\leqslant\mathrm{Im}\,z\leqslant M$. □

Corollary A4: Suppose that $|\mathrm{Im}\,\theta_k|<\theta_0$. Then the hybrid dielectric function Ω_k has at most finitely many zeros and they all lie in the strip $\min(-M,\mathrm{Im}\,\theta)\leqslant p\,\mathrm{Im}\,z\leqslant M$. If $\theta_0=|\mathrm{Im}\,\theta_k|$, then for any $\delta>0$, $\Omega_k(z)$ has at most finitely many zeros in the strip $\min(-M,-\theta_0+\delta)\leqslant p\,\mathrm{Im}\,z\leqslant M$.

Proof: This follows from the relations between Λ_k, ϵ_k, and Ω_k given above and Theorem A2. □

On the basis of Assumption 2 alone, we do not see how to eliminate the possible existence of accumulation points of zeros of $\epsilon_k(z)$ on the line segments $\mathrm{Im}\,z=-p\theta_0$ and $|\mathrm{Re}\,z|<R(-p\theta_0)$. Consequently, when the complex translation is performed in Sec. V, we restrict ourselves to θ such that $|\theta|<\theta_0$. This insures the finiteness of the point spectrum of $L_k(\theta)$.

APPENDIX B: GENERAL SPECTRAL DEFORMATION WHEN F_0 VANISHES ON OPEN SETS

The choice of the velocity group as the spectral deformation group is inappropriate, for example, in the case that supp F_0 (the support of F_0) is compact (or vanishes on some open set). We briefly indicate how to obtain results similar to those discussed in Secs. III and IV in more general situations where F_0 satisfies weaker analyticity assumptions. The method is an extension of that presented in Sec. II and is due to Hunziker.[10] The group $U(\theta)$ discussed in Sec. II is constructed from a flow on $\mathbb{R}^n$ generated by some vector field. Hunziker noted that it suffices to work with the first-order approximation to this flow. It is convenient to do this in the case that supp F_0 is compact, for example.

Let us suppose (for the sake of concreteness) that supp $F_0 \subset [-c,c]$ for some $0 < c < \infty$ and that F_0 is analytic on some open rectangle $R \equiv \{z \mid |\mathrm{Re}\, z| < c, \ -\mu < \mathrm{Im}\, z < \delta, 0 < \mu, \delta < \infty\}$. The choice of a rectangle is convenient, but not necessary. The essential point is that F_0 has some sufficiently large domain of analyticity about the point where spectral deformation is to be performed. The translation group used in Sec. III is generated by the constant vector field on $\mathbb{R}$. For the case at hand, we choose a vector field $\psi \in C_0^\infty(\mathbb{R})$ with supp $\psi \subset [-c/2, c/2]$ and $\|\psi\|_\infty < 1$. We consider the infinitesimal transformation on $\mathbb{R}$:

$$v \to v + t\psi(v) \equiv s_t(v).$$

For each t, $|t| \leqslant 1$, the map $s_t : \mathbb{R} \to \mathbb{R}$ is a smooth diffeomorphism. Let $J_t(v) \equiv 1 + t\psi'(v) > 0$ be the Jacobian. We define a transformation $U(t)$ on X_p by

$$(U(t)g) = J_t(v)^{1/p} g(s_t(v)).$$

Then $U(t)$, $|t| < 1$ is a bounded, invertible map on X_p with the bound

$$\|U(t)\|_{\Gamma,p} \leqslant 1 + t d_\psi$$

for some constant $0 < d_\psi < \infty$ depending on ψ. [Here, $U(t)$, like $U(\theta)$, is not an isometry on X_p; see (3.11).]

Even though the family of operators $\{U(t) \mid |t| < 1\}$ does not form a group, Hunziker (see, also, Cycon[28]) has shown that it is possible to recover the major aspects of the theory using such a family. In particular, there exists a dense set of analytic vectors for this family and the poles of the meromorphic continuation of matrix elements of the resolvent (between vectors from this set) are in one-to-one correspondence with the eigenvalues of an analytic family of operators. We indicate this for the linear Vlasov operator.

We first compute the conjugated linearized Vlasov operator and obtain, for $t \in \mathbb{R}$, $|t| < 1$, ad $g \in D(M_v)$,

$$\begin{aligned}(L_k(t)g)(v) &\equiv (U(t) L_k U(t)^{-1} g)(v) \\ &= -ik\Big[s_t(v)g(v) + \eta(k, s_t(v)) \\ &\qquad \times \int_{-\infty}^{\infty} g(v')\,dv'\Big]\end{aligned}$$

for $k \neq 0$. Note that this expression reduces to (3.14) upon taking $\psi(v) = 1$ and $t = \theta_k$. Clearly, the differomorphism s_t extends in t to an analytic map of $\{z \mid |z| \leqslant 1\} \subset \mathbb{C} \to \mathbb{C}$. For convenience, we assume that $\|\psi\|_\infty < 1$. Then for F_0 analytic on the rectangle R, the map

$$t \to \eta(k, s_t(v))$$

is analytic in t on the disk $|t| < \min(\mu, \delta, 1) \equiv \delta_0$. As in Sec. III, one easily shows that $L_k(\theta)$ is an analytic type-A family of operators on the disk $|t| < \delta_0$.

To see the effect of this scaling on the spectrum of L_k we compute the resolvent as in Sec. IV: The result is

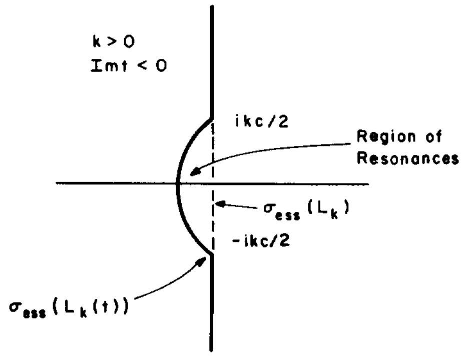


FIG. 13. The local spectral deformation of $\sigma_{\mathrm{ess}}(L_k)$ corresponding to the vector field ψ, with compact support for $k > 0$ and $\mathrm{Im}\, t < 0$.

$$\begin{aligned}(R_k(t,z)g)(v) &= (z + iks_t(v))^{-1}\Big[g(v) - \frac{\eta(k, s_t(v))}{\Omega_k(t, iz/k)} \\ &\qquad \times \int_{-\infty}^{\infty} \frac{g(v')}{z + iks_t(v')}\,dv'\Big],\end{aligned}$$

where

$$\Omega_k(t,z) \equiv 1 + \int_{-\infty}^{\infty} (s_t(v') - z)^{-1} \eta(k, s_t(v'))\,dv'. \tag{B1}$$

Now it is clear from Sec. IV and Appendix A that

$$\sigma_{\mathrm{ess}}(L_k(\theta)) = \{z \mid z + iks_t(v) = 0\}. \tag{B2}$$

Equation (B2) is precisely the locally distorted contour

$$\{z \mid z = k(t_2\psi(v) - iv),\ v \in \mathbb{R},\ t_2 \equiv \mathrm{Im}\, t\};$$

see Fig. 13. Note that for $|\mathrm{Im}\, z| > c|k|$ the $\sigma_{\mathrm{ess}}(L_k(\theta))$ coincides with $\sigma_{\mathrm{ess}}(L_k)$. Hence, we have obtained a *local* spectral deformation. An examination of the hybrid dielectric function $\Omega_k(t,z)$, (B1) in this case indicates that resonances lie between the imaginary axis [i.e., $\sigma_{\mathrm{ess}}(L_k)$] and the deformed spectrum (B2). It is obvious that by varying ψ we can obtain other local deformations provided that we respect the domain of analyticity of F_0.

Note added in proof: The method of spectral deformation has recently been used by H. Ye and A. Kaufman[29,30] to obtain an analytic solution for the second harmonic mode conversion problem in the ion-cyclotron-frequency heating of tokamaks.

[1] J. D. Crawford and P. D. Hislop, Ann. Phys. **189**, 265 (1989).
[2] J. Aguilar and J. M. Combes, Commun. Math Phys. **22**, 269 (1971).
[3] B. Simon, Phys. Lett. **71A**, 211 (1979).
[4] M. D. Arthur, W. Greenberg, and P. F. Zwiefel, Phys. Fluids **20**, 1296 (1977).
[5] M. D. Arthur, W. Greenberg, and P. F. Zwiefel, Phys. Fluids **22**, 1465 (1979).
[6] L. Landau, J. Phys. USSR **10**, 25 (1946).
[7] K. M. Case, Ann. Phys. 7, 349 (1959).
[8] N. G. van Kampen, Physica **23**, 647 (1957).
[9] I. M. Sigal, Ann. Inst. Henri Poincaré **41**, 103 (1984); **41**, 333 (1984).
[10] W. Hunziker, Ann. Inst. Henri Poincaré **45**, 339 (1986).
[11] P. D. Hislop and I. M. Sigal, in *Proceedings of the International Conference on Differential Equations and Mathematical Physics*, edited by I. Knowles and Y. Saito, Springer Lecture Notes in Math. (Springer–Verlag, Berlin, 1987); Memoirs of the AMS, **399** (1989).

[12]M. Reed and B. Simon, *Methods of Modern Mathematical Physics, IV* (Academic, New York, 1978).

[13]P. Briet, J. M. Combes, and P. Duclos, J. Math. Anal. Appl. **126**, 90 (1987).

[14]I. M. Sigal, *Exponential Bounds on Resonance States and Width of Resonances*, Adv. Appl. Math. **9**, 127 (1988).

[15]J. Howland, Pacific J. Math. **55**, 157 (1974).

[16]T. Kato, *Perturbation Theory for Linear Operators* (Springer, New York, 1976), 2nd ed.

[17]H. Weitzner, Phys. Fluids **6**, 1123 (1963).

[18]I. M. Sigal, J. Operator Theor. **13**, 119 (1985).

[19]M. Reed and B. Simon, *Methods of Modern Mathematical Physics, II* (Academic, New York, 1976).

[20]N. I. Muskhelishvili, *Singular Integral Equations* (Noordhoff, Leiden, 1977).

[21]M. Trocheris, in *Magneto-fluid and Plasma Dynamics, Proceedings of the Symposia in Applied Math.* (AMS, Providence, RI, 1967), Vol. 18.

[22]M. Trocheris, J. Math. **21**, 932 (1980); J. Math. **21**, 941 (1980).

[23]P. Degond, C. R. Acad. Sci. Paris **296**, 969 (1983); Trans. AMS **29**, 435 (1986).

[24]J. N. Haynes, Nuovo Cimento **30**, 1048 (1963).

[25]G. Backus, J. Math. Phys. **1**, 178 (1960).

[26]A. W. Sáenz, J. Math. Phys. **6**, 859 (1965).

[27]R. E. A. C. Paley and N. Wiener, *Fourier Transforms in the Complex Domain* (AMS, Providence, RI, 1934), Vol. XIX.

[28]H. Cycon, Helv. Phys. Acta **58**, 969 (1985).

[29]H. Ye and A. Kaufman, Phys. Rev. Lett. **61**, 2762 (1988).

[30]H. Ye, doctoral thesis, University of California, Berkeley, 1989.

Solution of the Dirac equation in Kasner's space-time

Sushil K. Srivastava
Department of Mathematics, North-Eastern Hill University, Bijni Complex, Bhagyakul, Shillong-793003, India

(Received 5 February 1988; accepted for publication 14 June 1989)

The Dirac equation for the spin-$\frac{1}{2}$ field in Kasner's space-time is discussed and various possible solutions are obtained. For further interpretation of the theory a current vector is derived and Gordon decomposition of the current for massive fields is discussed.

I. INTRODUCTION

Under the scheme of uniting quantum mechanics and general relativity, the search for solutions of relativistically covariant equations for spin-$\frac{1}{2}$ particles is very interesting. In 1985 Shishkin and Andrushkeirsh[1] solved the Dirac eqution in degenerate Kasner space-time. Recently, Barut and Duru[2] have gotten exact solutions of the Dirac equation in spatially flat Robertson–Walker space-times. Audrestsch and Schafer[3] have also discussed the dynamics of spin-$\frac{1}{2}$ particles in spatially homogeneous cosmological models.

In the present paper, the plan is to investigate solutions of the Dirac equation in Kasner space-time, which is a model for an anisotropically expanding universe with the line element

$$ds^2 = dt^2 - t^{2a_1} dx^2 - t^{2a_2} dy^2 - t^{2a_3} dz^2, \tag{1.1}$$

where the speed of light $c = 1$. This model is the vacuum solution of Einstein's equations if the real numbers a_1, a_2, a_3 satisfy the constraints

$$a_1 + a_2 + a_3 = a_1^2 + a_2^2 + a_3^2 = 1. \tag{1.2}$$

The focus here lies on getting the solution of the Dirac equation in the degenerate as well as nondegenerate cases. Here, a_1, a_2, and a_3 in (1.1) and (1.2), are taken to be $(1+\alpha)/(1+\alpha+\alpha^2)$, $\alpha(1+\alpha)/(1+\alpha+\alpha^2)$, and $-\alpha/(1+\alpha+\alpha^2)$, respectively (where α is real). It is interesting to see that when $\alpha \to \infty$, one gets the degenerate case, $a_1 = 0$, $a_2 = 1$, $a_3 = 0$.

Section II contains the definition of the tetrad components and the Dirac eqution in Kasner's model. In Sec. III, various solutions of Dirac equtions for chiral and nonchiral spinors are obtained. Section IV is devoted to a discussion on the current vector. Gordon decomposition of the current for massive spin-$\frac{1}{2}$ fields is discussed.

Hereafter $\hbar = c = 1$ is used as the fundamental unit.

II. DIRAC EQUATION IN KASNER SPACE-TIME

The tetrad components h^{μ}_{a} are defined as[4]

$$h^{\mu}_{a} h^{\nu}_{b} g_{\mu\nu} = \eta_{ab}, \tag{2.1}$$

where η_{ab} is the Minkowski metric and $g_{\mu\nu}$ is the metric given by the line element (1.1). Here μ,ν are curved space indices and a,b are flat space indices. So,

$$h^{0}_{0} = 1, \quad h^{1}_{1} = t^{-a_1}, \quad h^{2}_{2} = t^{-a_2}, \quad \text{and} \quad h^{3}_{3} = t^{-a_3}. \tag{2.2}$$

The Dirac equation, in curved space-time, for spinor field ψ is given as

$$i\gamma^{\mu}\psi_{;\mu} + m\psi = 0, \tag{2.3}$$

where γ^{μ}, the Dirac matrices in curved space, are

$$\gamma^{\mu} = h^{\mu}_{a}\tilde{\gamma}^{a} \tag{2.4}$$

and satisfy the anticommutation rule

$$\{\gamma^{\mu},\gamma^{\nu}\} = 2g^{\mu\nu}. \tag{2.5}$$

In (2.4) $\tilde{\gamma}^{a}$ are the standard Dirac matrices for flat space-time and satisfy the anticommutation rule[5]

$$\{\tilde{\gamma}^{a},\tilde{\gamma}^{b}\} = 2\eta^{ab}. \tag{2.6}$$

The covariant derivatives for ψ, in curved space-time used in (2.3), are given as

$$\psi_{;\mu} = (\partial_{\mu} - \Gamma_{\mu})\psi, \tag{2.7}$$

where

$$\Gamma_{\mu} = -\tfrac{1}{4}(\partial_{\mu} h^{\rho}_{a} + \{^{\rho}_{\sigma\mu}\} h^{\sigma}_{a}) g_{\nu\rho} h^{\nu}_{b} \tilde{\gamma}^{b}\tilde{\gamma}^{a} \tag{2.8}$$

and $\{^{\rho}_{\sigma\mu}\}$ is the affine connection.

Using the metric from (1.1) in (2.3), the Dirac equation for Kasner's space-time looks like

$$[\tilde{\gamma}^{0}\{\partial_0 + (1/2t)(a_1 t^{a_1} + a_2 t^{a_2} + a_3 t^{a_3})\} + t^{-a_1}\tilde{\gamma}^{1}\partial_1 + t^{-a_2}\tilde{\gamma}^{2}\partial_2 + t^{-a_3}\tilde{\gamma}^{3}\partial_3 - im]\psi = 0. \tag{2.9}$$

III. SOLUTION OF THE DIRAC EQUATION

Substituting $\psi = \Psi \exp[-\frac{1}{2}(t^{a_1} + t^{a_2} + t^{a_3})]$ in (2.9), one gets an equation for Ψ as

$$[\tilde{\gamma}^{0}\partial_0 + t^{-a_1}\tilde{\gamma}^{1}\partial_1 + t^{-a_2}\tilde{\gamma}^{2}\partial_2 + t^{-a_3}\tilde{\gamma}^{3}\partial_3 - im]\Psi = 0. \tag{3.1}$$

Setting

$$\overline{\Psi} = (2\pi)^{-3/2}\exp(ik_1 x + ik_2 y + ik_3 z) \times \begin{bmatrix} f_{\rm I}(k_1,k_2,k_3,t) \\ f_{\rm II}(k_1,k_2,k_3,t) \end{bmatrix} \tag{3.2}$$

in (3.1) the two-component spinors obey the following coupled equations:

$$(\partial_0 - im)f_{\rm I} + i(t^{-a_1}\sigma^{1}k_1 + t^{-a_2}\sigma^{2}k_2 + t^{-a_3}\sigma^{3}k_3)f_{\rm II} = 0 \tag{3.3a}$$

and

 0022-2488/89/122838-07$02.50

$$(\partial_0 + im)f_{\rm II} + i(t^{-a_1}\sigma^1 k_1 + t^{-a_2}\sigma^2 k_2 + t^{-a_3}\sigma^3 k_3)f_{\rm I} = 0, \tag{3.3b}$$

where σ^1, σ^2, and σ^3 are Pauli matrices.

A. Nondegenerate case (case I)

In this case, $a_1 \neq 0$, $a_2 \neq 0$, $a_3 \neq 0$. From (3.3a) and (3.3b), one gets the second-order differential equtien for $f_{\rm I}$

$$\left(\partial_0^2 + \frac{\dot q}{q}\partial_0 + m^2 - im\frac{\dot q}{q} - \frac{1}{q^2}\right)f_{\rm I} = 0 \tag{3.4a}$$

and $f_{\rm II}$ is expressed as

$$f_{\rm II} = -q(\partial_0 - im)f_{\rm I}. \tag{3.4b}$$

In (3.4a) and (3.4b), q is given by

$$q = -i\frac{(t^{-a_1}\sigma^1 k_1 + t^{-a_2}\sigma^2 k_2 + t^{-a_3}\sigma^3 k_3)}{t^{-2a_1}k_1^2 + t^{-2a_2}k_2^2 + t^{-2a_3}k_3^2}. \tag{3.5}$$

Introducing

$$\tau = \int^t \frac{dt'}{q(t')},$$

(3.4) is rewritten as

$$\frac{d^2 f_{\rm I}}{d\tau^2} + [m^2 q^2(t) - imq\dot q(t) - 1]f_{\rm I} = 0, \tag{3.6a}$$

$$f_{\rm II} = -q(t)\left[\frac{d}{d\tau} - im\right]f_{\rm I}. \tag{3.6b}$$

Now, (3.6a) yields the WKB solution

$$f_{\rm I} = \frac{1}{[m^2q^2 - imq\dot q - 1]^{1/4}}\left\{N_1 \exp\left(i\int d\tau\sqrt{m^2q^2 - imq\dot q - 1}\right) + N_2\exp\left(-i\int d\tau\sqrt{m^2q^2 - imq\dot q - 1}\right)\right\} \tag{3.7a}$$

and

$$\begin{aligned} f_{\rm II} = -\frac{q(t)}{(m^2q^2 - imq\dot q - 1)^{5/4}}\Big[&N_3\{-q/4(2mq\dot q - imq\ddot q - im\dot q^2) + i(m^2q^2 - imq\dot q - 1)^{3/2} \\ &- im(m^2q^2 - imq\dot q - 1)^{5/4}\}\exp\left(i\int d\tau\sqrt{m^2q^2 - imq\dot q - 1}\right) + N_4\{-(q/4)(2mq\dot q - imq\ddot q - im\dot q^2) \\ &- i(m^2q^2 - imq\dot q - 1)^{3/2} - im(m^2q^2 - imq\dot q - 1)^{5/4}\}\exp\left(-i\int d\tau\sqrt{m^2q^2 - imq\dot q - 1)}\right)\Big], \end{aligned} \tag{3.7b}$$

where N_1, N_2, N_3, N_4 are normalization constants. Corresponding to $f_{\rm I}$ and $f_{\rm II}$, ψ can be written as

$$\begin{aligned} \Psi_{{\rm I},\mathscr{S}} = &\frac{\exp\{ik_1x + ik_2y + ik_3z - \frac{1}{2}(t^{a_1} + t^{a_2} + t^{a_3})\}}{(2\pi)^{3/2}[m^2q^2 - imq\dot q - 1]^{1/4}} \\ &\times\left\{N_1 u_s \exp\left(i\int\frac{dt}{q}\sqrt{m^2q^2 - imq\dot q - 1}\right) + N_2\hat u_s\exp\left(-i\int\frac{dt}{q}\sqrt{m^2q^2 - imq\dot q - 1}\right)\right\}, \end{aligned} \tag{3.8a}$$

$$\begin{aligned} \Psi_{{\rm II},\mathscr{S}} = &\frac{-q\exp\{ik_1x + ik_2y + ik_3z - \frac{1}{2}(t^{a_1} + t^{a_2} + t^{a_3})\}}{(2\pi)^{3/2}(m^2q^2 - imq\dot q - 1)^{5/4}} \\ &\times\Big[N_3u_s\{-(q/4)(2mq\dot q - imq\ddot q - im\dot q^2) + i(m^2q^2 - imq\dot q - 1)^{3/2} - im(m^2q^2 - imq\dot q - 1)^{5/4}\} \\ &\times\exp\left(i\int\frac{dt}{q}\sqrt{m^2q^2 - imq\dot q - 1}\right) + N_4\,\hat u_s\{-(q/4)(2mq\dot q - imq\ddot q - im\dot q^2) - i(m^2q^2 - imq\dot q - 1)^{3/2} \\ &- im(m^2q^2 - imq\dot q - 1)^{5/4}\}\exp\left(-i\int\frac{dt}{q}\sqrt{m^2q^2 - imq\dot q - 1}\right)\Big] \end{aligned} \tag{3.8b}$$

with spin quantum number $s = \pm 1$ and

$$\hat u_1 = \begin{pmatrix}1\\0\\0\\0\end{pmatrix},\quad \hat u_{-1} = \begin{pmatrix}0\\1\\0\\0\end{pmatrix},$$

$$u_1 = \begin{pmatrix}0\\0\\-k_3\\-k_1 - ik_2\end{pmatrix} \quad\text{and}\quad u_{-1} = \begin{pmatrix}0\\0\\-k_1 + ik_2\\k_3\end{pmatrix}. \tag{3.9}$$

The normalization constants N_i are determined in such a way that asymptotically, i.e., in the flat space-time limit, ψ satisfies the usual $\delta_{ss'}\delta(k_1 - k_1')\delta(k_2 - k_2')\delta(k_3 - k_3')$ normalization. The norm of ψ is defined as

$$(\psi^k,\psi^{k'}) = \int_t t\, d^3x\, \bar{\psi}^k \gamma^0 \psi^{k'}, \tag{3.10}$$

where

$\bar{\psi} = \psi^\dagger \gamma^0.$

Using the above process for normalization of the wavefunction ψ,

$$N_1 = k^{-1}\exp(1.5)\left[\left(1+\frac{m^2}{k^2}\right)^2 + m^2\frac{(a_1k_1^2+a_2k_2^2+a_3k_3^2)^2}{k^8}\right]^{1/2}, \tag{3.11a}$$

$$N_2 = \left[\left(1+\frac{m^2}{k^2}\right)^2 + \frac{m^2(a_1k_1^2+a_2k_2^2+a_3k_3^2)^2}{k^8}\right]^{1/2}, \tag{3.11b}$$

$$N_3 = \frac{r^{5/4}\exp(1.5)}{\sqrt{(v+mr^{3/2}\sin(3\theta/2)-mr^{5/4}\sin(5\theta/4))^2+(\bar{v}r^{3/2}\cos(3\theta/2)-mr^{5/4}\cos(5\theta/4))^2}}, \tag{3.11c}$$

$$N_4 = \frac{r^{5/4}k\exp(1.5)}{\sqrt{(v+mr^{3/4}\sin(3\theta/2)-mr^{5/4}\sin(5\theta/4))^2+(\bar{v}+r^{3/2}\cos(3\theta/2)-mr^{5/4}\cos(5\theta/4))^2}}, \tag{3.11d}$$

where

$$k=\sqrt{k_1^2+k_2^2+k_3^2},$$

$$r^2=\left(1+\frac{m^2}{k^2}\right)^2+\frac{m^2(a_1k_1^2+a_2k_2^2+a_3k_3^2)^2}{k^8},$$

$$\tan\theta = -\frac{m(a_1k_1^2+a_2k_2^2+a_3k_3^2)}{k^4(1+m^2/k^2)},$$

$$v=\frac{m}{4k}[2k^4(a_1^2k_1^2+a_2^2k_2^2+a_3^2k_3^2)+(a_1k_1^2+a_2k_2^2+a_3k_3^2)\{k^4-2(a_1k_1^2+a_2k_2^2+a_3k_3^2)\}],$$

and

$$\bar{v} = -\frac{m(a_1k_1^2+a_2k_2^2+a_3k_3^2)}{2k^5}.$$

If ψ has a definite chirality (left-handed or right-handed), $m=0$. So, in this case, the Dirac equation for ψ can be solved exactly. As a result

$$f_{\rm I} = \widetilde{N}_1\exp\left[it\left(\frac{t^{-a_1}\sigma^1k_1}{1-a_1}+\frac{t^{-a_2}\sigma^2k_2}{1-a_2}+\frac{t^{-a_3}\sigma^3k_3}{1-a_3}\right)\right]+\widetilde{N}_2\exp\left[-it\left(\frac{t^{-a_1}\sigma^1k_1}{1-a_1}+\frac{t^{-a_2}\sigma^2k^2}{1-a_2}+\frac{t^{-a_3}\sigma^3k_3}{1-a_3}\right)\right] \tag{3.12a}$$

and

$$f_{\rm II} = i\frac{(t^{-a_1}\sigma^1k_1+t^{-a_2}\sigma^2k_2+t^{-a_3}\sigma^3k_3}{t^{-2a_1}k_1^2+t^{-2a_2}k_2^2+t^{-2a_3}k_3^2)}\times\left[\widetilde{N}_3\exp\left\{it\left(\frac{t^{-a_1}\sigma^1k_1}{1-a_1}+\frac{t^{-a_2}\sigma^2k_2}{1-a_2}+\frac{t^{-a_3}\sigma_3k_3}{1-a_3}\right)\right\}\right.$$
$$\left.+\widetilde{N}_4\exp\left\{-it\left(\frac{t^{-a_1}\sigma^1k_1}{1-a_1}+\frac{t^{-a_2}\sigma^2k_2}{1-a_2}+\frac{t^{-a_3}\sigma^3k_3}{1-a_3}\right)\right\}\right]. \tag{3.12b}$$

Now $\psi_{{\rm I},s}$ and $\psi_{{\rm II},s}$ can be written

$$\psi_{{\rm I},s} = (2\pi)^{-3/2}\exp\{ik_1x+ik_2y+ik_3z-\tfrac{1}{2}(t^{a_1}+t^{a_2}+t^{a_3}\}$$
$$\times\left[\widetilde{N}_1u_s\exp\left\{it\left(\frac{t^{-a_1}\sigma^1k_1}{1-a_1}+\frac{t^{-a_2}\sigma^2k_2}{1-a_2}+\frac{t^{-a_3}\sigma^3k_3}{1-a_3}\right)\right\}\right.$$
$$\left.+\widetilde{N}_2\hat{u}_s\exp\left\{-it\left(\frac{t^{-a_1}\sigma^1k_1}{1-a_1}+\frac{t^{-a_2}\sigma^2k_2}{1-a_2}+\frac{t^{-a_3}\sigma^3k_3}{1-a_3}\right)\right\}\right], \tag{3.13a}$$

$$\psi_{{\rm II},s} = i(2\pi)^{-3/2}\exp\{ik_1x+ik_2y+ik_3z-\tfrac{1}{2}(t^{a_1}+t^{a_2}+t^{a_3})\}$$
$$\times\left[\widetilde{N}_3u_s\exp\left\{it\left(\frac{t^{-a_1}\sigma^1k_1}{1-a_1}+\frac{t^{-a_2}\sigma^2k_2}{1-a_2}+\frac{t^{-a_3}\sigma^3k_3}{1-a_3}\right)\right\}\right.$$
$$\left.+\widetilde{N}_4\hat{u}_s\exp\left\{-it\left(\frac{t^{-a_1}\sigma^1k_1}{1-a_1}+\frac{t^{-a_2}\sigma^2k_2}{1-a_2}+\frac{t^{-a_3}\sigma^3k_3}{1-a_3}\right)\right\}\right], \tag{3.13b}$$

where u_s and $\hat{u}_s$ has the same meaning as given earlier in (3.9).

Applying the scheme of normalization given above, $\widetilde{N}_i$ are calculated as

$$\widetilde{N}_1 = k^{-1}\exp(1.5), \quad \widetilde{N}_2 = \exp(1.5), \quad \widetilde{N}_3 = \exp(1.5), \quad \widetilde{N}_4 = k\exp(1.5). \tag{3.14}$$

B. Degenerate case (case II)

In the degenerate case, one of a_1, a_2, a_3 is equal to 1 and other are vanishing. So, it is taken as $a_1 = 0$, $a_2 = 1$, and $a_3 = 0$. As a result, (3.3a) and (3.3b) are written as

$$(\partial_0 - im)f_{\rm I} + i(\sigma^1 k_1 + t^{-1}\sigma^2 k_2 + \sigma^3 k_3)f_{\rm II} = 0, \tag{3.15a}$$

$$(\partial_0 + im)f_{\rm II} + i(\sigma^1 k_1 + t^{-1}\sigma^2 k_2 + \sigma^3 k_3)f_{\rm I} = 0. \tag{3.15b}$$

Equations (3.15) imply that

$$\Big[\partial_0^2 + \frac{t^{-3}k_2^2}{(k_1^2 + k_2^2 t^{-2} + k_3^2)}\partial_0 + m^2 - \frac{imt^{-3}k_2^2}{(k_1^2 + k_2^2 t^{-2} + k_3^2)} + (k_1^2 + k_2^2 t^{-2} + k_3^2)\Big]f_{\rm I} = 0. \tag{3.16a}$$

and

$$f_{\rm II} = \frac{i(\sigma^1 k_1 + t^{-1}\sigma^2 k_2 + \sigma^3 k_3)}{k_1^2 + t^{-2}k_2^2 + k_3^2}(\partial_0 - im)f_{\rm I}. \tag{3.16b}$$

Case II (*a*): In the early universe, when t is very small (3.16) may be approximated as

$$[t^2\partial_0^2 + t\partial_0 + \{k_2^2 - imt + (m^2 + k_1^2 + k_3^2)t^2\}]f_{\rm I} = 0 \tag{3.17a}$$

and

$$f_{\rm II} = (it\sigma^2 k_2/k_2^2)(\partial_0 - im)f_{\rm I}. \tag{3.17b}$$

Equation (3.17a) yields the solution

$$f_{\rm I} = \exp(l't)t^{l-1/2}\Big[C_1\,{}_1F_1\Big(\frac{2ll' - im}{2l'}, 2l, -2l't\Big) + C_2(-2l't)^{1-2l}{}_1F_1\Big(\frac{2l' - im}{2l'}, 2-2l, -2l't\Big)\Big], \tag{3.18}$$

where ${}_1F_1(a,c,x')$ is the confluent hypergeometric function,[6] $l = \frac{1}{2}[1 \pm \sqrt{1 - 4k_2^2}]$, and $l' = \pm i\sqrt{m^2 + k_1^2 + k_3^2}$. Now, using the identity

$$\frac{d}{dx'}{}_1F_1(a,c,x') = {}_1F_1(a+1, c+1, x') \tag{3.19}$$

one can find out $f_{\rm II}$ from (3.17b) and (3.18). Now, $\psi_{\rm I}$ and $\psi_{\rm II}$ are written as

$$\psi_{{\rm I},s} = (2\pi)^{-3/2}\exp\{ik_1x + ik_2y + ik_3z - 1 - \tfrac{1}{2}t^1 + l't\}t^{l-1/2} \times\Big[C_1u_s\,{}_1F_1\Big(\frac{2ll' - im}{2l'}, 2l, -2l't\Big) + C_2\hat{u}_s\,{}_1F_1\Big(\frac{2l' - im}{2l'}, 2-2l, -2l't\Big)(-2l't)^{1-2l}\Big] \tag{3.20a}$$

and

$$\begin{aligned}\psi_{{\rm II},\mathscr{S}} &= i(2\pi)^{-3/2}k_2^{-2}\exp\{ik_1x + ik_2y + ik_3z - 1 - \tfrac{1}{2}t + l't\}t^{l+1/2}\sigma^2 k_2\\ &\times\Big[c_3u_s\Big\{\Big(l' + \Big(l - \frac{1}{2}\Big)t^{-1} - im\Big){}_1F_1\Big(\frac{2ll' - im}{2l'}, 2l, -2l't\Big) - 2l'\,{}_1F_1\Big(\frac{2ll' - im + 2l'}{2l'}, 2l+1, -2l't\Big)\Big\}\\ &+ C_4u_s\Big\{(-2l')^{1-2l}\Big(\frac{1}{2} + l' + 2\Big(l - \frac{1}{2}\Big)l't^{-2l} - im\Big){}_1F_1\Big(\frac{2l' - im}{2l'}, 2-2l, -2l't\Big)\\ &+ (-2l't)^{1-2l}(-2l')\,{}_1F_1\Big(\frac{4l' - im}{2l'}, 3-2l, -2l't\Big)\Big\}\Big],\end{aligned} \tag{3.20b}$$

where l and l' are defined as in (3.18); u_s and $\hat{u}_s$, for $\mathscr{S} = \pm 1$ are the matrices given in (3.9); and C_1, C_2, C_3, and C_4 are normalization constants, which are determined using the normalization scheme for ψ given above. So,

$$C_1 = k^{-1}\exp\Big(\frac{3}{2} - l'\Big)\;\Big|{}_1F_1\Big(\frac{2ll' - im}{2l'}, 2l, -2l'\Big)\Big|,$$

$$C_2 = (-2l')^{2l-1}\exp\Big(\frac{3}{2} - l'\Big)\;\Big|{}_1F_1\Big(\frac{2l' - im}{2l'}, 2-2l, -2l'\Big)\Big|,$$

$$C_3 = k^{-1}k_2\exp\Big(\frac{3}{2} - l'\Big)\;\Big|\Big(l' + l - \frac{1}{2} - im\Big)\Big|\;{}_1F_1\Big(\frac{2ll' - im}{2l'}, 2l, -2l'\Big) - 2l'\,{}_1F_1\Big|\frac{2ll' - im + 2l'}{2l'}, 1+2l, -2l'\Big|,$$

$$C_4 = \frac{(-2l')^{2l-1}\exp(\frac{3}{2} - l')}{|(l' + l'(l - \frac{1}{2}) - im)\,{}_1F_1((2l' - im/2l'), 2-2l, -2l') - 2l'\,{}_1F_1((4l' - im/2l'), 3-2l, -2l')|}.$$

Case II (b): If t is large, (3.16a) can be approximated as

$$[t^2\partial_0^2 + \{k_2^2 + (m^2 + k_1^2 + k_3^2)t^2\}]f_{\rm I} = 0 \tag{3.21a}$$

and

$$f_{\rm II} = it(\sigma^2 k_2/k_2^2)(\partial_0 - im)f_{\rm I}. \tag{3.21b}$$

Equation (3.21a) yields the solution

$$f_{\rm I} = t^l \exp(l't)[\widetilde{C}_1\, {}_1F_1(l,-2l,-2l't) + \widetilde{C}_2(-2l't)^{1-2l}\, {}_1F_1(1+l,2-2l,-2l't)], \tag{3.22a}$$

where $l = \frac{1}{2}(1 \pm \sqrt{1-4k_2^2})$ and $l' = i\sqrt{m^2 + k_1^2 + k_3^2}$. From (3.21b) and (3.22a),

$$\begin{aligned} f_{\rm II} = \frac{it\sigma^2 k_2}{k_2^2}[\widetilde{C}_3 t^l \exp(l't)\{(lt^{-1} + l')\,{}_1F_1(l,-2l,-2l't) \\ + (-2l')\,{}_1F_1(1+l,1-2l,-2l't) - im\,{}_1F_1(l,-2l,-2l't)\} \\ + \widetilde{C}_4(-2l')^{1-2l}t^{-l}\exp(l't)\{(lt'-l+1)\,{}_1F_1(1+l,2-2l,-2l't) \\ + t(-2l')\,{}_1F_1(2+l,3-2l,-2l't) - im\,{}_1F_1(1+l,2-2l,-2l't)\}]. \end{aligned} \tag{3.22b}$$

So, $\psi_{\rm I}$ and $\psi_{\rm II}$ are written as

$$\begin{aligned} \psi_{{\rm I},s} = (2\pi)^{-3/2}\exp(ik_1x + ik_2y + ik_3z - 1 - \tfrac{1}{2}t + l't)t^l \\ \times[\widetilde{C}_1 u_s\,{}_1F_1(l,-2l,-2l't) + \widetilde{C}_2\hat{u}_s(-2l't)^{1-2l}\,{}_1F_1(1+l,2-2l,-2l't)], \end{aligned} \tag{3.23a}$$

$$\begin{aligned} \psi_{{\rm II},s} = i(2\pi)^{-3/2}k_2^{-2}t^{l+1}\sigma^2 k_2\exp(ik_1x + ik_2y + ik_3z - 1 - \tfrac{1}{2}t + l't) \\ \times[\widetilde{C}_3 u_s\{(lt' + l')\,{}_1F_1(l,-2l,-2l't) - 2l'\,{}_1F_1(1+l,1-2l,-2l't) - im\,{}_1F_1(l,-2l,-2l't)\} \\ + \widetilde{C}_4\hat{u}_s(-2l')^{1-2l}t^{-2l}\{(l't - l + 1)\,{}_1F_1(1+l,2-2l,-2l't) \\ - 2l't\,{}_1F_1(2+l,3-2l,-2l't) - im\,{}_1F_1(1+l,2-2l,-2l't)\}]. \end{aligned} \tag{3.23b}$$

Here u_s and $\hat{u}_s$ have the same meaning as earlier. $\widetilde{C}_1$, $\widetilde{C}_2$, $\widetilde{C}_3$, and $\widetilde{C}_4$, being normalization constants, are determined according to normalization scheme mentioned above. So,

$$\begin{aligned} \widetilde{C}_1 &= k^{-1}\exp(\tfrac{3}{2} - l')/|{}_1F_1(l,-2l,-2l')|, \\ \widetilde{C}_2 &= (-2l')^{l-1}\exp(\tfrac{3}{2} - l')/|{}_1F_1(1+l,2-2l,-2l')|, \\ \widetilde{C}_3 &= k^{-1}k_2\exp(\tfrac{3}{2} - l')/|(l + l' - im)\,{}_1F_1(l,-2l,-2l') - 2l'\,{}_1F_1(1+l,1-2l,-2l')|, \\ \widetilde{C}_4 &= \frac{(-2l')^{-1+2l}k_2\exp(\tfrac{3}{2} - l')}{|(l' - l + 1 - im)\,{}_1F_1(1+l,2-2l,-2l') - 2l'\,{}_1F_1(2+l,3-2l,-2l')|}. \end{aligned} \tag{3.24}$$

Case II(c): If ψ obeys Weyl's symmetry, i.e., it has definite chirality,

$$\begin{aligned} f_{\rm I} = C_1'\exp(i[(\sigma^1 k_1 + \sigma^3 k_3)t + \sigma^2 k_2 \log t]) \\ + C_2'\exp(-i[(\sigma^1 k_1 + \sigma^3 k_3)t + \sigma^2 k_2\log t]) \end{aligned} \tag{3.25a}$$

and

$$\begin{aligned} f_{\rm II} = \frac{i(\sigma^1 k_1 + t^{-1}\sigma^2 k_2 + \sigma^3 k_3)}{(k_1^2 + t^2 k_2^2 + k_3^2)}[C_3'\exp(i[(\sigma^1 k_1 + \sigma^3 k_3)t + \sigma^2 k_2\log t]) \\ + C_4'\exp(-i[(\sigma^1 k_1 + \sigma^3 k_3)t + \sigma^2 k_2 \log t])]. \end{aligned} \tag{3.25b}$$

Now $\psi_{\rm I}$ and $\psi_{\rm II}$ can be written as

$$\begin{aligned} \psi_{{\rm I},s} = (2\pi)^{-3/2}\exp(ik_1x + ik_2y + ik_3z - 1 - \tfrac{1}{2}t)[C_1' u_s\exp(i[(\sigma^1 k_1 + \sigma^3 k_3)t + \sigma^2 k_2\log t]) \\ + C_2'\hat{u}_s\exp(-i[(\sigma^1 k_1 + \sigma^3 k_3)t + \sigma^2 k_2\log t])] \end{aligned} \tag{3.26a}$$

and

$$\begin{aligned} \psi_{{\rm II},s} = i(2\pi)^{-3/2}(\sigma^1 k_1 + t^{-1}\sigma^2 k_2 + \sigma^3 k_3)^{-1}\exp(ik_1x + ik_2y + ik_3z - 1 - \tfrac{1}{2}t) \\ \times[C_3' u_s\exp(i[(\sigma^1 k_1 + \sigma^3 k_3)t + \sigma^2 k_2\log t]) + C_4'\hat{u}_s\exp(-i[(\sigma^1 k_1 + \sigma^3 k_3)t + \sigma^2 k_2\log t])] \end{aligned} \tag{3.26b}$$

On normalization of ψ, C_1', C_2', C_3', and C_4' are determined as

$C_1' = k^{-1}\exp(1.5)$, $C_2' = \exp(1.5)$,

$C_3' = \exp(1.5)$ and $C_4' = k\exp(1.5)$.

IV. CURRENT

The current is defined as

$$j^\mu = \bar{\psi}\gamma^\mu\psi \tag{4.1}$$

which is divergence free as $j^\mu_{;\mu} = 0$. If ψ is a massive field, j^μ can be rewritten as

$$j^\mu = \bar{\psi}\gamma^\mu\psi = (1/2m)\bar{\psi}(i\partial_\gamma\gamma^\lambda\gamma^\mu - i\gamma^\mu\gamma^\lambda\partial_\lambda - i[\gamma^\lambda\Gamma_\lambda,\gamma^\mu])\psi. \tag{4.2}$$

Equation (4.2) can be reexpressed as

$$j^\mu = (1/2m)(\bar{\psi}\sigma^{\lambda\mu}\psi)_{,\lambda} - (i/4m)g^{\mu\lambda}\bar{\psi}\overleftrightarrow{\partial}_\lambda\psi - (i/4m)\bar{\psi}([\gamma^\lambda_{,\lambda},\gamma^\mu] + [\gamma^\lambda,\gamma^\mu_{,\lambda}])\psi - (i/2m)\bar{\psi}[\gamma^\lambda\Gamma_\lambda,\gamma^\mu]\psi, \tag{4.3}$$

where $\bar{\psi}\overleftrightarrow{\partial}_\lambda\psi = \bar{\psi}\partial_\lambda\psi - \psi\partial_\lambda\bar{\psi}$. In the space-time considered here

$$\gamma^\lambda_{,\lambda} = 0, \quad [\gamma^\lambda,\gamma^\mu_{,\lambda}] = [\gamma^0,\gamma^\mu_{,0}], \quad [\gamma^0,\gamma^0_{,0}] = 0,$$

$$[\gamma^0,\gamma^1_{,0}] = [\tilde{\gamma}^0,\tilde{\gamma}^1](-a_1t^{-a_1-1}), \quad [\gamma^0,\gamma^2_{,0}] = [\tilde{\gamma}^0,\tilde{\gamma}^2](-a^2t^{-a_2-1}),$$

$$[\gamma^0,\gamma^3_{,0}] = [\tilde{\gamma}^0,\tilde{\gamma}^3](-a_3t^{-a_3-1}), \sigma^{01} = (i/2)\quad [\gamma^0,\gamma^1] = it^{-a_1}\tilde{\gamma}_0\tilde{\gamma}_1,$$

$$\sigma^{02} = (i/2)[\gamma^0,\gamma^2] = it^{-a_2}\tilde{\gamma}_0\tilde{\gamma}_2, \quad \sigma^{03} = (i/2)[\gamma^0,\gamma^3] = it^{-a_3}\tilde{\gamma}_0\tilde{\gamma}_3,$$

$$\sigma^{12} = (i/2)[\gamma^1,\gamma^2] = (i/2)t^{-(1-a_3)}[\tilde{\gamma}^1,\tilde{\gamma}^2], \quad \sigma^{23} = (i/2)[\gamma^2_0\gamma^3] = (i/2)t^{-(1-a_1)}[\tilde{\gamma}^2,\tilde{\gamma}^3],$$

$$\sigma^{31} = (i/2)[\gamma^3,\gamma^1] = (i/2)t^{-(1-a_2)}[\tilde{\gamma}^3,\tilde{\gamma}^1], \quad \Gamma_0 = 0, \quad \Gamma_1 = -\frac{a_1}{2t}t^{a_1}\tilde{\gamma}^1\tilde{\gamma}^0,$$

$$\Gamma_2 = -(a_2/2t)t^{a_2}\tilde{\gamma}^2\tilde{\gamma}^0, \quad \text{and} \quad \Gamma_3 = -(a_3/2t)t^{a_3}\tilde{\gamma}^3\tilde{\gamma}^0.$$

Now, from (4.3)

$$j^0 = +(i/2m)t^{-a_1}\partial_1(\bar{\psi}\tilde{\gamma}_0\tilde{\gamma}_1\psi) + (it^{-a_2}/2m)\partial_2(\bar{\psi}\tilde{\gamma}_0\tilde{\gamma}_2\psi) + (i/2m)t^{-a_3}\partial_3(\bar{\psi}\tilde{\gamma}_0\tilde{\gamma}_3\psi) - (i/4m)\bar{\psi}\overleftrightarrow{\partial}_0\psi, \tag{4.4}$$

$$j^1 = -\frac{i}{2m}\frac{\partial}{\partial t}(\bar{\psi}t^{-a_1}\tilde{\gamma}_0\tilde{\gamma}_1\psi) - \frac{i}{4m}t^{-(1-a_3)}\partial_2(\bar{\psi}[\tilde{\gamma}^1,\tilde{\gamma}^2]\psi) + \frac{i}{4m}t^{-(1-a_2)}\partial_3(\bar{\psi}[\tilde{\gamma}^3,\tilde{\gamma}^1]\psi) - \frac{i}{4m}t^{-2a_1}\bar{\psi}\overleftrightarrow{\partial}_1\psi - \frac{i}{4m}t^{-1-a_1}\bar{\psi}[\tilde{\gamma}^0,\tilde{\gamma}^1]\psi, \tag{4.5}$$

$$j^2 = -\frac{i}{4m}\frac{\partial}{\partial t}(\bar{\psi}t^{-a_2}\tilde{\gamma}_0\tilde{\gamma}_2\psi) + \frac{i}{4m}t^{-(1-a_3)}\partial_1(\bar{\psi}[\tilde{\gamma}^1,\tilde{\gamma}^2]\psi) - \frac{i}{4m}t^{-(1-a_1)}\partial_3(\bar{\psi}[\tilde{\gamma}^2,\tilde{\gamma}^3]\psi) - \frac{i}{4m}t^{-2a_2}\bar{\psi}\overleftrightarrow{\partial}_2\psi + \frac{a_2 i}{4m}t^{-a_2-1}\bar{\psi}[\tilde{\gamma}^0,\tilde{\gamma}^2]\psi - \frac{i}{4m}t^{-a_2-1}\bar{\psi}[\tilde{\gamma}^0,\tilde{\gamma}^2]\psi, \tag{4.6}$$

and

$$j^3 = -\frac{i}{2m}\frac{\partial}{\partial t}(\bar{\psi}t^{-a_3}\tilde{\gamma}_0\tilde{\gamma}_3\psi) - \frac{i}{4m}t^{-(1-a_2)}\partial_1(\bar{\psi}[\tilde{\gamma}^3,\tilde{\gamma}^1]\psi) + \frac{i}{4m}t^{-(1-a_1)}\partial_2(\bar{\psi}[\tilde{\gamma}^2,\tilde{\gamma}^3]\psi) - \frac{i}{4m}t^{-2a_3}\bar{\psi}\overleftrightarrow{\partial}_3\psi + \frac{a_3 i}{4m}\bar{\psi}[\tilde{\gamma}^0,\tilde{\gamma}^3]\psi - \frac{i}{4m}t^{-1-a_3}\bar{\psi}[\tilde{\gamma}^0,\tilde{\gamma}^3]\psi. \tag{4.7}$$

Writing

$$j^0 = \partial_1P_1 + \partial_2P_2 + \partial_3P_3 + \rho_{\text{convective}}$$

$(\rho_{\text{convective}} = j^0_{\text{convective}})$ and

$$\mathbf{j} = -\frac{\partial\mathbf{P}}{\partial t} + \nabla\times\mathbf{M} + \mathbf{j}_{\text{convective}} + \frac{i(a_1-1)}{2mt}P_1 + \frac{i(a_2-1)}{2mt}P_2 + \frac{i(a_3-1)}{2mt}P_3$$

the polarization densities P_1, P_2, and P_3 are given by

$$P_1 = (i/2m)t^{-a_1}\bar{\psi}\tilde{\gamma}_0\tilde{\gamma}_1\psi,$$

$$P_2 = +(i/2m)t^{-a_2}\bar{\psi}\tilde{\gamma}_0\tilde{\gamma}_2\psi,$$

and (4.8)

$$P_3 = +(i/2m)t^{-a_3}\bar{\psi}\tilde{\gamma}_0\tilde{\gamma}_1\psi$$

and $\mathbf{M}$ are given by

$$M_1 = (i/4mt)[t^{a_2}(\bar{\psi}[\tilde{\gamma}^3,\tilde{\gamma}^1]\psi) - t^{a_3}(\bar{\psi}[\tilde{\gamma}^1,\tilde{\gamma}^2]\psi)],$$

$$M_2 = (i/4mt)[t^{a_3}(\bar{\psi}[\tilde{\gamma}^1,\tilde{\gamma}^2]\psi) - t^{a_1}(\bar{\psi}[\tilde{\gamma}^2,\tilde{\gamma}^3]\psi)],$$

and (4.9)

$$M_3 = (i/4mt)[t^{a_1}(\bar{\psi}[\tilde{\gamma}^2,\tilde{\gamma}^3]\psi) - t^{a_2}(\bar{\psi}[\tilde{\gamma}^3,\tilde{\gamma}^1]\psi)]$$

(with regard to external electromagnetic field, $\mathbf{M}$ has the meaning of magnetization current density).

Connecting expressions for polarization density and magnetization density with solutions obtained in Sec. III, it is interesting to note that the magnitudes of P_1, P_2, and P_3 become almost equal for large t. Similarly magnitudes of M_1, M_2, and M_3 also come very close for large t. This shows that anisotropy might be possible for small t but when time increases anisotropy decreases.

In case of massless fields, Gordon decomposition of the current (discussed above for massive fields) is not possible. But here also, one finds that by using the solutions obeying Weyl constraint, the magnitude of the three space components of the current come close with advancement in time (t).

[1]G. V. Shishkin and I. E. Andrushkeirsh, Phys. Lett. A **110**, 84 (1985).
[2]A. O. Barut and I. H. Duru, Phys. Rev. D **36**, 3705 (1987).
[3]J. Audretsch and G. Safer, J. Phys. A **11**, 1583 (1978).
[4]E. A. Lord, *Tensors, Relativity and Cosmology* (Tata McGraw-Hill, India, 1976).
[5]J. D. Bjorken and S. D. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1964).
[6]G. M. Murphy, *Ordinary Differential Equations* (Van Nostrand, Englewood Cliffs, NJ, 1960).

Approximate relativistic quantum Hamiltonians for N interacting particle systems

E. Ruiz
Grupo de Física Teórica, Universidad de Salamanca, 37008 Salamanca, Spain

(Received 14 April 1989; accepted for publication 9 August 1989)

Time evolution of relativistic particle systems can be accomplished by means of a Schrödinger wave equation, provided the Hamiltonian of the N particle system satisfies some commutator equations involving the other generators of a suitable representation of the Poincaré group on the initial data space of the Schrödinger equation. This set of operator equations is solved in the N free particle case. Further, the structure of N interacting particle Hamiltonians is worked out to include second-order relativistic corrections, i.e., $(1/c^2)$-terms. It is shown that Breit and Barker–O'Connell Hamiltonians (nonzero spin particles) and the Bažański Hamiltonian (zero-spin particles) are particular cases of this more general Hamiltonian.

I. INTRODUCTION

Bel and Ruiz[1] have developed in a previous paper (hereafter we refer to it as BR) relativistic quantum mechanics for N particle systems based on a Schrödinger-like wave equation. Our purpose here is to apply this theoretical framework to two different classes of particle systems. First, we deal with free particle systems. In this case we are able to obtain exact expressions for all the infinitesimal generators of the BR representation of the Poincaré group, and the Hamiltonian as well. This example is a first test of the consistency of the theory and a helpful guide to the more interesting, but also more involved, interacting particle systems we deal with throughout Secs. III and IV. There we consider particle systems whose interaction reduces to a purely Newtonian potential in the nonrelativistic limit (i.e., two-body potential depending only on the distances between particles). We obtain the first relativistic corrections to the Hamiltonian. Moreover, we prove that the standard nonrelativistic scalar product is a suitable one for these kinds of particle systems up to this order of approximation. (This result is exact in the free particle system case.) This approximate Hamiltonian contains three arbitrary functions of some definite combinations of particle operators that depend on the type of interaction we want the Hamiltonian to describe. In Sec. V, we give the expressions of the unknown functions that fit the two most widely known second-order relativistic Hamiltonians: the Breit Hamiltonian and the Barker–O'Connell one.

We devote the rest of this section to outlining the principal features of the theoretical framework we use in this paper; for further details we refer the reader to the above-mentioned BR paper.

The BR starting point is a $(3N+1)$-dimensional surface in the momentum-space of the N particle system ($a, b,\ldots = 1,2,\ldots,N$ label the particle)

$$\sum : F_{a'}(k_b^2 \equiv -k_b^{\mu} k_{b\mu}) = c_{a'} \quad (a', b', \ldots = 2,3,\ldots,N). \tag{1.1}$$

Such a surface is invariant under the standard action of the Lorentz group on the momentum-space. It is assumed that Σ admits the parametric equations

$$k_a^0 = h_a(E, \mathbf{k}_b^2), \quad \mathbf{k}_a^2 = k_a^i k_{ai}, \tag{1.2}$$

where the h_a's obey the following constraint[2]:

$$\epsilon^a h_a(E, \mathbf{k}_b^2) = E \quad (\epsilon^a = +1). \tag{1.3}$$

Restricting the action of the Lorentz group to Σ, we get a nonlinear action of it as a family of one-one mappings of Σ onto Σ. The pullback under these mappings of functions $\varphi(E\,\mathbf{k}_b)$ defines a representation of the Lorentz group on that function space. Adding to these transformations the time-space translations, defined in the usual way it is done in momentum-space, we get the BR representation of the full Poincaré group [see BR equations (2.10), (2.12), and (2.13)].

Fourier transforming the operators of the above-mentioned representation, we get the position-space representation, that is, a representation of the Poincaré group acting on functions of the time coordinate and the spatial coordinates of all the particles $\psi(t, \mathbf{x}_b)$.

That tool makes it possible to discuss the relativistic invariance of a Schrödinger-like wave equation

$$i \frac{\partial}{\partial t} \psi = H\psi, \tag{1.4}$$

where

$$\psi = \psi_{\{P\}} \chi^{\{P\}}. \tag{1.5}$$

Here $\{P\} = P_1 P_2 \ldots P_N$ denotes a multi-index and $\chi^{\{P\}}$ is a basis of the spin-space of an N particle system; i.e., a basis of the tensor product of N complex vector spaces, one for each particle, of respective dimensions $2s_a + 1$ (s_a is the value of the spin of the particle labelled a). Equation (1.4) is said to be Poincaré invariant if their solutions go into solutions under the action of the position-space representation.

It turns out that (1.4) is Poincaré invariant if H does not depend on time and

$$[P^i, H] = [J^k, H] = 0, \tag{1.6a}$$

$$[\bar{K}^i, H] = iP^i, \tag{1.6b}$$

where P_i, J_j, and $\bar{K}_l$ are, respectively, the time-space translation, spatial rotation, and boost infinitesimal generators of the representation induced by the position-space representation on the initial data space of the Schrödinger equation (i.e., Heisenberg's representation). The operators P_i and J_j

have the familiar expressions

$$P^i = \epsilon^a p^i_a \quad \left(p_{ai} \equiv -i\frac{\partial}{\partial x^{ai}}\right),$$
$$J^i = \epsilon^i_{jk} x^{aj} p^k_a + S^i, \tag{1.7}$$

where S_i are the spatial rotation generators of a finite-dimensional representation of the Lorentz group on the spin-space of the N particle system (direct sum of Lorentz group representations on one-particle spin-spaces). The operator $\bar{K}_i$ has a rather involved expression that is given in the Appendix.

Let us summarize. To obtain relativistic invariant Schrödinger equations implies obtaining solutions of the highly nontrivial set of operator equations (1.6). That is the question we are going to deal with throughout the following sections.

II. FREE PARTICLE SYSTEMS

We consider in this section a system composed of N noninteracting particles. Even though it may be regarded up to a certain extent as an academic topic, we have at least two reasons for dedicating some attention to it. First, it can be exactly solved. Second, it actually provides useful information to treat the interacting case. We may further point out that, if interaction dies off when spatial separation between particles increases, the behavior of interacting systems must tend to that of free particle systems when the distances between any one particle and the others are large enough so their mutual influence can be neglected.

We are not going to solve Eqs. (1.6) imposing some reasonable conditions. We just propose to check the following Hamiltonian,[3]

$$H_F = g^a(\mathbf{p}_a^2 + m_a^2)^{1/2} \equiv g^a\omega_a \quad [(g_a)^2 = +1], \tag{2.1}$$

where m_a is a constant: the mass of the particle a.[4] We allow the square roots on the left-hand side of (2.1) to take both signs for the sake of completeness. Nevertheless we shall consider only the positive sign (real massive particles) in the following sections. Anyhow, the results we shall derive there can be obtained for a different choice of signs following the same line of reasoning.

It is straightforward to check that H_F verifies the pair of equations (1.6a). Equation (1.6b) requires an explicit expression for the boost generator. Setting $W = 0$ in the formula for $\bar{K}_i$ given in the Appendix (A18) [or (A19)], we get

$$\bar{K}^i_F = x^{ai}\bar{h}^{(0)}_a + Q^i \equiv x^{ai} h_a(H_F,\mathbf{p}_b^2) + Q^i \tag{2.2}$$

(Q_i is the boost infinitesimal generator of the finite-dimensional representation of the Lorentz group we have mentioned in Sec. I). The commutator of this operator with H_F reads

$$[\bar{K}^i_F, H_F] = ig^a\omega_a^{-1} h_a(H_F,\mathbf{p}_b^2)p^i_a. \tag{2.3}$$

We now recall the definition of the surface Σ (1.1) and its parametric equations (1.2) and (1.3). If we choose the constants $c_{a'} = F_{a'}(m_b{}^2)$, the implicit function theorem ensures that $h_a(H_F,p_a^2) = g_a\omega_a$. Hence we conclude that H_F verifies (1.6b).[5] Therefore H_F is a Poincaré invariant Hamiltonian, as was expected.

The operator P_i is Hermitian with respect to the standard nonrelativistic product

$$(\psi_1,\psi_2) = \int \psi_1^{*\{P\}}\psi_{2\{P\}}\, d\mathbf{x}_1 \wedge d\mathbf{x}_2 \wedge \cdots \wedge d\mathbf{x}_N, \tag{2.4}$$

where we have introduced the shorthand notation $d\mathbf{x}_a = dx_a{}^1 \wedge dx_a{}^2 \wedge dx_a{}^3$. [Hereafter, we shall say that an operator is Hermitian when it is Hermitian with respect to the scalar product (2.4).] However, it is well known that no representation of the Lorentz group on a finite-dimensional vector space can be unitary. Therefore S_i and Q_i may be nonunitary, and consequently J_i and $\bar{K}_{Fi}$ as well. As far as S_i closes the Lie algebra of the rotation group we can always assume that it is Hermitian. Then, J_i is also Hermitian. But $\bar{K}_{Fi}$ is clearly not, because neither Q_i nor the spin independent part of the operator is Hermitian.

Now we proceed to prove that it is possible to achieve Hermicity by means of a nonunitary transformation. We shall do it in two steps. Firstly, we perform the following change of representation:

$$\psi \to M\psi = \prod_{a=1}^{N} (\omega_a/m_a)^{1/2}\psi. \tag{2.5}$$

The operator M commutes with H_F, P_i, and J_k. Therefore they have the same expressions in the new representation. On the other hand,

$$\bar{K}^i_F \to M\bar{K}^i_F M^{-1} = \tfrac{1}{2}g^a(x^i_a\omega_a + \omega_a x^i_a) + Q^i. \tag{2.6}$$

To perform the second step, we have to specify the operators S_i and Q_j. Let S_{ai} and Q_{aj} be the infinitesimal generators of the Lorentz group representation on the spin-space of the particle a. We fix all these individual representations by making $Q_{ai} = \pm iS_{ai}$. For instance, if the particle a has spin $s_a = \frac{1}{2}$, $S_{ai} = \frac{1}{2}\sigma_i$ (σ_i's are Pauli's matrices), then $Q_{ai} = \pm\frac{1}{2}\sigma_i$, yielding to the $(0,\frac{1}{2})$ or $(\frac{1}{2},0)$, depending on the choice of sign, representations of the Lorentz group.[6] Having in mind that S_{ai} and Q_{aj} are one particle operators (i.e., they only act on their respective spin-spaces, ignoring the others), we can write

$$S^i = \epsilon^a S^i_a, \quad Q^i = i\delta^a S^i_a \quad [(\delta_a)^2 = +1]. \tag{2.7}$$

We try a transformation quite similar to the one that is used to go from the Dirac to the Foldy–Wouthuysen representation,[7] that is,

$$\psi \to L\psi = \prod_{a=1}^{N} \exp[\theta_a(|\mathbf{p}_a|)p^k_a S_{ak}]\psi, \tag{2.8}$$

where the function $\theta_a(|\mathbf{p}_a|)$ $[|\mathbf{p}_a| = (p_a{}^k p_{ak})^{1/2}]$ is defined by

$$\tanh(\theta_a|\mathbf{p}_a|) = \delta_a g_a(|\mathbf{p}_a|/\omega_a). \tag{2.9}$$

This transformation leaves unchanged H_F, P_i, and J_k, as well as the previous one. However, $\bar{K}_{Fi}$ comes to an explicit Hermitian form,

$$\bar{K}^i_F \to L\bar{K}^i_F L^{-1} = g^a[\tfrac{1}{2}(x^i_a\omega_a + \omega_a x^i_a) + (m_a + \omega_a)^{-1}\epsilon_{ikm}p^k_a S^m_a]. \tag{2.10}$$

The infinitesimal generators of the Poincaré group (1.7), (2.1), and (2.10) coincide with those of the Foldy–

Wouthuysen representation, either the positive or negative energy part of it, when N is set equal to 1, $s_1 = \frac{1}{2}$, and S_1 and Q_1 are the infinitesimal generators of the $(\frac{1}{2}, 0)$ or $(0, \frac{1}{2})$ representations of the Lorentz group we have mentioned above.

We would like to point out that expressions (1.7), (2.1), and (2.10) can be obtained by simply applying the correspondence principle to the constant of motion associated to the invariance under the Poincaré group of a classical N free particle system.[8]

III. APPROXIMATION SCHEME. NONRELATIVISTIC QUANTUM HAMILTONIANS

In BR the (local) equivalence between representations based on different surfaces (1.1) was proven. An extremely convenient choice of Σ is

$$\sum : k_{a'}^2 - k_1^2 = m_{a'}^2 - m_1^2. \tag{3.1}$$

The parametric equations of this surface can be written in a half-explicit form (we restrict ourselves to the region $k_a{}^0 > 0$)

$$k_{a'}^0 = f_{a'}(k_1^0, \mathbf{k}_b^2) = +[(k_1^0)^2 + \omega_{a'}^2 - \omega_1^2]^{1/2},$$
$$k_1^0 + \epsilon^{a'} f_{a'}(k_1^0, \mathbf{k}_b^2) = E. \tag{3.2}$$

The last equation defines implicitly $k_1{}^0 = h_1(E, \mathbf{k}_b{}^2)$; substituting this function in the other one, we get the rest of the parametric equations of Σ: $k_{a'}{}^0 = f_{a'}[h_1(E,\mathbf{k}_b{}^2), \mathbf{k}_c{}^2] \equiv h_{a'}(E, \mathbf{k}_d{}^2)$. In fact, we do not need to know how these functions are, because they enter into the expression of $\bar{K}_i$ [see (A18)] only by means of its derivatives with respect to E at the point $E = H_F \equiv \epsilon^a \omega_a$, and such derivatives can be obtained directly from (3.2).

It is easily shown that

$$\bar{h}_a^{(n)} \equiv \frac{\partial^n h_a}{\partial E}(H_F, \mathbf{k}_b^2) = \left(\Omega \frac{\partial}{\partial k_1^0}\right)^n f_a(k_1^0, \mathbf{k}_b^2)\Big|_{k_1^0 = \omega_1}, \tag{3.3}$$

where

$$\Omega^{-1} \equiv \epsilon^a \frac{\partial f_a}{\partial k_1^0}(k_1^0, \mathbf{k}_b^2), \tag{3.4}$$

and we have introduced the definition $f_1(k_1{}^0, \mathbf{k}_b{}^2) = k_1{}^0$.

Using a slightly different version of the formula (A8), we expand the operator on the right-hand side of (3.3) to get

$$\bar{h}_a^{(n)} = \sum_{l_1=0}^{1} \sum_{l_2=0}^{2-s_1} \cdots \sum_{l_{n-1}=0}^{n-1-s_{n-2}} \binom{1}{l_1} \times \binom{2-s_1}{l_2} \cdots \binom{n-1-s_{n-2}}{l_{n-1}} \times \Omega \Omega^{(l_1)} \cdots \Omega^{(l_{n-1})} f_a^{(n-s_{n-1})}\Big|_{k_1^0 = \omega_1}, \tag{3.5}$$

where $s_p = l_1 + l_2 + \cdots + l_p$ $(p \leqslant n)$ and $\Omega^{(r)} \equiv \partial^r \Omega / \partial k_1{}^0$.

We now proceed to use the *Faà* di Bruno formula[9] (an expression for the nth derivative of a composite function) to evaluate $\Omega^{(r)}$ and $f_a{}^{(m)}$ at $k_1{}^0 = \omega_1$. It leads to

$$f_a^{(m)}(\omega_1, \mathbf{k}_b^2) = \omega_1^{-m+1} P_a(\omega_1/\omega_{c'}),$$
$$\Omega^{(r)}(\omega_1, \mathbf{k}_b^2) = \omega_1^{-r} Q(\omega_1/\omega_{c'}), \tag{3.6}$$

where P_a is a polynomial in the $n-1$ variables $\omega_1/\omega_{c'}$ and Q is a rational function in the same set of variables. Inserting (3.6) into (3.5) we get

$$\bar{h}_a^{(n)} = \omega_1^{-n+1} R_a(\omega_1/\omega_{c'}), \tag{3.7}$$

where R_a is a rational function as well.

Restoring conventional units, where c is the speed of light in a vacuum, and assuming that $k_a{}^0$ and consequently E have energy dimensions, we have

$$\omega_a = c(m_a^2 c^2 + \mathbf{k}_a^2)^{1/2} \simeq m_a c^2 + 0(1). \tag{3.8}$$

Thus taking into account (3.7), we conclude that the leading term in the expansion of $\bar{h}_a{}^{(n)}$ in powers of $1/c^2$ is proportional to $(1/c^2)^{n-1}$.

This result lends support to an approximation scheme based on such a kind of expansion. To be precise, we assume that the Hamiltonian of the N particle system $H = H_F + W$ can be expanded in a formal power series of $1/c$,[10] that is,

$$H_F = \epsilon^a \omega_a \simeq Mc^2 + \sum_{a=1}^{N} \frac{\mathbf{p}_a^2}{2m_a} - \frac{1}{c^2} \sum_{a=1}^{N} \frac{\mathbf{p}_a^4}{8m_a^3} + \cdots,$$
$$W \simeq W_{(0)} + \frac{1}{c} W_{(1)} + \frac{1}{c^2} W_{(2)} + \cdots, \tag{3.9}$$

here M denotes the total mass of the system, $M = \epsilon^a m_a$. It induces a similar expansion of the boost generator $\bar{K}_a$. Moreover, if we let $\bar{K}_i$ have units of mass times a length, as it is used to be, and recall the dependence of $\bar{h}_a{}^{(n)}$ on $1/c^2$, it turns out by simple inspection of the formula (A19) that the nth term in the expansion of $\bar{K}_i$ is completely determined by the $(n-2)$th term in the expansion of H (3.9),

$$\bar{K}^i \simeq \sum_{a=1}^{N} m_a x_a^i + \frac{1}{c} \sum_{a=1}^{N} S_a^i + \frac{1}{c^2} \sum_{a=1}^{N} x_a^i \frac{\mathbf{p}_a^2}{2m_a} + \frac{\mu}{c^2} \sum_{a=1}^{N} x_a^i \frac{W_{(0)}}{m_a} + \cdots \quad \left(\frac{1}{\mu} = \sum_{a=1}^{N} \frac{1}{m_a}\right) \tag{3.10}$$

(S_{ai} has units of angular momentum). Neither P_i nor J_k have to be expanded.

Inserting (3.9) and (3.10) into Eqs. (1.6) and equating to zero terms of the same order in 1/c, we obtain an infinite set of operator equations for the infinite set of operators $W_{(n)}$. It has the structure of a ladder, in the very sense that the nth order equation is well posed [i.e., it is useful to determine the operator $W_{(n)}$] if and only if all the preceding equations have been solved; that is, we know the operators $W_{(m)}$ $m < n$.

The zeroth-order equations read

$$[P^i, W_{(0)}] = [J^k, W_{(0)}] = [\bar{K}^i_{(0)}, W_{(0)}] = 0, \tag{3.11}$$

where $\bar{K}_{(0)i}$ obviously stands for the first term on the right-hand side of (3.10): the infinitesimal generator of the Galileo transformations in nonrelativistic quantum mechanics (or despite a factor $1/M$ the classical definition of the center of mass). Therefore, we recover at this order of approximation the Lie algebra of the standard representation of the Galileo group used in nonrelativistic quantum mechanics.

The most general solution of (3.11) is an arbitrary function of all the scalars we can build up by combining the following operators:

$$x^i_{ab} \equiv x^i_a - x^i_b, \quad v^i_{ab} \equiv \frac{p^i_a}{m_b} - \frac{p^i_b}{m_b}, \quad S^i_a. \tag{3.12}$$

That leads to the well-known structure of a Galileo invariant nonrelativistic quantum Hamiltonian describing a system composed of N particles having arbitrary and unnecessary equal spins whose interaction can be described, in general, by a momentum dependent potential.[11]

We shall devote the next section to work out the first corrections to the nonrelativistic Hamiltonian.

IV. FIRST- AND SECOND-ORDER RELATIVISTIC CORRECTIONS

Since $\bar{K}^i$ has a $(1/c)$-term [see (3.10)], we must consider a term of this kind in the expansion of W. It has to verify the following set of equations:

$$[P^i, W_{(1)}] = [J^i, W_{(1)}] = 0, \tag{4.1a}$$

$$[\bar{K}^i_{(1)}, W_{(1)}] = -i \sum_{a=1}^{N} [S^i_a, W_{(0)}]. \tag{4.1b}$$

Even though we only know how the structure of $W_{(0)}$ is, we may evaluate the right-hand side of (4.1b) and try to solve these equations. However, we shall not do that. We recall that the presence of such a term was representation dependent in the free particle case [see (2.2) and (2.7)]. Here, it is as well. Let us perform the same change of representation we used in the free particle case (2.8), but retaining terms of order not higher than $1/c^2$. Then we get

$$\bar{K}^i \to L\bar{K}^i L^{-1} \simeq \sum_{a=1}^{N} m_a x^i_a + \frac{1}{c^2} \sum_{a=1}^{N} x^i_a \frac{\mathbf{p}^2_a}{2m_a} + \frac{\mu}{c^2} \sum_{a=1}^{N} x^i_a \frac{W_{(0)}}{m_a} + \frac{1}{2c^2} \epsilon^i{}_{kl} \sum_{a=1}^{N} \frac{1}{m_a} p^k_a S^l_a + \cdots, \tag{4.2}$$

$$H \to LHL^{-1} \simeq Mc^2 + \sum_{a=1}^{N} \frac{\mathbf{p}^2_a}{2m_a} + W_{(0)} + \frac{1}{c} {}^1W_{(1)} + \cdots .$$

P_i and J_k commute with L. (It will be true for every transformation we shall perform hereafter.) As far as we have gotten rid of the $(1/c)$-term of $\bar{K}^i$ in the new representation, we have modified the expansion of Eqs. (1.6): ${}^1W_{(1)}$, the new $(1/c)$-term in the expansion of H, does not have to satisfy (4.1) as $W_{(1)}$ has to do, but rather the same equations that $W_{(0)}$ did.

It is clear that we can obtain $W_{(1)}$ by transforming back LHL^{-1}. Nevertheless, it is more convenient in order to get the next term in the expansion of W to continue using the last representation (4.2) instead of the previous one (3.9), (3.10). That way, we avoid the annoying presence of ${}^1W_{(1)}$ in the equations for ${}^1W_{(2)}$, that is ${}^1W_{(1)}$ will not appear until the third-order approximation.

Hereafter we shall restrict ourselves to interactions that are purely Newtonian at the lowest order of approximation, that is, $W_{(0)}$ is a two-body potential depending only on the distances between pairs of particles, $r_{ab} \equiv +(x_{ab}{}^k x_{abk})^{1/2}$ $(a \neq b)$,

$$W_{(0)} = \frac{1}{2} \sum_{a=1}^{N} \sum_{a=1}^{N} V_{ab}(r_{ab}), \qquad V_{[ab]} = V_{aa} = 0. \tag{4.3}$$

Before solving, not even writing, the equations for ${}^1W_{(2)}$, we are going to search for a more suitable representation of the generator algebra. We turn back to formula (4.2) and examine it more carefully. Let us assume that the interaction is separable (i.e., $V_{ab} \to 0$ when $r_{ab} \to +\infty$) and picture a situation in which some of the particles are distant enough from the rest that we can ignore the influence of one group of particles on the other group. One may expect that the operator $\bar{K}^i$ can be split into two pieces each one depending on variables of one of the two clusters, but not on variables of the other one. Unfortunately, the expression (4.2) does not exhibit this physically reasonable clustering property. However, it is not an intrinsic property of the theory; it is just a peculiar feature of the representation we are using. This undesirable fact can be removed by an appropriate change of representation,

$$\psi \to C\psi \simeq \left\{ 1 - \frac{i}{4Mc^2} \frac{\mu}{\sigma} P_k \sum_{b=1}^{N} \sum_{c=1}^{N} (x^k_b - x^k_c) \Phi_{bc} + \cdots \right\} \Psi,$$

$$\Phi_{bc} = \sum_{d=1}^{N} (\rho_b V_{cd} - \rho_c V_{bd}), \tag{4.4}$$

where we have introduced the shorthand notations

$$\sigma \equiv \prod_{a=1}^{N} m_a, \quad \rho_a \equiv \frac{\sigma}{m_a} = \prod_{b \neq a} m_b. \tag{4.5}$$

A straightforward calculation leads to

$$\bar{K}^i \to C\bar{K}^i C^{-1} \simeq \sum_{a=1}^{N} m_a x^i_a + \frac{1}{c^2} \sum_{a=1}^{N} x^i_a \frac{\mathbf{p}^2_a}{2m_a} + \frac{1}{2c^2} \sum_{a=1}^{N} \sum_{b=1}^{N} V_{ab} x^i_a + \frac{1}{2c^2} \epsilon^i_{jk} \sum_{a=1}^{N} \frac{1}{m_a} p^j_a S^k_a + \cdots . \tag{4.6}$$

Moreover, CHC^{-1} differs from H in terms of order $1/c^2$. We collect all of them and ${}^1W_{(2)}$ in a unique term that we call ${}^2W_{(2)}$. Therefore, we have

$$H \to CHC^{-1} \simeq Mc^2 + \sum_{a=1}^{N} \frac{\mathbf{p}^2_a}{2m_a} + \frac{1}{2} \sum_{a=1}^{N} \sum_{b=1}^{N} V_{ab} + \frac{1}{c} {}^1W_{(1)} - \frac{1}{8c^2} \sum_{a=1}^{N} \frac{\mathbf{p}^4_a}{m^3_a} + \frac{1}{c^2} {}^2W_{(2)} + \cdots . \tag{4.7}$$

Finally, we bring $\bar{K}^i$ to a Hermitian form by applying the transformation (2.5) to (4.6),

$$\bar{K}^i \simeq \sum_{a=1}^N m_a x_a^i + \frac{1}{2c^2} \sum_{a=1}^N \left(x_a^i \frac{\mathbf{p}_a^2}{2m_a} + \frac{\mathbf{p}_a^2}{2m_a} x_a^i \right) + \frac{1}{2c^2} \sum_{a=1}^N \sum_{b=1}^N x_a^i V_{ab} + \frac{1}{2c^2} \sum_{a=1}^N \epsilon^i_{jk} \frac{p_a^j}{m_a} S_a^k + \cdots . \tag{4.8}$$

H differs from (4.7) in terms of order $1/c^2$, which we take into account by renaming the indeterminate $(1/c^2)$-piece of H: instead of ${}^2W_{(2)}$, now we write ${}^3W_{(2)}$.

Inserting (4.7) and (4.8) into Eqs. (1.6) (we recall that these equations are invariant under changes of representation) and omitting terms of order higher than $1/c^2$, we get

$$[P^i, {}^3W_{(2)}] = [J^k, {}^3W_{(2)}] = 0,$$

$$[\bar{K}^i_{(0)}, {}^3W_{(2)}] + \frac{1}{4} \sum_{a=1}^N \sum_{b=1}^N \sum_{c=1}^N \left\{ \left[x_a^i \frac{\mathbf{p}_a^2}{2m_a} + \frac{p_a^2}{2m_a} x_a^i + \epsilon^i_{jk} \frac{p_a^j}{m_a} S_a^k, V_{bc} \right] + 2\left[x_a^i V_{ab}, \frac{\mathbf{p}_a^2}{2m_a} \right] \right\} = 0. \tag{4.9}$$

A straightforward but lengthy calculation makes the above equations into the following:

$$[P^i, \chi_{(2)}] = [J^k, \chi_{(2)}] = [\bar{K}^l_{(0)}, \chi_{(2)}] = 0, \tag{4.10}$$

where

$$\chi_{(2)} = {}^3W_{(2)} + \frac{1}{8} \sum_{a=1}^N \sum_{b=1}^N \left\{ \frac{1}{m_a m_b} (p_{ak} p_{bl} V^{kl}_{ab} + V^{kl}_{ab} p_{ak} p_{bl}) + 4\epsilon^{klm} \frac{\partial V_{ab}}{\partial x_a^k} \frac{p_{al}}{m_a^2} S_{am} \right\}, \tag{4.11}$$

$$V^{kl}_{ab} = V_{ab} \delta^{kl} - \frac{\partial V_{ab}}{\partial x_k^a} x^1_{ab}.$$

Thus we conclude that $\chi_{(2)}$ is an arbitrary function of all the scalars we can build up with the operators (3.12) as well as $W_{(0)}$ and ${}^1W_{(1)}$.

The Hamiltonian up to this order of approximation reads

$$H \simeq Mc^2 + \sum_{a=1}^N \frac{\mathbf{p}_a^2}{2m_a} + \frac{1}{2} \sum_{a=1}^N \sum_{b=1}^N V_{ab} + \frac{1}{c} {}^1W_{(1)} - \frac{1}{c^2} \sum_{a=1}^N \frac{p_a^4}{8m_a^3} - \frac{1}{8c^2} \sum_{a=1}^N \sum_{b=1}^N \left\{ \frac{1}{m_a m_b} (p_{ak} p_{bl} V^{kl}_{ab} + V^{kl}_{ab} p_{ak} p_{bl}) + 4\epsilon^{klm} \frac{\partial V_{ab}}{\partial x_a^k} \frac{p_{al}}{m_a^2} S_{am} \right\} + \frac{1}{c^2} \chi_{(2)} \cdots . \tag{4.12}$$

We would like to remark that this equation only shows what an approximate relativistic N particle Hamiltonian must look like, whenever the interaction between particles has a reasonable nonrelativistic limit. If we want the Hamiltonian (4.12) to describe a system of particles acting on each other by means of a definite interaction (e.g., electromagnetic, gravitational, etc.), we should complete it by giving the adequate expressions to the indeterminate operators V_{ab}, ${}^1W_{(1)}$, and $\chi_{(2)}$.

Let us say something about these operators. The first one V_{ab} is, as we said before, a Newtonian potential energy, that is, it governs the dynamic of the system in the nonrelativistic limit. Moreover, V_{ab} also determines without ambiguity some $(1/c^2)$-terms of the relativistic Hamiltonian. The other operators ${}^1W_{(1)}$ and $\chi_{(2)}$ are obviously pure relativistic terms. However, as far as they are solutions of Eqs. (3.11) [or (4.10)], they have the structure of a general nonrelativistic potential energy, that is, they are functions of the distances between particles and the velocities of every particle relative to the others, apart from the spin operators of all of them. The above-mentioned structure suggests that ${}^1W_{(1)}$ and $\chi_{(2)}$ take into account the contributions to the energy related to the dynamic of a particle in an external field. In fact, if we let the mass of one of the particles be much smaller than the masses of the rest (it amounts to making the linear momenta of all the particles except the lightest one approximately equal to zero), all the Newtonian potential energy dependent second-order terms goes to zero except the spin-orbit interaction energy. Then, the potential energy of the particle in the field created by the others must be included in ${}^1W_{(1)}$ and $\chi_{(2)}$. (It may include static three-body interaction terms if we are dealing with a nonlinear theory.) Certainly, it does not prevent other interaction terms of a different kind from occurring in ${}^1W_{(1)}$ and $\chi_{(2)}$. However, we think that this situation will not, in general, appear at this order of approximation.

Let us learn more about the meaning of ${}^1W_{(1)}$ and $\chi_{(2)}$ by comparing (4.12) with some well-known second-order relativistic Hamiltonians.

V. SOME RELATIVISTIC HAMILTONIANS

We show in this section how the more widely known relativistic Hamiltonians can be accomplished by conveniently fixing the free operator functions that the general Hamiltonian we have derived in the preceding section (4.12) contains.

First of all, we consider Breit's Hamiltonian. It describes a system of N electromagnetically interacting particles having charges e_a and gyromagnetic factors λ_a. The operators should have the following expressions in order to fit Breit's Hamiltonian[12]:

$$V_{ab} = \frac{e_a e_b}{r_{ab}}, \quad {}^1W_{(1)} = 0,$$

$$\chi_{(2)} = -\frac{1}{2} \sum_{a=1}^N \sum_{b \neq a} \frac{e_a e_b}{m_a r_{ab}^3} \lambda_a \epsilon_{klm} x^k_{ab} v^m_{ab} S^l_a + \frac{1}{8} \sum_{a=1}^N \sum_{b \neq a} \frac{e_a e_b}{m_a m_b r_{ab}^3} \lambda_a \lambda_b S_{ak} S_{bl} \times (\delta^{kl} - 3r_{ab}^{-2} x^k_{ab} x^1_{ab}) + \text{“contact terms.”} \tag{5.1}$$

Substituting this expression into (4.12), we get a straightforward generalization of Breit's Hamiltonian (any number of particles is allowed and the values of the particle spins are not restricted to $\frac{1}{2}$), that is[13]

$$H_{\rm BREIT} \approx Mc^2 + \sum_{a=1}^{N} \frac{\mathbf{p}_a^2}{2m_a} + \frac{1}{2} \sum_{a=1}^{N} \sum_{b\neq a} \frac{e_a e_b}{r_{ab}} - \frac{1}{c^2} \sum_{a=1}^{N} \frac{\mathbf{p}_a^4}{8m_a^3} - \frac{1}{4c^2} \sum_{a=1}^{N} \sum_{b\neq a} \frac{e_a e_b}{m_a m_b r_{ab}} (\delta^{kl} + r_{ab}^{-1} x_{ab}^k x_{ab}^l) p_{ak} p_{bl}$$
$$- \frac{1}{2c^2} \sum_{a=1}^{N} \sum_{b\neq a} \frac{e_a e_b}{m_a^2 r_{ab}^3} (\lambda_a - 1) \epsilon_{klm} x_{ab}^k p_a^l S_a^m + \frac{1}{2c^2} \sum_{a=1}^{N} \sum_{b\neq a} \frac{e_a e_b}{m_a m_b r_{ab}^3} \lambda_a \epsilon_{klm} x_{ab}^k p_b^l S_a^m$$
$$+ \frac{1}{8c^2} \sum_{a=1}^{N} \sum_{b\neq a} \frac{e_a e_b}{m_a m_b r_{ab}^3} \lambda_a \lambda_b S_{ak} S_{bl} (\delta^{kl} - 3 r_{ab}^{-2} x_{ab}^k x_{ab}^l) + \text{"contact terms"} + \cdots . \qquad (5.2)$$

Our second example is the Barker–O'Connell Hamiltonian. It described a system of N nonzero spin particles interacting by means of gravitational forces in the framework of Einstein's theory.[14] Now the indeterminate operator should have the expressions

$$V_{ab} = -G \frac{m_a m_b}{r_{ab}}, \quad {}^1W_{(1)} = 0,$$
$$\chi_{(2)} = \frac{1}{2} \sum_{a=1}^{N} \sum_{b\neq a} \sum_{c\neq a} \frac{G^2 m_a m_b m_c}{r_{ab} r_{ac}} - \frac{3}{8} \sum_{a=1}^{N} \sum_{b\neq a} G m_a m_b (r_{ab}^{-1} \mathbf{v}_{ab}^2 + \mathbf{v}_{ab}^a r_{ab}^{-1})$$
$$+ 2 \sum_{a=1}^{N} \sum_{b\neq a} \frac{G m_b}{r_{ab}^3} \epsilon_{klm} x_{ab}^k v_{ab}^l S_a^m - \frac{1}{2} \sum_{a=1}^{N} \sum_{b\neq a} \frac{G}{r_{ab}^3} S_{ak} S_{bl} (\delta^{kl} - 3 r_{ab}^{-2} x_{ab}^k x_{ab}^l) + \text{"contact terms."} \qquad (5.3)$$

Inserting (5.3) into (4.12), we obtain

$$H_{\rm B\text{-}OC} \approx Mc^2 + \sum_{a=1}^{N} \frac{\mathbf{p}_a^2}{2m_a} - \frac{1}{2} \sum_{a=1}^{N} \sum_{b\neq a} \frac{G m_a m_b}{r_{ab}} - \frac{1}{c^2} \sum_{a=1}^{N} \frac{\mathbf{p}_a^4}{8m_a^3} - \frac{3}{4c^2} \sum_{a=1}^{N} \sum_{b\neq a} \frac{G m_b}{m_a} (r_{ab}^{-1} \mathbf{p}_a^2 + \mathbf{p}_a^2 r_{ab}^{-1})$$
$$+ \frac{7}{8c^2} \sum_{a=1}^{N} \sum_{b\neq a} G (r_{ab}^{-1} p_a^k p_{bk} + p_a^k p_{bk} r_{ab}^{-1}) + \frac{1}{8c^2} \sum_{a=1}^{N} \sum_{b\neq a} G (r_{ab}^{-3} x_{ab}^k x_{ab}^l p_{ak} p_{al} + p_{ak} p_{bl} x_{ab}^k x_{ab}^l r_{ab}^{-3})$$
$$+ \frac{3}{2c^2} \sum_{a=1}^{N} \sum_{b\neq a} \frac{G m_b}{m_a r_{ab}^3} \epsilon_{klm} x_{ab}^k p_a^l S_a^m - \frac{2}{c^2} \sum_{a=1}^{N} \sum_{b\neq a} \frac{G}{r_{ab}^3} \epsilon_{klm} x_{ab}^k p_b^l S_a^m$$
$$- \frac{1}{2c^2} \sum_{a=1}^{N} \sum_{b\neq a} \frac{G}{r_{ab}^3} S_{ak} S_{bl} (\delta^{kl} - 3 r_{ab}^{-2} x_{ab}^k x_{ab}^l) + \frac{1}{2c^2} \sum_{a=1}^{N} \sum_{b\neq a} \sum_{c\neq a} \frac{G^2 m_a m_b m_c}{r_{ab} r_{ac}} + \text{"contact terms"} + \cdots . \qquad (5.4)$$

Strictly speaking, this is not the Barker–O'Connell Hamiltonian.[15] If we drop all the spin dependent terms in (5.4) (i.e., system of spinless particles), we recover the Einstein–Infeld–Hoffmann Hamiltonian.[16] If we set the number of particles $N = 2$ and ignore the quadratic term in the gravitational constant G, (5.4) reduces to the Hamiltonian obtained by Ibáñez and Martın.[17] By writing the last mentioned Hamiltonian in the center of mass $P_i = 0$ and making a coordinate transformation,[18] it coincides with the Barker–O'Connell Hamiltonian if the quadrupole interaction terms are omitted.

Another important Hamiltonian is the Bažański Hamiltonian.[19] It takes account of the interaction on N spinless charged particles in the framework of Einstein's theory up to the inclusion of $1/c^2$ terms. Even though we are not going to write it explicitly here, let us say that the interaction terms in this Hamiltonian that are not included in the Darwin Hamiltonian or the Einstein–Infeld–Hoffmann Hamiltonian, so that they describe the coupling between the electromagnetic and gravitational fields, are functions just of the operators r_{ab}. Therefore, we need only to add them to the operator $\chi_{(2)}$ given by (5.3) after setting $S_{ai} = 0$ to obtain the adequate $\chi_{(2)}$ for this new Hamiltonian [the operator $\chi_{(2)}$ corresponding to the electromagnetic case (5.1) vanishes for spinless particles]. Obviously V_{ab} must be the sum of the electromagnetic and gravitational Newtonian potential energies.

VI. REMARKS

Our aim in this paper has been to show that the BR relativistic quantum theory has a good agreement with the more basic and widely accepted results on relativistic N particle systems. On the other hand, we also have attempted to elucidate as to what are the utilities of a theory of this kind. It is clear that the BR theory is a likely framework. It tells us the structure that an N interacting particle Hamiltonian must have in order to be Poincaré invariant. However, it does not explain to us what we should do to work out the Hamiltonian that describes a system of particles interacting by means of the particular forces we are interested in.

This does not mean that the BR framework may not be useful to solve some problems. It can be done if we complete the BR framework with some good guessing work. We have pointed out in the last paragraph of Sec. IV that the indeterminate operators that the Hamiltonian (4.12) includes may be related to a simpler dynamical problem: the motion of a single particle in an external field. In fact, it can be checked that $\chi_{(2)}$ only contains terms of this kind in both of the cases we have dealt with in the previous section. Moreover, $\chi_{(2)}$ can be obtained from two-body terms (one particle in the field created by another) by making them symmetrical under arbitrary changes of particle labelling; the only exception is the G^2 term in the Barker–O'Connell Hamiltonian that is a three-body term, even though it also takes account of the potential energy of a particle in a field.

Another important feature of the BR scheme is that it gives explicit expressions for the infinitesimal generators of the Poincaré group; mainly, it shows how to build up the boost generator, provided that the Hamiltonian is known. It has permitted showing that the boost generator is Hermitian and exhibits good clustering properties in the representation; it is commonly used to write down the Breit and the Barker–O'Connell Hamiltonians.

Finally, we would like to make some comments on the set of equations (1.6), the expansion in powers of $1/c$, and the relativistic invariant scalar product. Though we have demonstrated that there are approximate solutions of Eqs. (1.6), we have stopped at the crucial point. Beyond the second-order approximation, one can ignore neither radiative effects nor that Lagrangians should depend on accelerations and higher-order derivatives of the position coordinates.[20] Fortunately, these nonstandard classical Lagrangians[21] admit a kind of Legendre's transformation; it leads to a Hamiltonian that is a function of the canonical coordinates and momenta, but does not depend on the derivatives of these quantities.[22] Therefore, one may expect that there also exists quantum partners to these classical Hamiltonians; that is, there are solutions of Eqs. (1.6) up to any order of approximation. Anyhow, it is far from being evident, due to the highly nonlinear character of the system (1.6). The nature of the relativistic scalar product is another question related to the problem of existence of solutions. We have proved that one can use the nonrelativistic scalar product in the relativistic scheme at least up to the second-order of approximation. It raises the question of whether it will be compatible with higher relativistic corrections. These two problems will be the matter of a forthcoming paper.

ACKNOWLEDGMENTS

The author would like to thank Professor Bel for many fruitful and pleasant discussions.

This work has been supported by a grant from the Ministerio de Educación y Ciencia under the Plan de Formación de Personal Investigador scheme.

APPENDIX: THE BOOST GENERATOR

The Boost generator formula given in BR [BR equation (3.4)] is not very useful for the sort of perturbation calculus to which this paper is devoted. We derive in this Appendix a different expression for this infinitesimal generator that may be regarded as a power series expansion in some nonspecified coupling constant.

Our departing point is the position-space representation that is linked to the momentum-space representation by means of the invariant Fourier transform [BR equation (5.17)] instead of the ordinary Fourier transform. As we said in the beginning of the paper, there is a representation on the initial data space of the Schrödinger equation associated to the position-space representation (induced representation). A formal expression for the boost generator of this last representation can be easily obtained from the BR equation (5.19) by using the technique explained in the BR appendix. [It differs slightly from the above-mentioned BR equation (3.4) since we are using another position-space representation.] This procedure leads to

$$\bar{K}^i = x^{ai} \sum_{n=0}^{\infty} \frac{1}{n!} \frac{\partial^n h_a}{\partial E}(0,\mathbf{p}_b^2) H^n + Q^i. \tag{A1}$$

Let us assume that the functions $h_a(E,\mathbf{k}_b{}^2)$, which define the parametric equations of the surface Σ (1.1), are analytic in a neighborhood of the subset $E = H_F = \epsilon^a \omega_a$, that is,

$$h_a(E,\mathbf{k}_b^2) = \sum_{n=1}^{\infty} \frac{1}{n!} \bar{h}_a^{(n)} (E - H_F)^n,$$
$$\bar{h}_a^{(n)} \equiv \frac{\partial^n h_a}{\partial E}(H_F,\mathbf{k}_b^2). \tag{A2}$$

Then, the derivatives of $h_a(E,\mathbf{k}_b{}^2)$ are series of positive powers of E and therefore their respective Fourier transforms are well-defined operators (at least in all that it may concern to E),

$$\frac{\partial^n h_a}{\partial E}(T,\mathbf{p}_b^2) = \sum_{r=0}^{\infty} \sum_{m=n+r}^{\infty} \frac{(-1)^{m-n-r}}{r!(m-n-r)!} \bar{h}_a^{(m)} H_F^{m-n-r} T^r \tag{A3}$$

($T = i\partial/\partial t$ is the position-space version of the E operator). The operators (A3) acting on initial data space (functions independent of t) is what we have called $(\partial^n h_a/\partial E)(0,\mathbf{p}_b{}^2)$ in formula (A1). Thus a formal expression for these last operators can be easily obtained from (A3) by simply setting $T = 0$.

Combining (A1) and (A3), we get

$$\bar{K}^i = x^{ai} \sum_{n=0}^{\infty} \sum_{m=n}^{\infty} \frac{(-1)^{m-n}}{n!(m-n)!} \bar{h}_a^{(m)} H_F^{m-n} H^n + Q^i. \tag{A4}$$

We now write the Hamiltonian as a sum of two terms, the free Hamiltonian H_F and an interaction term W. Then, the nth power of H reads

$$H^n = (H_F + W)^n = \sum_{r=0}^{n} \{H_F^{n-r}, W^r\}, \tag{A5}$$

where

$$\{H_F^{n-r}, W^r\} \equiv \sum_{k_1 + \cdots + k_{r+1} = n-r} H_F^{k_{r+1}} W H_F^{k_r} \cdots W H_F^{k_1},$$
$$\{H_F^n, W^0\} \equiv H_F^n \quad (r=0). \tag{A6}$$

We use the identity

$$W H_F^k = \sum_{r=0}^{k} \binom{k}{r} H_F^{k-r} [W,H_F]^{(r)},$$
$$[W,H_F]^{(r)} \equiv \underbrace{[[\cdots[W,H_F]H_F]\cdots H_F]H_F]}_{(r\text{-times})} \tag{A7}$$

to bring all the terms on the right-hand side of (A6) to the following form:

$$H_F^{k_{r+1}} W H_F^{k_r} \cdots W H_F^{k_1}$$
$$= \sum_{l_1=0}^{m_1} \sum_{l_2=0}^{m_2-s_1} \cdots \sum_{l_r=0}^{m_r-s_{r-1}} \binom{m_1}{l_1} \times \binom{m_2-s_1}{l_2} \cdots \binom{m_r-s_{r-1}}{l_r}$$
$$\times H_F^{m_{r+1}-s_r} [W,H_F]^{(l_r)} \cdots [W,H_F]^{(l_1)}, \quad (A8)$$

where we have introduced the notations $s_p = l_1 + \cdots + l_p$ and $m_P = k_1 + \cdots + k_P$.

Substituting this expression into (A6) and changing the order of the sums over the k_P's and l_P's indices, we get

$$\{H_F^{n-r}, W^r\} = \sum_{l_1=0}^{n-r} \sum_{l_2=0}^{n-r-s_1} \cdots \sum_{l_r=0}^{n-r-s_{r-1}} D_{(l_1,\cdots,l_r)}^{(n-r)} H_F^{n-r-s_r} \times [W,H_F]^{(l_r)} \cdots [W,H_F]^{(l_1)}, \quad (A9)$$

where

$$D_{(l_1,\cdots,l_r)}^{(n-r)} \equiv \sum_{m_r=s_r}^{n-r} \sum_{m_{r-1}=s_{r-1}}^{m_r} \cdots \sum_{m_1=s_1}^{m_2} \binom{m_1}{l_1} \binom{m_2-s_1}{l_2} \times \cdots \times \binom{m_r-s_{r-1}}{l_r}. \quad (A10)$$

Let us come back to (A4). Substituting (A5) into (A4) and reordering the sums over the n, m, and r indices, we obtain

$$\bar{K}^i = x^{ai} \sum_{r=0}^{\infty} \sum_{m=r}^{\infty} \sum_{n=r}^{\infty} \frac{(-1)^{m-n}}{n!(m-n)!} \bar{h}_a^{(m)} \times H_F^{m-n} \{H_F^{n-r}, W^r\} + Q^i. \quad (A11)$$

We now take into account (A9) and rewrite (A11) moving the sum over n to the right side of the block of sums over the l_p's indices; so we get

$$\bar{K}^i = x^{ai} \sum_{m=0}^{\infty} S^{(m,0)} \bar{h}_a^{(m)} H_F^m + x^{ai} \sum_{r=1}^{\infty} \sum_{m=r}^{\infty} \sum_{l_1=0}^{m-r} \sum_{l_2=0}^{m-r-s_1} \cdots \sum_{l_r=0}^{m-r-s_{r-1}} S_{(l_1,\cdots,l_r)}^{(m,r)} \times \bar{h}_a^{(m)} H_F^{m-r-s_r} [W,H_F]^{(1_r)} \times \cdots \times [W,H_F]^{(l_1)} + Q^i, \quad (A12)$$

where

$$S^{(m,0)} \equiv \sum_{n=0}^{m} \frac{(-1)^{m-n}}{n!(m-n)!} = \delta_0^m,$$
$$S_{(l_1,\cdots,l_r)}^{(m,r)} \equiv \sum_{n=r+s_r}^{m} \frac{(-1)^{m-n}}{n!(m-n)!} D_{(l_1,\cdots,l_r)}^{(n-r)} \quad (r \geqslant 1, s_r \leqslant m-r). \quad (A13)$$

Changing the order of the sums one more time, we obtain

$$\bar{K}^i = x^{ai} \bar{h}_a^{(0)} + x^{ai} \sum_{r=1}^{\infty} \sum_{l_1=0}^{\infty} \cdots \sum_{l_r=0}^{\infty} \sum_{m=r+s_r}^{\infty} S_{(l_1,\cdots,l_r)}^{(m,r)} \times \bar{h}_a^{(m)} H_F^{m-n-s_r} [W,H_F]^{(l_r)} \cdots [W,H_F]^{(l_1)} + Q^i. \quad (A14)$$

Now we proceed to evaluate the numerical coefficients (A13). First of all we calculate (A10) by applying successively the combinatorial identity

$$\sum_{l=\alpha}^{\beta} \binom{l}{\alpha} = \binom{\beta+1}{\alpha+1}. \quad (A15)$$

It leads to

$$D_{(l_1,\cdots,l_r)}^{(n-r)} = \binom{n}{r+s_r} \prod_{p=1}^{r} \binom{s_P+p-1}{l_P}. \quad (A16)$$

Inserting this expression into (A13), we have

$$S_{(l_1,\cdots,l_r)}^{(m,r)} = \delta_{r+s_r}^m \prod_{p=1}^{r} \frac{1}{l_P!(s_P+p)} \equiv \delta_{r+s_r}^m \Gamma_{(l_1,\cdots,l_r)}^{(r)}. \quad (A17)$$

It turns out that the sum over m in (A14) is trivial. Therefore we finally obtain

$$\bar{K}^i = x^{ai} \bar{h}_a^{(0)} + x^{ai} \sum_{r=1}^{\infty} \sum_{l_1=0}^{\infty} \cdots \sum_{l_r=0}^{\infty} \Gamma_{(l_1,\cdots,l_r)}^{(r)} \times \bar{h}_a^{(r+s_r)} [W,H_F]^{(l_r)} \cdots [W,H_F]^{(l_1)} + Q^i. \quad (A18)$$

A more comprehensive formula can be derived by isolating in the preceding one all the terms in the multiple sum corresponding to $l_1 = l_2 = \cdots = l_r = 0$, that is

$$\bar{K}^i = x^{ai} \sum_{r=0}^{\infty} \frac{1}{r!} \bar{h}_a^{(r)} W^r + x^{ai} \sum_{r=1}^{\infty} \sum_{\substack{l_1,\cdots,l_r=0 \\ s_r>0}}^{\infty} \Gamma_{(l_1,\cdots,l_r)}^{(r)} \times \bar{h}_a^{(r+s_r)} [W,H_F]^{(l_r)} \cdots [W,H_F]^{(l_1)} + Q^i. \quad (A19)$$

The first term on the right-hand side of this expression is the usual Taylor expansion at the point $E = H_F$; the second one contains all of the quantum corrections to the classical formula coming from the fact that W and H_F do not commute.

[1] L. Bel and E. Ruiz, J. Math. Phys. **28**, 18 (1988).

[2] Summation over repeated upper and lower indices is assumed. Nevertheless we sometimes make explicit the summation symbol, mainly in the last sections.

[3] This Hamiltonian is the sum of N standard one particle Hamiltonians. See, for instance, S. Schweber, *An Introduction to Relativistic Quantum Field Theory* (Harper & Row, New York, 1961).

[4] The action of the operator H_F is perfectly well defined in the momentum-space. See, for instance, Ref. 3.

[5] These calculations are more clearly and rigorously made using the momentum-space representation [BR equation (5.19)]. This last procedure requires solving the Schrödinger equation in the momentum-space $(H_F - E)\phi(E,\mathbf{k}_b) = 0$; the general solution of this equation is $\phi(E,\mathbf{k}_a) = \theta(\mathbf{k}_a)\delta(E - H_F)$.

[6] See, for instance, S. Schweber (Ref. 3).

[7] L. L. Foldy and S. A. Wouthuysen, Phys. Rev. **78**, 29 (1950); J. Bjorken and S. Drell, *Relativistic Quantum Mechanics* (McGraw-Hill, New York, 1965).

[8] We refer to the work on these classical systems by L. Bel and J. Martin, Ann. Inst. Henri Poincaré A **33**, 409 (1980). The description of the system is made by means of an extension to N particles of the one-particle symplectic form worked out by J. M. Souriau, *Structure des Systèmes Dynamiques* (Dunod, Paris, 1970).

[9] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1972).

[10] It amounts to saying that the system will be accurately described by some few terms of the series whenever the energies involved are much smaller than the rest masses of the particles. It means that we are dealing with true massive particles, $m_a > 0$ for $a = 1,2,\cdots,N$; zero-mass particles are excluded from our analysis.

[11]See, for instance, K. Gotfried, *Quantum Mechanics* (Benjamin, London, 1974).

[12]Even though contact terms can obviously be included in $\chi_{(2)}$, we have preferred to omit them, since their usual expressions come from spin-$\frac{1}{2}$ particle calculations and perhaps they may be rather different for some other kind of particles.

[13]H. A. Bethe and E. E. Salpeter, *Quantum Mechanics of One and Two Electron Atoms* (Springer–Verlag, Berlin, 1957); V. Berestetski, E. Lifchitz, and L. Pitayevski, *Théorie Quantique Relativiste* (Mir, Moscow, 1972); a classical version of the Breit Hamiltonian that allows arbitrary values of the intrinsic angular momenta (classical spin) can be found in L. Bel and J. Martín, Ann. Inst. Henri Poincaré A **34**, 235 (1981).

[14]For gravitational interaction in theories other than general relativity (PPN with parameters β and γ), see B. M. Barker and R. F. O'Connell, Phys. Rev. D **14**, 861 (1976). The two-particle Hamiltonian that is given in this paper can be achieved by making the numerical constant that precedes each multiple sum in the expression of the operator $\chi_{(2)}$ corresponding to the Barker–O'Connell Hamiltonian (5.3) into a simple function of the parameters β and γ.

[15]B. M. Barker and R. F. O'Connell, Phys. Rev. D **12**, 329 (1975); Gen. Relativ. Gravit. **11**, 149 (1979).

[16]L. Landau and E. Lifchitz, *Théorie du Champ* (Mir, Moscow, 1966).

[17]J. Ibáñez and J. Martín, Gen. Relativ. Gravit. **14**, 439 (1982).

[18]See the first source in Ref. 15.

[19]S. Bažański, in *Recent Developments in General Relativity* (Pergamon, New York, 1962); B. M. Barker and R. F. O'Connell, J. Math. Phys. **18**, 1818 (1977).

[20]J. Martín and J. L. Sanz, J. Math. Phys. **19**, 780 (1978); **20**, 25 (1979).

[21]The electromagnetic Lagrangian up to order $1/c^4$ is discussed in B. M. Barker and R. F. O'Connell, Ann. Phys. (NY) **129**, 358 (1980); the gravitational Lagrangian in general relativity up to the same order is discussed in T. Damour, in *Gravitational Radiation*, edited by N. Deruelle and T. Piran (North-Holland, Amsterdam, 1983).

[22]X. Jaén, J. Llosa, and A. Molina, Phys. Rev. D **34**, 2302 (1986).

Quantum mechanics on p-adic fields

Ph. Ruelle[a] and E. Thiran[b]
Institut de Physique Théorique, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium
D. Verstegen
NIKHEF-H, P.O. Box 41882, 1009 DB Amsterdam, The Netherlands
J. Weyers
Institut de Physique Théorique, Université Catholique de Louvain, B-1348 Louvain-la-Neuve, Belgium

(Received 8 April 1989; accepted for publication 21 June 1989)

A formulation of quantum mechanics on p-adic number fields is presented. Quantum amplitudes are taken as complex functions of p-adic variables and it is shown how the Weyl approach to quantum mechanics can be generalized to the p-adic case. The p-adic analogs of simple one-dimensional systems (free particle, compact and noncompact oscillators) are defined by a "group of motion," which is an Abelian subgroup of SL $(2,\mathbb{Q}_p)$. In each case the evolution operator is a unitary representation of the appropriate group. Its spectrum is given by characters and its eigenstates are calculated.

I. INTRODUCTION

To describe the physical world one obviously needs a number field. In classical physics, the field of real numbers, $\mathbb{R}$, is clearly singled out, while in quantum mechanics it is the field of complex numbers, $\mathbb{C}$, that is of fundamental importance.[1]

A rather remarkable mathematical theorem[2] states that, besides $\mathbb{R}$ and $\mathbb{C}$, the only other number fields with reasonable algebraic and topological properties—which will be stated explicitly below—are the fields $\mathbb{Q}_p$ (see Refs. 3 and 4) of p-adic numbers or algebraic extensions thereof. The essential difference between $\mathbb{R}$ (or $\mathbb{C}$) and the $\mathbb{Q}_p$'s lies in a property of their respective metric: in the former case, the metric satisfies the usual triangular inequality, while in the latter, a much stronger inequality, often called ultrametricity, holds.

It is precisely via the ultrametric structure of the ground states of spin glasses[5] that, to our knowledge, $\mathbb{Q}_p$ fields were first mentioned in the physics literature.

In a completely different context, Volovich and collaborators[6] have suggested various uses of p-adic fields either as the expression of fundamentally new characteristics of space and time at the Planck scale or as tools to investigate the behavior of various physical systems. This has led to a growing interest in analyzing properties of a physical theory when defined over p-adic[7–9] (or even finite[10]) number fields. The relevance of such investigations is certainly obvious in the context of string theories[11]: since the world sheet parameters are intrinsically nonobservable, it is certainly legitimate and worthwhile to investigate the structure of strings when these parameters are p-adic. Interestingly enough some of the arithmetic ingredients of string theory such as modular forms have p-adic analogs as well.[12] One of the first important uses of p-adic numbers in string theory was by Freund and Witten (see Ref. 11). Their work certainly strengthens the idea that a "fundamental theory" is not only independent of the parametrization one uses, which is certainly a key input of general relativity or conformal field theories, but also that, in some sense, it does not really depend on the number field in which the parametrization is expressed. We find this idea rather fascinating.

Clearly if one wants to use $\mathbb{Q}_p$ in string theory, it is useful and important to learn first how to formulate quantum mechanics when the underlying degrees of freedom or parameters are p-adic.

Our purpose in this paper is to develop this formulation *ab initio* in the study of simple one-dimensional systems whose classical (real) Lagrangian is quadratic. This includes the cases of a free particle, a particle in a constant field, and a compact or noncompact harmonic oscillator. In the p-adic version of these problems quantum amplitudes and wave functions will be taken as complex valued functions of p-adic variables. This point of view was originally suggested, among other possibilities, by Volovich.[6] Freund and Olson (see Ref. 7) then proposed that matrix elements of the quantum evolution operator be taken as proportional to additive characters, $\exp(2\pi i S_{\rm cl})$, where $S_{\rm cl}$ is the classical action expressed in terms of initial and final p-adic positions. The present authors with Alacoque later elaborated[8] on this idea by emphasizing the group theoretical structure; using the p-adic Gaussian integral, we checked that the Freund–Olson proposal indeed gives the correct evolution operator. In this paper we will justify the Freund–Olson ansatz by considering the Weyl formulation of quantum mechanics,[13] which involves finite transformation operators instead of infinitesimal ones. This makes it particularly suited for formulating quantum mechanics on p-adic fields. Indeed, an important by-product of ultrametricity is that the fields $\mathbb{Q}_p$ are disconnected: they are the union of disjoint closed sets.[4] It follows that one does not really have the usual notion of a path in $\mathbb{Q}_p$: adding many "infinitesimal" displacements does not build up a finite one! An unavoidable consequence of this

---
[a] Research assistant at Institut Interuniversitaire des Sciences Nucleaires (IISN).
[b] Research assistant at Fonds National de Recherche Scientifique (FNRS).

fact is that "infinitesimal generators" cannot be properly defined on $\mathbb{Q}_p$ and this clearly precludes a naive Schrödinger-like formulation of p-adic quantum mechanics.

To make this paper as self-contained as possible, we start in Sec. II with a brief summary of the salient features of $\mathbb{Q}_p$. We then sketch in Sec. III the Weyl formulation of ordinary quantum mechanics (i.e., over $\mathbb{R}$) as well as its generalization to $\mathbb{Q}_p$.

When studied over $\mathbb{R}$, the classical evolution matrices of the systems we consider here belong to Abelian subgroups of $SL(2,\mathbb{R})$. Going from $\mathbb{R}$ to $\mathbb{Q}_p$ thus leads us to study Abelian subgroups of $SL(2,\mathbb{Q}_p)$. This is done, in some detail, in Sec. IV and Appendix C. As in the real case, we will distinguish three classes of subgroups (parabolic, hyperbolic, and elliptic[14]) and the elucidation of the precise structure of these subgroups will allow for an easy description of their unitary characters (i.e., one-dimensional unitary representations).

In Sec. V we determine the eigenvalues and eigenfunctions of the evolution operators for the p-adic free particle (parabolic group) and noncompact oscillator (hyperbolic group). In Sec. VI we solve the corresponding problem for the p-adic compact harmonic oscillators (elliptic groups). In all cases the eigenvalues of the evolution operators are unitary characters of the relevant subgroup of $SL(2,\mathbb{Q}_p)$.

Section VII contains our conclusions.

Several Appendices are included to deal with more technical aspects of our study. Appendix A is devoted to p-adic integration with particular emphasis on Gaussian integrals. In Appendix B we consider the case when the classical evolution is inhomogeneous, namely, the problem of a particle in a constant force field. Appendix C, as already mentioned, continues Sec. IV and contains, in particular, an explicit description of the elliptic subgroups in the more difficult cases. We are not aware of any reference containing these results. Appendix D contains the explicit evaluation of some integrals and, finally, Appendix E presents a method for calculating the trigonometric sums that appear in the harmonic oscillator eigenfunctions.

II. THE p-ADIC NUMBER FIELDS $\mathbb{Q}_p$

In this section we briefly review definitions and properties of various number fields.[4] Our description of $\mathbb{Q}_p$, the field of p-adic numbers, is, of course, non exhaustive. The interested reader should consult the mathematical literature for more details and proofs. We begin with a general description of the field $\mathbb{Q}_p$, then discuss quadratic extensions $\mathbb{Q}_p(\sqrt{\tau})$, and finally collect useful formulas for p-adic "circles."

A. General properties

The simplest example of a number field is $\mathbb{F}_p$, the set of integers modulo p, where p is a prime number. The field $\mathbb{F}_p$ contains p elements, which can be taken as

$$\mathbb{F}_p = \{0,1,...,p-1\}; \tag{2.1}$$

$\mathbb{F}_p$ is of characteristic p: for any $x \in \mathbb{F}_p$, one has $px = 0$.

One can show that $\mathbb{F}_p^\times$, its multiplicative group, is isomorphic to the cyclic group

$$\{\omega,\omega^2,...,\omega^{p-1} = 1\}, \tag{2.2}$$

with ω a primitive $(p-1)$th root of 1.

The field of rational numbers, $\mathbb{Q}$, is significantly richer than $\mathbb{F}_p$. It contains an infinite number of elements and is of characteristic 0.

The "absolute value," which we will denote $|\ |_\infty$, defines a norm on the field $\mathbb{Q}$, i.e., it satisfies the properties

$$|x|_\infty = 0 \quad \text{iff} \quad x = 0, \tag{2.3a}$$

$$|xy|_\infty = |x|_\infty |y|_\infty, \tag{2.3b}$$

$$|x+y|_\infty \leqslant |x|_\infty + |y|_\infty. \tag{2.3c}$$

With the help of this norm, one can define a distance on $\mathbb{Q}$,

$$d(x,y) = |x-y|_\infty, \tag{2.4}$$

and start playing topological games. $\mathbb{Q}$ is a discrete field: it is, so to speak, full of holes. However, one can fill up these holes by "completing $\mathbb{Q}$ with respect to the norm $|\ |_\infty$": the resulting field is $\mathbb{R}$, the field of real numbers.

There are other norms that one can define on $\mathbb{Q}$. Indeed let p be an arbitrary prime number. Every rational number x can be written as

$$x = p^\alpha a/b, \tag{2.5}$$

where $\alpha,a,b \in \mathbb{Z}$ and p does not divide a nor b. The integer α is called the ordinal of x (at p). The p-adic norm $|x|_p$ of x is defined as follows:

$$|x|_p = p^{-\alpha} = p^{-\operatorname{ord} x}. \tag{2.6}$$

One checks that $|\ |_p$ is indeed a norm, i.e., Eqs. (2.3) are satisfied. In fact, a stronger form of Eq. (2.3c) is seen to hold, namely,

$$|x+y|_p \leqslant \max\{|x|_p,|y|_p\}. \tag{2.3c$'$}$$

Norms with the property in Eq. (2.3c) are called Archimedean and those with property in Eq. (2.3c') non-Archimedean or ultrametric.

With respect to the p-adic norm [Eq. (2.6)], $\mathbb{Q}$ is still discrete, of course, but, again, it can be completed: the resulting number fields are the $\mathbb{Q}_p$'s, the fields of p-adic numbers (one distinct field for every rational prime p).

Here $\mathbb{R}$ and $\mathbb{Q}_p$ are the only number fields that can be built in this way: one can prove (Ostrowski's Theorem[4]) that any norm on $\mathbb{Q}$ is equivalent to $|\ |_\infty$ or to one of the $|\ |_p$.

The abstract construction of $\mathbb{R}$ and $\mathbb{Q}_p$ from $\mathbb{Q}$ makes clear that ultrametricity, i.e., Eq. (2.3c') instead of Eq. (2.3c), is really at the heart of the quite different properties of real and p-adic numbers.

A very concrete and practical realization of $\mathbb{Q}_p$ is given by the set of power series,

$$\mathbb{Q}_p = \left\{ x = p^\alpha \left(\sum_{j=0}^{\infty} x_j p^j \right) \middle| 0 \leqslant x_j \leqslant p-1,\ x_0 \neq 0,\ \alpha \in \mathbb{Z} \right\}. \tag{2.7}$$

The x_j's are the "digits" of the p-adic number x and the integer α is the ordinal of x. All series in (2.7) are convergent with respect to the p-adic norm $|\ |_p$. Indeed, because of ultrametricity, a series Σa_n is convergent iff $|a_n|_p \to 0$. Let us, for example, rewrite $x = -1$ in this way: $-1 = (p-1) + (-1)\cdot p$. Iterating this equation, one gets

$$-1 = \sum_{j=0}^{\infty} (p-1)p^j. \tag{2.8}$$

We note that while $|\ |_\infty$, when extended from $\mathbb{Q}$ to $\mathbb{R}$, can take any (positive) value, $|\ |_p$ remains valued on a discrete set: the integral powers of p. An immediate consequence is that $\mathbb{Q}_p$ is only partially ordered (while $\mathbb{R}$ is totally ordered).

Subsets of $\mathbb{Q}_p$ that will often be used in the following are

$$\mathbb{Z}_p = \{x \in \mathbb{Q}_p |\ |x|_p \leqslant 1\}, \tag{2.9}$$

which is the ring of p-adic integers, and

$$\mathbb{Z}_p^\times = \{x \in \mathbb{Q}_p |\ |x|_p = 1\}, \tag{2.10}$$

which is the multiplicative group of p-adic "units."

Concerning the topological properties of $\mathbb{Q}_p$, the following remarks[4] will be sufficient for our purposes. Call

$$U(a;n) = \{x \in \mathbb{Q}_p |\ |x-a|_p \leqslant p^{-n}\} \tag{2.11}$$

a p-adic "ball" with center a. In this notation, $Z_p = U(0;0)$ while $U(0;n)$ is the set $p^n\mathbb{Z}_p = \{x \in \mathbb{Q}_p; |x|_p \leqslant p^{-n}\}$. Note that every point of $U(a;n)$ is a center, i.e., if $b \in U(a;n)$, then $U(a;n) = U(b;n)$. This is ultrametricity at work!

The p-adic balls are both open and closed sets,

$$U(a;n) = \{x \in Q_p |\ |x-a|_p < p^{-n+1}\}, \tag{2.12}$$

and they are disconnected sets, namely,

$$U(a;n) = \bigcup_{x=0}^{p-1} U(a + p^n x; n+1) \tag{2.13}$$

is the union of closed disjoint sets. Finally one can show that $U(a;n)$ is compact. Thus $\mathbb{Q}_p$ as well as $\mathbb{R}$ is a locally compact topological field. Note that $\mathbb{Z}$ is dense in $\mathbb{Z}_p$ as well as $\mathbb{Q}$ in $\mathbb{Q}_p$.

One can now start doing analysis on the $\mathbb{Q}_p$'s. We will make essential use of additive (χ) and multiplicative (π) characters,[3] which are continuous complex valued functions of a p-adic variable,

$$\chi,\pi\colon\ \mathbb{Q}_p \to \mathbb{C}, \tag{2.14}$$

with the property

$$\chi(x+y) = \chi(x)\chi(y) \tag{2.15}$$

and

$$\pi(xy) = \pi(x)\pi(y). \tag{2.16}$$

Furthermore unitary characters have the property

$$|\chi(x)| = 1, \quad |\pi(x)| = 1. \tag{2.17}$$

Characters will be extensively discussed in Sec. IV. Thus we give only a couple of examples here:

$$\pi(x) = |x|_p^s, \tag{2.18}$$

with s real or complex are multiplicative characters while

$$\chi(x) = \exp(2\pi i x) \tag{2.19}$$

is a unitary additive character. The precise meaning of Eq. (2.19) is the following: for x given by Eq. (2.7),

$$\chi(x) = \exp\left(2\pi i p^\alpha \left(\sum_{j=0}^{-1-\alpha} x_j p^j\right)\right). \tag{2.20}$$

In other words, for $x \in \mathbb{Z}_p$, $\chi(x) = 1$. It is important to realize that the "series expansion" of Eq. (2.19) makes no sense.

Integration over $\mathbb{Q}_p$ is discussed in Appendix A.

B. Quadratic extensions of Q_p

The algebraic structures of $\mathbb{R}$ and $\mathbb{Q}_p$ are quite different. The field $\mathbb{R}$ is clearly divided into two quadratic classes: a (nonzero) real number y is either the square of another real number or minus such a square. One writes

$$y = x^2 \quad \text{or} \quad y = -x^2. \tag{2.21}$$

Squaring a generic element x of $\mathbb{Q}_p$ as given by Eq. (2.7) leads to

$$x^2 = p^{2\alpha}(x_0^2 + 2x_0x_1p + \cdots). \tag{2.22}$$

Comparing this with the digital expansion of an arbitrary p-adic number y,

$$y = p^\beta(y_0 + y_1 p + \cdots), \tag{2.23}$$

shows that two necessary conditions for y to be a square are its ordinal is even ($\beta = 2\alpha$) and its first digit y_0 is a square modulo p. Hensel's lemma[4] guarantees that these necessary conditions are also sufficient for $p \neq 2$.

Since $y_0 \in \mathbb{F}_p^\times$ can be written as $\omega^k \pmod p$ with ω a primitive $(p-1)$th root of unity, half the elements of $\mathbb{F}_p^\times$ are squares (ω^k with k even) while the other half are not (k odd). This property is formalized by the Legendre symbol

$$(y_0/p)_L = (-1)^k, \quad \text{if } y_0 \equiv \omega^k \bmod p.$$

The Legendre symbol is a multiplicative character of order 2. To give a simple example,

$$\mathbb{F}_5^\times = \{1,2,3,4\} = \{2^4, 2, 2^3, 2^2\}$$

and

$$(1/5)_L = (4/5)_L = 1$$

while

$$(2/5)_L = (3/5)_L = -1.$$

Hence given an arbitrary p-adic number y, one and only one of the four numbers y, εy, py, or εpy with $(\varepsilon/p)_L = -1$, will be a square. Thus $\mathbb{Q}_p$ ($p \neq 2$) has four quadratic classes labeled by the parity of the ordinal and the quadratic class of the first digit. One writes

$$y = x^2 \quad \text{or} \quad y = \tau x^2, \quad \text{where } \tau = \varepsilon, p, \varepsilon p. \tag{2.24}$$

In the case $p = 2$, Eq. (2.22) reads $x^2 = p^{2\alpha}(1 + O(p^3))$, where the notation $O(p^k)$ means $|O(p^k)|_p \leqslant p^{-k}$. The parity of the ordinal and the value of the first two significant digits (x_1 and x_2) distinguish the eight quadratic classes:

$$y = x^2 \quad \text{or} \quad y = \tau x^2, \quad \text{where } \tau = -1, \pm 2, \pm 3, \pm 6. \tag{2.25}$$

From Eq. (2.21) it follows that $\mathbb{R}$ has a unique quadratic extension obtained by adding $\sqrt{-1}$, the solution of the equation $x^2 + 1 = 0$:

$$\mathbb{C} = \mathbb{R}(\sqrt{-1}) = \{z = x_1 + \sqrt{-1}\,x_2 | x_1, x_2 \in \mathbb{R}\}. \tag{2.26}$$

Once we have $\mathbb{C}$, we are done with extensions. In $\mathbb{C}$ every polynomial equation has a solution: the field $\mathbb{C}$ is algebraically closed. It is also complete with respect to $|\ |_\infty$, which is defined as follows:

$$|z|_\infty = |z\bar{z}|_\infty^{1/2} = |x_1^2 + x_2^2|_\infty^{1/2}, \tag{2.27}$$

where $\bar{z}$, the conjugate of $z = x_1 + \sqrt{-1}x_2$, is given by

$$\bar{z} = x_1 - \sqrt{-1}x_2.$$

Note also that any (nontrivial) extension of $\mathbb{R}$ is identical to $\mathbb{C}$.

In $\mathbb{Q}_p$ one proceeds along similar lines. From Eqs. (2.24) and (2.25), it follows that for $p\neq 2$, $\mathbb{Q}_p$ has *three* distinct quadratic extensions while for $p=2$, $\mathbb{Q}_2$ has *seven* of them. They can all be written as

$$\mathbb{Q}_p(\sqrt{\tau}) = \{z = x_1 + \sqrt{\tau}x_2 | x_1, x_2 \in \mathbb{Q}_p\}, \tag{2.28}$$

with $\tau = \varepsilon$, p, or εp, for $p\neq 2$, and $\tau = -1, \pm 2, \pm 3, \pm 6$, for $p=2$. The three distinct quadratic extensions of $\mathbb{Q}_5$, for example, are $\mathbb{Q}_5(\sqrt{2})$, $\mathbb{Q}_5(\sqrt{5})$, and $\mathbb{Q}_5(\sqrt{10})$.

One shows that two quadratic extensions $\mathbb{Q}_p(\sqrt{\tau})$ and $\mathbb{Q}_p(\sqrt{\tau'})$ are identical iff τ/τ' is a square in $\mathbb{Q}_p$. For example, $\mathbb{Q}_5(\sqrt{10})$ and $\mathbb{Q}_5(\sqrt{15})$ are identical fields.

In contrast with $\mathbb{C} = \mathbb{R}(\sqrt{-1})$, none of the extensions $\mathbb{Q}_p(\sqrt{\tau})$ are algebraically closed and, in fact, the extension process (i.e., adding to the field a "solution" of an algebraic equation) can be continued "indefinitely," e.g., with the equations

$$x^n = p \tag{2.29}$$

leading to $\mathbb{Q}_p(\sqrt[n]{p})$, etc.

The conjugate z of an element $z = x_1 + \sqrt{\tau}x_2$ of $\mathbb{Q}_p(\sqrt{\tau})$ is again defined as

$$\bar{z} = x_1 - \sqrt{\tau}x_2 \tag{2.30}$$

and the norm $|\ |_p$ is uniquely extended to $\mathbb{Q}_p(\sqrt{\tau})$ via the formula[4]

$$|z|_p = |z\bar{z}|_p^{1/2} = |x_1^2 - \tau x_2^2|_p^{1/2}. \tag{2.31}$$

The possible values of $|z|_p$ are thus integer or half-integer powers of p.

It is a theorem[2] that $\mathbb{R}$, $\mathbb{C}$, and finite extensions of $\mathbb{Q}_p$ exhaust all possibilities of Abelian locally compact number fields of characteristic 0. The fields $\mathbb{R}$ and $\mathbb{C}$ are connected fields while any finite extension of $\mathbb{Q}_p$ is not.

The quadratic field extensions $\mathbb{C}$ and $\mathbb{Q}_p(\sqrt{\tau})$ are two-dimensional vector spaces over their base field. Particularly important subsets of these extensions correspond to "circles," which we now proceed to describe.[3]

C. *p*-adic circles

The circle of square radius ρ in the complex plane is defined as

$$C_{-1}(\rho) = \{z\in\mathbb{C} | z\bar{z} = x_1^2 - (-1)x_2^2 = \rho\}. \tag{2.32}$$

In $\mathbb{Q}_p(\sqrt{\tau})$, one defines analogously

$$C_\tau(\rho) = \{z\in\mathbb{Q}_p(\sqrt{\tau}) | z\bar{z} = x_1^2 - \tau x_2^2 = \rho\}. \tag{2.33}$$

This equation allows one to define, on $\mathbb{Q}_p$, the functions $\operatorname{sgn}_\tau \rho$ associated with each quadratic extension $\mathbb{Q}_p(\sqrt{\tau})$:

$$\operatorname{sgn}_\tau \rho = \begin{cases} +1, & \text{if } \exists x_1, x_2 \in \mathbb{Q}_p\text{: } x_1^2 - \tau x_2^2 = \rho, \\ -1, & \text{otherwise.}\end{cases} \tag{2.34}$$

In other words, $C_\tau(\rho)$ is not empty iff $\operatorname{sgn}_\tau \rho = +1$, but, in contrast to Eq. (2.32), $\operatorname{sgn}_\tau \rho = +1$ does not necessarily imply that ρ is a square of an element of $\mathbb{Q}_p$.

An explicit calculation[15] leads to the following results: for $t = p^{\operatorname{ord} t}(t_0 + t_1 p + t_2 p^2 + \dots)$, $\tau = p^{\operatorname{ord}\tau}(\varepsilon_0 + \varepsilon_1 p + \varepsilon_2 p^2 + \dots)$, one finds, for $p\neq 2$,

$$\operatorname{sgn}_\tau t = \left(\frac{t_0}{p}\right)_{\mathrm{L}}^{\operatorname{ord}\tau} \cdot \left(\frac{\varepsilon_0}{p}\right)_{\mathrm{L}}^{\operatorname{ord} t} \cdot \left(\frac{-1}{p}\right)_{\mathrm{L}}^{\operatorname{ord}\tau\cdot\operatorname{ord} t}; \tag{2.35}$$

while, for $p=2$ ($\varepsilon_0 = t_0 = 1$),

$$\operatorname{sgn}_\tau t = (-1)^{\varepsilon_1 t_1} \cdot (-1)^{(t_1+t_2)\operatorname{ord}\tau + (\varepsilon_1+\varepsilon_2)\operatorname{ord} t}. \tag{2.36}$$

All these functions are multiplicative characters of order 2, i.e.,

$$\operatorname{sgn}_\tau(ut) = \operatorname{sgn}_\tau(u)\cdot\operatorname{sgn}_\tau(t),$$
$$\operatorname{sgn}_\tau(t^2) = 1. \tag{2.37}$$

A special role is played by the unit circle C_τ,

$$C_\tau = C_\tau(1) = \{z\in\mathbb{Q}_p(\sqrt{\tau}) | z\bar{z} = 1\}, \tag{2.38}$$

whose elements form a multiplicative group. Writing

$$z = c + \sqrt{\tau}s, \quad \text{with } c^2 - \tau s^2 = 1, \tag{2.39}$$

the group law $zz' = z''$ reads

$$c'' = cc' + \tau ss', \tag{2.40a}$$
$$s'' = sc' + cs'. \tag{2.40b}$$

The circle C_τ is most usefully described through the so-called "rational parametrization[3]":

$$c = (1+\tau t^2)/(1-\tau t^2), \quad s = 2t/(1-\tau t^2). \tag{2.41}$$

When $t = s/(c+1)$ runs over $\mathbb{Q}_p$, $z = c + \sqrt{\tau}s$ runs over the circle C_τ. In terms of t, the group law now becomes

$$t'' = (t + t'')/(1 + \tau tt'). \tag{2.42}$$

The detailed structure of the groups C_τ will be discussed in Sec. IV.

III. *P*-ADIC QUANTUM MECHANICS *à la* WEYL

In the following we will be mainly interested in one-dimensional systems described by one of the following quadratic Lagrangians:

$$L_1 = \tfrac{1}{2}m\dot{x}^2, \tag{3.1a}$$
$$L_2 = \tfrac{1}{2}m\dot{x}^2 - \tfrac{1}{2}kx^2, \tag{3.1b}$$
$$L_3 = \tfrac{1}{2}m\dot{x}^2 - Fx. \tag{3.1c}$$

We will mostly be concerned with the free particle and harmonic oscillator problems [Eqs. (3.1a) and (3.1b)]. The case of a particle in a constant field [Eq. (3.1c)] is discussed in Appendix B. For the harmonic oscillator, we will distinguish the compact ($k>0$) and noncompact ($k<0$) cases.

The classical equations of motion are easily solved and yield, the Eqs. (3.1a) and (3.1b),

$$\begin{pmatrix} x(t) \\ p(t) \end{pmatrix} = \begin{pmatrix} a(t) & b(t) \\ c(t) & d(t) \end{pmatrix} \begin{pmatrix} x(0) \\ p(0) \end{pmatrix} = M(t) \begin{pmatrix} x(0) \\ p(0) \end{pmatrix}. \quad (3.2)$$

By Liouville's theorem, $M(t)$ is "volume preserving" and is thus an element of the group $\mathrm{SL}(2,\mathbb{R})$, i.e., $\det M(t) = 1$.

More specifically, for the free particle,

$$M_f(t) = \begin{pmatrix} 1 & t/m \\ 0 & 1 \end{pmatrix}; \quad (3.3a)$$

for the compact harmonic oscillator ($k = m\omega^2$),

$$M_{\text{c.o.}}(t) = \begin{pmatrix} \cos \omega t & (1/m\omega)\sin \omega t \\ -m\omega \sin \omega t & \cos \omega t \end{pmatrix}; \quad (3.3b)$$

while in the noncompact case ($k = -m\omega^2$),

$$M_{\text{n.c.}}(t) = \begin{pmatrix} \cosh \omega t & (1/m\omega)\sinh \omega t \\ m\omega \sinh \omega t & \cosh \omega t \end{pmatrix}. \quad (3.3c)$$

For our purposes it is important to realize that the matrices given by Eqs. (3.3) belong to specific Abelian subgroups of $\mathrm{SL}(2,\mathbb{R})$.

In the Heisenberg picture of quantum mechanics, the classical variables $x(t)$ and $p(t)$ become operators $X(t)$ and $P(t)$ which act on a Hilbert space. For the quadratic systems considered here, the time development of $X(t)$ and $P(t)$ is still given by Eqs. (3.2) and (3.3). The commutation relation

$$[X(0),P(0)] = ih/2\pi \quad (3.4)$$

is imposed at some initial time $t = 0$ and the unimodularity of $M(t)$ guarantees that it will hold at any time.

An explicit realization on the Hilbert space of functions of a real variable is provided by

$$X(0)f(x) = xf(x), \quad (3.5a)$$

$$P(0)f(x) = -\frac{ih}{2\pi}\frac{d}{dx}f(x). \quad (3.5b)$$

One also introduces a time evolution operator $U(t)$ such that

$$U^{-1}(t)X(0)U(t) = X(t) = a(t)X(0) + b(t)P(0), \quad (3.6a)$$

$$U^{-1}(t)P(0)U(t) = P(t) = c(t)X(0) + d(t)P(0). \quad (3.6b)$$

This operator is easily shown to be

$$U(t) = \exp\{-(2\pi i/h)Ht\}, \quad (3.7)$$

where H is the Hamiltonian expressed in terms of $X(0)$ and $P(0)$. The proof relies on the possibility of expanding $U(t)$ in powers of t and uses the canonical commutation relation Eq. (3.4).

There are several obstacles to generalizing this quantization procedure to p-adic variables. To respect the laws of quantum mechanics one must remain in a Hilbert space of complex valued functions, but then a relation like Eq. (3.5a) becomes meaningless: x is p-adic while $f(x)$ is complex!

One could try $|x|_p f(x)$ or $|x|_p \operatorname{sgn}_\tau(x) f(x)$ as right-hand sides of Eq. (3.5a), but the operator $X(0)$ then seems much too "coarse" to admit a conjugate operator $P(0)$ satisfying Eq. (3.4)!

However, the rules of ordinary quantum mechanics can be expressed equally well in terms of *finite transformation* operators, namely, the Weyl operators[13]:

$$W_K(\alpha) = \exp\{(2\pi i/h)\alpha P(0)\}, \quad (3.8a)$$

$$W_X(\beta) = \exp\{(2\pi i/h)\beta X(0)\}. \quad (3.8b)$$

The commutation relation [Eq. (3.4)] is now expressed as

$$W_X(\beta)W_K(\alpha) = \exp\{-(2\pi i/h)\alpha\beta\}W_K(\alpha)W_X(\beta). \quad (3.9)$$

An explicit realization of Eqs. (3.8) is given, e.g., in configuration space, where

$$W_X(\beta)|x\rangle = \exp\{(2\pi i/h)\beta x\}|x\rangle, \quad (3.10)$$

$$W_K(\alpha)|x\rangle = |x-\alpha\rangle, \quad (3.11)$$

and Eqs. (3.6) of quantum evolution now read

$$U^{-1}(M)W_K(\alpha)U(M) = \exp\{-(2\pi i/2h)\alpha^2 cd\}W_K(\alpha d)W_X(\alpha c), \quad (3.12a)$$

$$U^{-1}(M)W_X(\beta)U(M) = \exp\{-(2\pi i/2h)\beta^2 ab\}W_K(\beta b)W_X(\beta a), \quad (3.12b)$$

with M the classical evolution matrix given in Eq. (3.2).

This "Weyl formulation" of quantum mechanics, where one represents the Heisenberg group rather than its algebra, provides an appropriate framework for a generalization to p-adic variables.[16] As already mentioned, p-adic quantum mechanics still lives in a Hilbert space of square integrable complex functions of a p-adic variable:

$$\mathscr{H} = \{\varphi(x)\,|\,\langle\varphi|\varphi\rangle < \infty\}, \quad (3.13)$$

with the scalar product

$$\langle\varphi|\psi\rangle = \int_{\mathbb{Q}_p} dx\, \varphi^*(x)\psi(x). \quad (3.14)$$

The details of p-adic integration are discussed in Appendix A.

As usual, in order to handle continuous spectra, one must also introduce a space S of well-behaved test functions (locally constant and with compact support) and a space S' of functionals on S (the distributions). This can be done straightforwardly as shown in Ref. 3.

One can thus use the standard (improper) "configuration space" basis $|x\rangle$, which is now indexed by elements x of $\mathbb{Q}_p$:

$$\langle x|x'\rangle = \delta(x-x'), \quad (3.15)$$

$$\mathbb{1} = \int_{\mathbb{Q}_p} dx|x\rangle\langle x|. \quad (3.16)$$

Equations (3.10) and (3.11) thus provide an explicit realization of the Weyl operators even when x, α, and β are p-adic: the complex phase $\exp(2\pi i\beta x/h)$ is simply reinterpreted as a unitary additive character on the field $\mathbb{Q}_p$ and its precise meaning is given by Eq. (2.20).

For the simple systems considered in this paper, one can go one step further and take Eqs. (3.12) as dynamical equations even when the variables are p-adic: the phase factor is again reinterpreted as a character while the entries a, b, c, and d, of the "evolution" matrix M are simply taken as p-adic numbers. In other words, we may define the p-adic free particle and compact or noncompact harmonic oscillator by Eqs. (3.12), where M is now an element of an Abelian subgroup of $\mathrm{SL}(2,Q_p)$, which is the "analog" of the real Abelian subgroups of $\mathrm{SL}(2,R)$ given by Eqs. (3.3). These subgroups will be discussed in Sec. IV.

From Eqs. (3.10)–(3.12) one can explicitly determine the matrix elements of U up to an overall factor, whether the variables are real of p-adic:

$$\begin{aligned}\langle y|U|x\rangle &= \langle 0|W_K(y)U|x\rangle \\ &= \chi(-y^2cd/2h)\langle 0|UW_K(yd)W_K(yc)|x\rangle \\ &= \chi(-y^2cd/2h)\chi(ycx/h)\langle 0|UW_K(yd-x)|0\rangle. \end{aligned} \tag{3.17}$$

If $b\neq 0$, Eq. (3.12b) with $\beta=(yd-x)/b$ gives at once

$$\begin{aligned}&\langle 0|W_X((yd-x)/b)U|0\rangle \\ &\quad=\langle 0|U|0\rangle \\ &\quad=\chi\Big(-\frac{a}{2h}\frac{(yd-x)^2}{b}\Big)\langle 0|UW_K(yd-x)|0\rangle;\end{aligned}$$

hence

$$\begin{aligned}\langle y|U|x\rangle &= \chi\Big(-\frac{y^2cd}{2h}\Big)\chi\Big(\frac{ycx}{h}\Big)\chi\Big(\frac{a}{2h}\frac{(yd-x)^2}{b}\Big)\langle 0|U|0\rangle \\ &= \chi\Big(-\frac{y^2cd}{2h}+\frac{ycx}{h}+\frac{a}{2h}\frac{(yd-x)^2}{b}\Big)\langle 0|U|0\rangle \\ &= \chi\Big(\frac{1}{2hb}(ax^2+dy^2-2xy)\Big)\,v(M). \end{aligned} \tag{3.18}$$

If $b=0$, Eq. (3.12b) reads

$$W_X(\beta)U=UW_X(\beta a).$$

Hence

$$\begin{aligned}\langle y|U|x\rangle &= \langle y|W_X(-\beta)UW_X(\beta a)|x\rangle \\ &= \chi((\beta/h)(ax-y))\langle y|U|x\rangle,\end{aligned}$$

which shows that the matrix element vanishes unless $y=ax$. Equation (3.17) with $ad=1$ also gives

$$\begin{aligned}\langle y|U|x\rangle &= \chi(-a^2x^2cd/2h+acx^2/h) \\ &\quad\times\langle 0|UW_K(x(ad-1))|0\rangle\delta(y-ax) \\ &= \chi((ac/2h)x^2)v(M)\delta(y-ax).\end{aligned} \tag{3.19}$$

We have gone through the detailed derivation of these simple formulas in order to show explicitly that all manipultions remain meaningful in the p-adic case. Note that we do recover the Freund–Olson ansatz (see Ref. 7): for the Lagrangians given in Eqs. (3.1), matrix elements of U are proportional to unitary additive characters with the classical action as argument.

The modulus of the overall factor $v(M)$ is fixed by unitarity:

$$\int dz\langle z|U(M)|x\rangle^*\langle z|U(M)|y\rangle=\delta(x-y). \tag{3.20}$$

One finds

$$\begin{aligned}&\langle y|U(M)|x\rangle \\ &\quad=f(M)|hb|_p^{-1/2}\chi((1/2hb)(ax^2+dy^2-2xy)), \\ &\quad\quad b\neq 0,\end{aligned} \tag{3.21a}$$

$$\begin{aligned}&\langle y|U(M)|x\rangle \\ &\quad=f(M)|a|_p^{1/2}\chi(acx^2/2h)\delta(y-ax), \\ &\quad\quad b=0.\end{aligned} \tag{3.21b}$$

With the phase factor $f(M)$ these formulas define a projective representation of the full $\mathrm{SL}(2,Q_p)$ group. We show in Secs. V and VI that for the quadratic systems we consider one can choose $f(M)$ so as to obtain a true representation of the relevant Abelian subgroup of $\mathrm{SL}(2,Q_p)$.

To conclude this section let us emphasize once again that, in the formulation of quantum mechanics over p-adic fields adopted here, there is no use nor need to define p-adic "momentum" or "Hamiltonian" operators. In the real case they are infinitesimal generators of space and time translations, but, since Q_p is a disconnected field, these infinitesimal transformation operators become meaningless. On the other hand, finite transformations remain meaningful and the corresponding Weyl and evolution operators are p-adically well defined.

We can now proceed with a more precise identification of what we mean by p-adic free particle and harmonic oscillators.

IV. ABELIAN SUBGROUPS OF SL(2,$\mathbb{Q}_p$) AND THEIR CHARACTERS

For an arbitrary $\mathrm{SL}(2,\mathbb{Q}_p)$ matrix, the secular equation has either a degenerate solution, two distinct solutions, or no solution at all in the field $\mathbb{Q}_p$. In the latter case it admits two conjugate solutions in one of the quadratic extensions $\mathbb{Q}_p(\sqrt{\tau})$. By anaolgy with $\mathrm{SL}(2,\mathbb{R})$ we will call the corresponding elements of $\mathrm{SL}(2,\mathbb{Q}_p)$, parabolic, hyperbolic, and τ elliptic, respectively.

We now define the quantum mechanical p-adic free particle by the evolution operator $U(M)$ [Eqs. (3.21)], where M belongs to the following parabolic Abelian subgroup of $\mathrm{SL}(2,\mathbb{Q}_p)$:

$$G_1=\left\{M=\begin{pmatrix}1 & t/m\\ 0 & 1\end{pmatrix}\ ;\ t,m\in\mathbb{Q}_p\right\}. \tag{4.1}$$

Multiplication of elements of G_1 corresponds to addition for the parameter t: G_1 is thus isomorphic to the additive group $\mathbb{Q}_p^+$ and hence noncompact. Note, also, that G_1 is conjugate to the subgroup

$$\widetilde{G}_1=S^{-1}G_1S=\left\{\widetilde{M}=\begin{pmatrix}1 & 0\\ -t/m & 1\end{pmatrix}\right\}, \tag{4.2}$$

with

$$S=\begin{pmatrix}0 & 1\\ -1 & 0\end{pmatrix}.$$

Similarly the noncompact p-adic harmonic oscillator (for simplicity, we have taken $m\omega=2h=1$) is defined through matrices M that belong to the following hyperbolic subgroup of $\mathrm{SL}(2,\mathbb{Q}_p)$:

$$G_2=\left\{M=\begin{pmatrix} c & s\\ s & c\end{pmatrix},\quad c,s\in\mathbb{Q}_p,\ c^2-s^2=1\right\}. \qquad (4.3)$$

The eigenvalues of a typical element of G_2 are $c+s$ and $c-s$. Since $c^2-s^2=1$, they are each other's inverse; hence the group G_2 is conjugate to

$$\widetilde{G}_2=T^{-1}G_2T=\left\{\widetilde{M}=\begin{pmatrix} a & 0\\ 0 & a^{-1}\end{pmatrix};\ a\in\mathbb{Q}_p^\times\right\}, \qquad (4.4)$$

with

$$T=\begin{pmatrix} 1 & -\frac{1}{2}\\ 1 & \frac{1}{2}\end{pmatrix}.$$

From Eq. (4.4) it is obvious that G_2 is isomorphic to the multiplicative group $\mathbb{Q}_p^\times$ and thus noncompact.

Finally we define the various (three for $p\neq 2$ and seven for $p=2$) compact p-adic harmonic oscillators (again, $m\omega=2h=1$) via matrices M that belong to the following τ-elliptic subgroups of $\mathrm{SL}(2,\mathbb{Q}_p)$:

$$G_\tau=\left\{M=\begin{pmatrix} c & s\\ \tau s & c\end{pmatrix},\ c,s\in\mathbb{Q}_p,\right.$$
$$\left. c^2-\tau s^2=1;\ \tau \text{ not a square}\right\}. \qquad (4.5)$$

The eigenvalues of an element of G_τ are $c\pm\sqrt{\tau}s$, which belong to the unit circle C_τ [Eq. (2.38)] in the quadratic extension $\mathbb{Q}_p(\sqrt{\tau})$. Clearly G_τ and C_τ are isomorphic.

Our next problem is to compute the *spectrum* of the quantum p-adic free particle and harmonic oscillators, which are now completely defined.

From the eigenvalue equation

$$U(M)|\psi_\lambda\rangle=\lambda(M)|\psi_\lambda\rangle, \qquad (4.6)$$

one deduces that

$$\lambda(M_1)\lambda(M_2)=\lambda(M_1M_2) \qquad (4.7)$$

and

$$|\lambda(M)|_\infty=1. \qquad (4.8)$$

The eigenvalues of the quantum evolution operators are thus *unitary characters* of the groups G_1, G_2, and G_τ or equivalently of $\mathbb{Q}_p^+$, $\mathbb{Q}_p^\times$, and C_τ. In the remainder of this section and in Appendix C we give a complete description of all these characters. These results will then be used in the following sections to determine eigenvalues and eigenfunctions explicitly.

Let us recall that characters of a topological group G themselves form a group $\widehat{G}$, the dual of G. If G is compact, which is the cases for C_τ, $\widehat{G}$ is discrete and vice versa; if G is a direct product, so is $\widehat{G}$. When dealing with characters, cyclic groups are particularly easy to handle: specifying the value of a character for a generator determines its value for any element. Thus if one succeeds in decomposing a group into a direct product of (finite or infinite) cyclic groups, its dual group is easy to describe. This is the strategy used below for $\mathbb{Q}_p^\times$ and C_τ.

A. The additive group $\mathbb{Q}_p^+\approx G_1$

Every character is of the form[3]

$$\chi_\sigma(x)=\chi(\sigma x)=\exp\{2\pi i\sigma x\}, \qquad (4.9)$$

with σ an arbitrary p-adic number. Thus $\widehat{G}_1$ is isomorphic to G_1. The completeness and orthogonality properties of the characters are expressed as

$$\int_{\widehat{\mathbb{Q}}_p} d\sigma\,\chi_\sigma(x)\chi_\sigma^*(x')=\delta(x-x'), \qquad (4.10a)$$

$$\int_{\mathbb{Q}_p} dx\,\chi_\sigma(x)\chi_{\sigma'}^*(x)=\delta(\sigma-\sigma'). \qquad (4.10b)$$

Precisely as in the case of the real additive group, the characters Eq. (4.9) are the "Fourier modes" that can be used to Fourier transform complex functions of a p-adic variable.[3]

B. The multiplicative group $\mathbb{Q}_p^\times\approx G_2$

Any element of $\mathbb{Q}_p^\times$ can be written as

$$x=p^ky, \qquad (4.11)$$

with $k=\mathrm{ord}\,x$ a rational integer and y a p-adic unit, i.e., an element of $\mathbb{Z}_p^\times$. For $p\geqslant 3$, a p-adic unit can be uniquely decomposed as

$$y=\omega^ju,\quad j=1,\ldots,p-1, \qquad (4.12)$$

where u is a p-adic unit whose first digit is 1, i.e.,

$$u\in O_1=\{1+O(p)\}=1+p\mathbb{Z}_p, \qquad (4.13)$$

and ω is a primitive $(p-1)$th root of 1 in $\mathbb{Z}_p^\times$. Such roots exist in $\mathbb{F}_p^\times$, as we have seen in Eq. (2.2), and hence exist in $\mathbb{Z}_p^\times$ by Hensel's lemma.[4]

The (multiplicative) group O_1 has subgroups O_m, with

$$O_m=\{1+O(p^m)\}=1+p^m\mathbb{Z}_p \qquad (4.14)$$

and

$$O_1\supset O_2\supset\cdots\supset O_m\supset\cdots\ . \qquad (4.15)$$

For increasing m, the subgroups O_m thus contain elements of O_1 that are closer and closer to 1.

From Eqs. (4.11) and (4.12) it follows that

$$\mathbb{Q}_p^\times\approx\mathbb{Z}\otimes\mathbb{Z}_p^\times, \qquad (4.16)$$

and, for $p\geqslant 3$,

$$\mathbb{Z}_p^\times\approx Z_{p-1}\otimes O_1. \qquad (4.17)$$

The dual group is thus also a direct product. Following the notation of Ref. 3 we will write π for a general character on $\mathbb{Q}_p^\times$ and θ for a character on the compact subgroup $\mathbb{Z}_p^\times$:

$$\pi(p^ky)=\exp\{2\pi ik\rho\}\theta(y), \qquad (4.18)$$

with $\rho\in[0,1]$.

For every θ character, $\theta(1)=1$. Hence by continuity every character on $\mathbb{Z}_p^\times$ eventually becomes trivial on some subgroup O_m. This leads to a classification of θ characters according to their rank.[3]

(i) The trivial character $\theta(y)=1$ on $\mathbb{Z}_p^\times$ is of rank 0.

(ii) Characters of rank 1 are trivial on O_1 but not on $\mathbb{Z}_p^\times$, i.e., not on Z_{p-1}.

They read

$$\theta(\omega^j u) = \theta(\omega^j) = \chi\left(\frac{jl}{p-1}\right) = \exp\left\{2\pi i \frac{jl}{p-1}\right\}, \quad (4.19)$$

with $1 \leqslant l \leqslant p-2$. There are thus $(p-2)$ θ characters of rank 1. For example, the Legendre symbol is the rank 1 character with $l = (p-1)/2$.

(iii) Characters of rank m ($\geqslant 2$) are trivial on O_m but not on O_{m-1}. For an element y of $\mathbb{Z}_p^\times$,

$$\begin{aligned} y &= \omega^j(1 + u_1 p + \cdots + u_{m-1}p^{m-1} + u_m p^m + \cdots) \\ &= \omega^j(1 + u_1 p + \cdots + u_{m-1}p^{m-1})h, \end{aligned} \quad (4.20)$$

where $h \in O_m$, the value of the character can only depend on j and on the digits $u_1,\ldots,u_{m-1}$, i.e.,

$$\theta(\omega^j u) = \theta(\omega^j(1 + u_1 p + \cdots + u_{m-1}p^{m-1})). \quad (4.21)$$

In other words, a θ character of rank m is a character on $Z_{p-1} \otimes O_1/O_m$, where

$$O_1/O_m = \{1 + u_1 p + \cdots u_{m-1}p^{m-1},\ 0 \leqslant u_i \leqslant p-1\}. \quad (4.22)$$

This finite group is of order p^{m-1}. It is cyclic and any element g with $|g-1|_p = p^{-1}$ can be taken as a generator (see Appendix C).

In summary, a θ character of rank $m \geqslant 2$ reads

$$\begin{aligned} \theta(\omega^j u) &= \theta(\omega^j g^n) \\ &= \chi\left(\frac{jl}{p-1}\right)\chi\left(\frac{an}{p^{m-1}}\right), \\ & 1 \leqslant l \leqslant p-1, \quad 1 \leqslant a \leqslant p^{m-1}, \quad p \nmid a. \end{aligned} \quad (4.23)$$

There are thus $(p-1)^2 p^{m-2}$ such characters.

The set of all π characters defines a basis on the Hilbert space of square integrable functions on the group $\mathbb{Q}_p^\times$ with invariant measure $dx/|x|_p$. Their completeness and orthogonality properties are expressed as

$$\int d\pi\, \pi(x)\pi^*(x') = \delta\left(1 - \frac{x}{x'}\right) = |x|_p \delta(x - x'), \quad (4.24)$$

where

$$\int d\pi = (1 - p^{-1})^{-1} \int_0^1 d\rho \sum_{\theta \text{ characters}}, \quad (4.25)$$

$$\int_{\mathbb{Q}_p^\times} \frac{dx}{|x|_p} \pi(x)\pi'^*(x) = \delta(\pi - \pi'). \quad (4.26)$$

The case $p = 2$ is discussed in Appendix C.

C. The elliptic groups C_τ

The analysis of the C_τ groups is in many respects similar to what was done for the multiplicative group $\mathbb{Q}_p^\times$. One factors C_τ as a direct product of a finite cyclic group and an infinite group C_τ^1. The former is the set of all elements of finite order and the latter, in analogy with Eqs. (4.14) and (4.15), contains a sequence of subgroups C_τ^m, whose elements get closer and closer to the unit element.

We will give here the details of this analysis for the two cases $p \geqslant 5$, $\tau = p$ or εp and $p \geqslant 3$, $\tau = \varepsilon$. The other cases are treated in Appendix C.

When $p \geqslant 5$, $\tau = p$ or εp, from $c^2 - \tau s^2 = 1$ it follows that $|c^2|_p = 1$, $|\tau s^2|_p < 1$; hence $|1 - c^2| < 1$. Therefore either $|1 - c|_p < 1$ or $|1 + c|_p < 1$. In terms of the parameter t [Eq. (2.41)], one has

$$\text{when } |t|_p \leqslant 1, \quad (c,s) = (1 + O(p), O(1)); \quad (4.27)$$

while

$$\begin{aligned} \text{for } |t|_p \geqslant p, \quad (c,s) &= (-1 + O(p), O(1)) \\ &= (-1,0)\cdot(1 + O(p), O(1)). \end{aligned} \quad (4.28)$$

Thus

$$C_\tau = Z_2 \otimes C_\tau^1, \quad (4.29)$$

with

$$Z_2 = \{(c,s) = (\pm 1, 0)\}, \quad (4.30)$$

and C_τ^1 is the subgroup of elements of C_τ for which $|1 - c|_p < 1$. We will say that C_τ splits into two "sheets."

From the decreasing sequence of subgroups

$$C_\tau \equiv C_\tau^0 \supset C_\tau^1 \supset C_\tau^2 \supset \cdots \supset C_\tau^m \supset \cdots, \quad (4.31)$$

where

$$C_\tau^m = \{(c,s) = (1 + O(p^{2m-1}), O(p^{m-1}))\}, \quad m \geqslant 1, \quad (4.32)$$

one defines, as before, the rank m of a character $\eta(z)$ on the circle $z\bar{z} = 1$. A character of rank m is trivial on C_τ^m but not on C_τ^{m-1}. The determination of the characters proceeds as in the case of the multiplicative group $\mathbb{Q}_p^\times$. Obviously there is one (trivial) character of rank 0, $\eta(z) = 1$, and one character of rank 1,

$$\eta(z = (-)^j h) = \eta((-1)^j) = (-)^j, \quad (4.33a)$$

with $h \in C_\tau^1$.

For $m \geqslant 2$, any element of C_τ^1 can be factorized as follows:

$$\begin{aligned} z = [&(1 + c_1 p + \cdots c_{m-1}p^{m-1}) \\ &+ \sqrt{\tau}(s_0 + s_1 p + \cdots + s_{m-2}p^{m-2})] \\ &\cdot [1 + O(p^{2m-1}) + \sqrt{\tau}O(p^{m-1})], \end{aligned}$$

where the second factor belongs to C_τ^m. Hence an η character of rank $m \geqslant 2$ only depends on the $(m-1)$ first significant digits of c and s. They can thus be viewed as characters on the finite group $Z_2 \otimes C_\tau^1/C_\tau^m$. Once again we show in Appendix C that C_τ^1/C_τ^m is a cyclic group of order p^{m-1} and that any element $g = c + \sqrt{\tau}s$ with $|s|_p = 1$ and $|1 - c|_p = p^{-1}$ can be taken as a generator.

There are thus $2(p-1)\,p^{m-2}$ characters of rank $m \geqslant 2$ given by ($h \in C_\tau^m$)

$$\eta(z = (-)^j g^n h) = \eta((-1)^j g^n) = (-)^{lj}\chi(bn/p^{m-1}), \quad (4.33b)$$

$$l = 0,1, \quad 1 \leqslant b \leqslant p^{m-1}, \quad p \nmid b.$$

The case $p \geqslant 3$, $\tau = \varepsilon$ is analyzed in a similar fashion. Since $|\varepsilon|_p = 1$, the equation $c^2 - \varepsilon s^2 = 1$ now implies $|c|_p \leqslant 1$, $|s|_p \leqslant 1$. Writing

$$\begin{aligned} c &= c_0 + c_1 p + c_2 p^2 + \cdots, \\ s &= s_0 + s_1 p + s_2 p^2 + \cdots, \end{aligned}$$

one has $c_0^2 - \varepsilon s_0^2 = 1 \pmod p$ and this equation admits $p+1$ solutions, $(c_0,s_0) = (\pm 1,0)$, and $p-1$ solutions with $s_0 \neq 0$. These $p+1$ solutions can be shown to form a cyclic group mod p. One can determine the higher digits of c and s so as to obtain a finite cyclic subgroup of C_ε. Let ω be a generator of this subgroup. Any element of C_ε can thus be written uinquely as

$$(c,s) = \omega^{j}\cdot(\tilde{c},\tilde{s}), \quad 1 \leqslant j \leqslant p+1, \tag{4.34}$$

with $(\tilde{c},\tilde{s}) = (1+O(p^2), O(p))$. In other words, C_ε factorizes into the product of a cyclic group Z_{p+1} times a group of elements close to the identity

$$C_\varepsilon = Z_{p+1} \otimes C_\varepsilon^1. \tag{4.35}$$

We show in Appendix C that this latter group is isomorphic to the additive group $p\mathbb{Z}_p$.

The $p+1$ "sheets" of C_ε can be characterized in terms of the rational parameter [Eq. (2.41)]:

$$|t|_p < 1, \quad (c,s) = (1+O(p^2), O(p));$$
$$|t|_p > 1, \quad (c,s) = (-1+O(p^2), O(p));$$
$$|t|_p = 1, \quad p-1 \text{ sheets with } |c|_p \leqslant 1, \ |s|_p = 1$$

corresponding to the $p-1$ possible values of t_0 ($t_0 \neq 0$) in $t = t_0 + t_1 p + t_2 p^2 + \cdots$.

Once again one has

$$C_\varepsilon = C_\varepsilon^0 \supset C_\varepsilon^1 \supset C_\varepsilon^2 \supset \cdots \supset C_\varepsilon^m \supset \cdots, \tag{4.36}$$

where

$$C_\varepsilon^m = \{(c,s) = (1+O(p^{2m}), O(p^m))\}, \quad m \geqslant 1. \tag{4.37}$$

As before, an η character of rank m is trivial on C_ε^m: it is a character of the finite group $Z_{p+1} \otimes C_\varepsilon^1 / C_\varepsilon^m$, which is now of order $(p+1)p^{m-1}$. (Note that elements of C_ε^1 are now taken mod p^m.)

In Appendix C we show that $C_\varepsilon^1/C_\varepsilon^m$ is cyclic. Let g be a generator. Then with $h \in C_\varepsilon^m$ and $m \geqslant 2$,

$$\eta(z = \omega^j g^n h) = \chi\left(\frac{lj}{p+1}\right)\chi\left(\frac{bn}{p^{m-1}}\right),$$
$$1 \leqslant l \leqslant p+1, \quad 1 \leqslant b \leqslant p^{m-1}, \quad p \nmid b, \tag{4.38}$$

while for $m=1$, the values of l are restricted to $1 \leqslant l \leqslant p$. There are thus, respectively, one character of rank 0, p characters of rank 1, and $(p^2-1)\,p^{m-2}$ characters of rank m.

For *all* circles, the invariant measure on C_τ is given by

$$d\mu = b_\tau \frac{dt}{|1-\tau t^2|_p}. \tag{4.39}$$

The constant b_τ is fixed by the normalization

$$\int_{C_\tau} d\mu = 1$$

and a simple calculation yields

$$b_\tau = \begin{cases} (1+p^{-1})^{-1}, & \text{for } p \neq 2, \quad \tau = \varepsilon, \\ \frac{1}{3}, & \text{for } p = 2, \quad \tau = -3, \\ \frac{1}{2}, & \text{otherwise.} \end{cases}$$

The discrete set of η characters gives a basis for the Hilbert space of square integrable functions on the circle C_τ.

Completeness and orthogonality properties read

$$\sum_\eta \eta(z)\eta^*(z') = \delta\left(1 - \frac{z}{z'}\right) = \delta(z-z'), \tag{4.41}$$

$$\int_{C_\tau} d\mu\, \eta(z)\eta^{*\prime}(z) = \delta_{\eta,\eta'}. \tag{4.42}$$

V. SPECTRUM OF THE NONCOMPACT EVOLUTION GROUPS

In Sec. IV we defined the quantum mechanical p-adic free particle and noncompact oscillator by their evolution group G_1 and G_2 [see Eqs. (4.1) and (4.3)]. We now proceed to give the complete solution of these problems; namely, we obtain the full spectral decomposition of the corresponding evolution operators.

A glance at Eqs. (3.21) shows that the spectral decomposition of $U(M)$ is almost trivial when M is of the form $\left(\begin{smallmatrix} a & 0 \\ c & d \end{smallmatrix}\right)$. Since matrices belonging to the conjugated subgroups $\widetilde{G}_1$ and $\widetilde{G}_2$ [Eqs. (4.2) and (4.4)] are precisely of this form we will first determine the spectrum of these groups and then "conjugate back" to G_1 and G_2.

A. Free particle

For an element

$$\widetilde{M} = \begin{pmatrix} 1 & 0 \\ -t/m & 1 \end{pmatrix}$$

of $\widetilde{G}_1$, we write $\widetilde{U}_t$ for $U(\widetilde{M})$ and Eq. (3.21b) then reads

$$\langle y|\widetilde{U}_t|x\rangle = f(\widetilde{M})\chi(-tx^2/2hm)\delta(x-y). \tag{5.1}$$

The phase factor $f(\widetilde{M})$ can consistently be taken as 1. With this choice, $\widetilde{U}_t$ defines a true representation of $\widetilde{G}_1$. Eigenvectors and eigenvalues of $\widetilde{U}_t$ are then given by

$$\varphi_k(x) = \delta(x-k), \tag{5.2}$$

$$\lambda_k(\widetilde{M}) = \chi(-tk^2/2mh), \tag{5.3}$$

since

$$(\widetilde{U}\varphi_k)(y) = \int dx \langle y|\widetilde{U}_t|x\rangle\langle x|\varphi_k\rangle$$
$$= \lambda(-tk^2/2mh)\varphi_k(y).$$

To go back to the group G_1 we still need the representation $U(S)$ of the conjugation matrix $S\left(\begin{smallmatrix} 0 & 1 \\ -1 & 0 \end{smallmatrix}\right)$.

From Eq. (3.21a), one may set

$$\langle y|U(S)|z\rangle = |h|_p^{-1/2}\chi(-yz/h). \tag{5.4}$$

Note that it is simply the kernel of the Fourier transform.

For the quantum evolution operator $U_t = U(M)$ corresponding to the element

$$M = \begin{pmatrix} 1 & t/m \\ 0 & 1 \end{pmatrix}$$

of G_1, we already know from Eq. (3.21a) that

$$\langle y|U_t|x\rangle = f(M)|ht/m|_p^{-1/2}\chi((m/2ht)(x-y)^2). \tag{5.5a}$$

Using

$$U_t = U(S)\widetilde{U}_t U(S^{-1})$$

and the results of Appendix A completely determines $f(M)$ (see Ref. 8):

$$f(M) = \tau_p^{-1}(2mht). \tag{5.5b}$$

With this phase factor, U_t is a true representation of G_1.

The eigenfunctions $\psi_k(x)$ of U_t are given by

$$\psi_k(x) = \int dy \langle x|U(S)|y\rangle\langle y|\varphi_k\rangle$$
$$= |h|_p^{-1/2}\chi(-xk/h). \tag{5.6}$$

These "plane waves" correspond to the doubly degenerate eigenvalue

$$\lambda_k(M) = \lambda_k(\widetilde{M}) = \chi(-tk^2/2mh) \equiv \chi_\sigma(t), \tag{5.7}$$

with $\sigma = -k^2/2mh$. Once h and m are given, the *quadratic class* of σ is fixed: for $p \neq 2$, one-fourth of all possible characters of $\mathbb{Q}_p^+$ appear as eigenvalues of U_t, while, for $\mathbb{Q}_2^+$, only one-eighth of them do. This is analogous to the real case where one-half of the characters of $\mathbb{R}^+$ are eigenvalues of the evolution operator. Orthogonality and completeness of the wave functions follow immediately from Eq. (4.10) and the spectral decomposition of U_t reads

$$\langle y|U_t|x\rangle = |h|_p^{-1}\int dk\,\chi\Big(\frac{xk}{h}\Big)\chi\Big(\frac{-yk}{h}\Big)\chi\Big(\frac{-tk^2}{2mh}\Big). \tag{5.8}$$

The problem of a particle in a constant field can be treated in a similar way. Details are given in Appendix B.

B. Noncompact oscillator

Essentially, the previous analysis can be repeated step by step.

For an element

$$\widetilde{M} = \begin{pmatrix} a & 0 \\ 0 & a^{-1}\end{pmatrix}$$

of $\widetilde{G}_2$, the propagator $\widetilde{U}_a$ reads

$$\langle y|\widetilde{U}_a|x\rangle = |a|_p^{1/2}\delta(y-ax), \tag{5.9}$$

where, again, the phase factor $f(\widetilde{M})$ can be taken equal to 1: Eq. (5.9) then defines a true representation of $\widetilde{G}_2$.

For π, an arbitrary multiplicative character,

$$\varphi_\pi(x) = |x|_p^{-1/2}\pi^{-1}(x) \tag{5.10}$$

and

$$\lambda_\pi(\widetilde{M}) = \pi(a) \tag{5.11}$$

are the eigenfunctions and eigenvalues of $\widetilde{U}_a$. The conjugation kernel $U(T)$ of the matrix $T = \left(\begin{smallmatrix} 1 & -1/2 \\ 1 & 1/2\end{smallmatrix}\right)$ can be taken as

$$\langle y|U(T)|x\rangle = |2|_p\chi(-2(x^2+\tfrac{1}{2}y^2-2xy)). \tag{5.12}$$

For an element $M = \left(\begin{smallmatrix} c & s \\ s & c\end{smallmatrix}\right)$ of G_2, the quantum evolution operator

$$U(M) = U(T)\widetilde{U}_{c+s}U(T^{-1}) \equiv U(c,s) \tag{5.13}$$

finally becomes[8]

$$\langle y|U(c,s)|x\rangle = |s/2|_p^{-1/2}\tau_p^{-1}(s(s+c))$$
$$\times\chi([c(x^2+y^2)-2xy]/s). \tag{5.14}$$

The eigenfunction of $U(c,s)$ for the eigenvalue $\lambda_\pi(M) = \pi(c+s)$ is

$$\psi_\pi(x) = \int dy\langle x|U(T)|y\rangle\langle y|\varphi_\pi\rangle$$
$$= |2|_p\chi(-x^2)\int\frac{dy}{|y|_p^{1/2}}\pi^{-1}(y)\chi(-2y^2+4xy). \tag{5.15}$$

It follows from Eqs. (4.24)–(4.26) that they are orthogonal and complete.

One can check that each unitary multiplicative character π appears once and only once in the spectrum of $U(c,s)$.

For multiplicative characters π that do not involve a θ character [see Eq. (4.18)], the corresponding eigenfunctions are easily calculated with the help of Appendix A. For $\pi(x) = 1$ and $p \neq 2$, for example, one finds

$$\psi_{\pi\equiv 1}(x) = \begin{cases} 1+p^{-1/2}, & \text{for } |x|_p \leqslant 1, \quad (5.16a) \\ |x|_p^{-1/2}\{\chi(x^2)+\chi(-x^2)\}, & \text{for } |x|_p \geqslant p. \quad (5.16b)\end{cases}$$

When nontrivial θ characters are present, the task of evaluating Eq. (5.15) is considerably harder since one must eventually compute nontrivial Gauss' sums. We will not pursue the problem here.

VI. SPECTRUM AND WAVE FUNCTIONS FOR THE COMPACT ELLIPTIC GROUPS

We have seen in Sec. IV that compact p-adic harmonic oscillators are defined by τ-elliptic subgroups of $SL(2,\mathbb{Q}_p)$. These cannot be diagonalized within $SL(2,\mathbb{Q}_p)$; hence the conjugation trick used in the noncompact cases is not applicable anymore.

For a matrix $M = \left(\begin{smallmatrix} c & s \\ \tau s & c\end{smallmatrix}\right)$ belonging to C_τ, we determined in Sec. III [Eq. (3.21a)] the evolution operator $U(M) = U_\tau(c,s)$ up to a phase $f(M) = f_\tau(c,s)$:

$$\langle y|U_t(c,s)|x\rangle$$
$$= \left|\frac{2}{s}\right|_p^{1/2} f_\tau(c,s)\chi\Big(\frac{c}{s}(x^2+y^2)-\frac{2}{s}xy\Big). \tag{6.1}$$

We will first fix the phase factor, then compute the complete spectrum with its degeneracies, and finally illustrate the calculation of wave functions.

A. Determination of $f_\tau(c,s)$

The phase factor $f_\tau(c,s)$ can again be chosen so that $U_\tau(c,s)$ defines a true representation of the group C_τ. As discussed in Sec. IV its eigenvalues are then C_τ characters. This leads immediately to the following functional equation:

$$f_\tau(c,s)f_\tau(c',s')\tau_p(ss's'') = f_\tau(c'',s''), \tag{6.2}$$

where $(c'',s'') = (cc'+\tau ss', sc'+s'c)$. [The factor τ_p disappears from Eq. (6.2) when $ss's''=0$.]

Equation (6.2) determines $f_\tau(c,s)$ up to an arbitrary character of C_τ, which we take equal to 1.

Let us consider the case $p \geqslant 3$, $\tau = \varepsilon$. We have shown [Eq. (4.35)] that

TABLE I. Solutions for the phase factor $f_\tau(c,s)$. [For $p=2$, it may happen that the phase factors associated to two equivalent extensions are different. This is the origin of a mistake in Eq. (32) of Ref. (8), where the result mentioned for $\tau=-3$ is in fact the phase for $\tau=5$.]

Extension $\mathbb{Q}_p(\sqrt{\tau})$	$f_\tau(c,s)$
$p\geqslant 3$, $\tau=\varepsilon$; $p=2$, $\tau=-1$ or 3	$\tau_p^{-1}(cs)$, $\lvert c\rvert_p>\lvert s\rvert_p$ 1, $\lvert c\rvert_p\leqslant\lvert s\rvert_p$
$p\geqslant 3$, $\tau=\alpha p(\alpha=1,\varepsilon)$; $p=2$, $\tau=\pm 2,\pm 6$	$\tau_p^{-1}(cs)$
$p=2$, $\tau=-3$	$\tau_2^{-1}(cs)$, $\lvert c\rvert_2>\lvert s\rvert_2$ $-\tau_2(cs)$, $\lvert c\rvert_2=\lvert s\rvert_2=2$

$$C_\varepsilon = Z_{p+1}\otimes C^1_\varepsilon .$$

Therefore we solve Eq. (6.2) separately on each factor group and then use the group law to obtain the general solution.

For $(\tilde c,\tilde s)$ belonging to C^1_ε, the solution reads

$$f_\varepsilon(\tilde c,\tilde s)=\tau_p^{-1}(\tilde s). \tag{6.3}$$

This is easily shown from Eqs. (A28) and (A30), using the fact that $(\tilde c,\tilde s)=(1+O(p^2),O(p))$ and that τ_p depends only on the first digit of its argument.

For the finite group Z_{p+1}, we can consistently choose

$$f_\varepsilon(\omega^j)=1,\quad 1\leqslant j\leqslant p+1.$$

Using Eq. (6.2) for $(c,s)\in Z_{p+1}$ and $(c',s')\in C^1_\varepsilon$ gives the solution quoted in Table I.

Exactly the same strategy can be applied in all other cases except $p=2$, $\tau=3$, where the solution is, however, easy to guess.

B. The spectrum of U_τ and its degeneracies

In terms of the "rational" parametrization of the circle [Eq. (2.41)], the propagator reads

$$\langle y|U_\tau(t)|x\rangle = \left|\frac{1-\tau t^2}{t}\right|_p^{1/2} f_\tau(t)\chi\left(\frac{(x-y)^2}{2t}+\frac{\tau t}{2}(x+y)^2\right). \tag{6.4}$$

Its spectral decomposition

$$\langle y|U_\tau(t)|x\rangle = \sum_\eta\sum_{i=1}^{d(\eta)}\langle y|\psi_{\eta,i}\rangle\langle\psi_{\eta,i}|x\rangle\eta\left(\frac{1+\sqrt{\tau}t}{1-\sqrt{\tau}t}\right) \tag{6.5}$$

allows us to calculate the degeneracy $d(\eta)$ of a given C_τ character η using the formulas [Eqs. (4.41) and (4.42)] of harmonic analysis on the group C_τ:

$$\begin{aligned} d(\eta) &= \int_{\mathbb{Q}_p}dx\int_{C_\tau}d\mu\langle x|U_\tau(t)|x\rangle\eta^*\left(\frac{1+\sqrt{\tau}t}{1-\sqrt{\tau}t}\right)\\ &= b_\tau\int_{\mathbb{Q}_p}dx\int_{\mathbb{Q}_p}dt\,\frac{f_\tau(t)}{|t(1-\tau t^2)|_p^{1/2}}\\ &\quad\times\chi(2\tau t x^2)\eta^*\left(\frac{1+\sqrt{\tau}t}{1-\sqrt{\tau}t}\right). \end{aligned}\tag{6.6}$$

The computation of this integral is rather lengthy. An explicit evaluation is presented in Appendix D for the circle C_ε when $p\geqslant 3$. The results are summarized in Tables II and III, which show that all eigenvalues are nondegenerate, as in the noncompact case, and that, roughly speaking, half the C_τ characters enter in the spectral decomposition of U_τ.

C. Eigenfunctions

The last problem is now the computation of the eigenfunctions. Since no eigenvalue is degenerate, the spectral decomposition given in Eq. (6.5) leads to the following relation:

TABLE II. Spectrum of the compact harmonic oscillators for $p\geqslant 3$.

Extension	Circle $Z_k\otimes C^1_\tau$	Generator $z=\omega^j g^n$	Multiplicity $d(\eta_m)$ of η characters of rank m $\eta_m(\omega^j g^n)=\chi\left(\frac{lj}{k}\right)\chi\left(\frac{bn}{p^{m-1}}\right)$
$Q_p(\sqrt{\varepsilon})$	$C_\varepsilon=Z_{p+1}\otimes C^1_\varepsilon$		$d(\eta_m)=\frac{1}{2}(1+(-)^m)$
$Q_p(\sqrt{\alpha p})$ $\alpha=1$ or ε $p\neq 3$	$C_{\alpha p}=Z_2\otimes C^1_{\alpha p}$	$g=\dfrac{1+\sqrt{\alpha p}u}{1-\sqrt{\alpha p}u}$ $\lvert u\rvert_p=1$	$d(\eta_0)=1$, $d(\eta_1)=0$ $d(\eta_m)=\frac{1}{2}\left[1+\left(\frac{-2ub}{p}\right)_L\left(\frac{-\alpha}{p}\right)_L^{m-1}\right]$ $(m\geqslant 2)$
$Q_3(\sqrt{3})$	$C_3=Z_2\otimes C^1_3$	$g=\dfrac{1+\sqrt{3}u}{1-\sqrt{3}u}$ $\lvert u\rvert_3=1$	$d(\eta_0)=1$, $d(\eta_1)=0$ $d(\eta_2)=\frac{1}{2}(1+(-ub/3)_L)$ $d(\eta_m)=\frac{1}{2}[1+(-)^m(ub/3)_L]$ $(m\geqslant 3)$
$Q_3(\sqrt{-3})$	$C_{-3}=Z_6\otimes C^1_{-3}$	$\omega=\frac{1}{2}+\frac{1}{2}\sqrt{-3}$ $g=\dfrac{1+\sqrt{-3}u}{1-\sqrt{-3}u}$ $\lvert u\rvert_3=\frac{1}{3}$	$d(\eta_0)=1$ $d(\eta_1)=\begin{cases}1, & \text{for } l=1,4\\ 0, & \text{for } l=2,3,5\end{cases}$ $d(\eta_m)=\frac{1}{2}[1+(ub/3)_L]$ $(m\geqslant 2)$

TABLE III. Spectrum of the compact harmonic oscillators for $p = 2$.

Extension	Circle	Generators	Multiplicity $d(\eta_m)$ of η characters of rank m
	$Z_k \otimes C^1_\tau$	$z = \omega^j g^n$	$\eta_m(\omega^j g^n) = \chi\left(\frac{lj}{k}\right)\chi\left(\frac{bn}{p^{m-1}}\right)$ $b = 1 + b_1 2 + b_2 2^2 + \cdots$
$Q_2(\sqrt{2\alpha})$ $\alpha = 1 + \alpha_1 \cdot 2 + \alpha_2 \cdot 2^2$ $(\tau = 2\alpha = \pm 2, \pm 6)$	$Z_2 \otimes C^1_{2\alpha}$	$g = \dfrac{1+\sqrt{2\alpha u}}{1-\sqrt{2\alpha u}}$ $u = 1 + 1\cdot 2 + (1+\alpha_1)2^2 + \cdots$	$d(\eta_0) = 1,\quad d(\eta_1) = 0$ $d(\eta_2) = \frac{1}{2}[1 + (-)^l]$ $d(\eta_3) = \frac{1}{2}[1 + (-)^{l+\alpha_1+b_1}]$ $d(\eta_4) = \frac{1}{2}[1 + (-)^{b_2+(1+b_1)\alpha_1}]$ $d(\eta_m) = \frac{1}{2}[1 + (-)^{m(\alpha_1+\alpha_2)+b_2+(1+b_1)(1+\alpha_1)}]\quad (m \geqslant 5)$
$Q_2(\sqrt{3})$	$Z_2 \otimes C^1_3$	$g = \dfrac{1+\sqrt{3}u}{1-\sqrt{3}u}$ $u = 1 + O(2^3)$	$d(\eta_0) = 1$ $d(\eta_1) = 0$ (see Appendix C) $d(\eta_2) = 1$ (see Appendix C) $d(\eta_3) = \frac{1}{4}(1 + (-)^l)(1 + (-)^{b_1})$ $d(\eta_4) = \frac{1}{2}(1 + (-)^{l+b_1})$ $d(\eta_m) = \frac{1}{2}(1 + (-)^{m+b_1})\quad (m \geqslant 5)$
$Q_2(\sqrt{-1})$	$Z_4 \otimes C^1_{-1}$	$\omega = \sqrt{-1}$ $u = 2 + O(2^3)$	$d(\eta_0) = 1,\quad d(\eta_1) = \frac{1}{2}(1 + (-)^l)$ $d(\eta_2) = \begin{cases} 1, & \text{for } l = 3,4 \\ 0, & \text{for } l = 1,2 \end{cases}$ $d(\eta_m) = \frac{1}{2}[1 + (-)^{b_1+1}]\quad (m \geqslant 3)$
$Q_2(\sqrt{-3})$	$Z_6 \otimes C^1_{-3}$	$\omega = \frac{1}{2} + \frac{1}{2}\sqrt{-3}$	$d(\eta_0) = 1$ $d(\eta_1) = \begin{cases} 1, & \text{for } l = 2,3,4 \\ 0, & \text{for } l = 1,5 \end{cases}$ $d(\eta_m) = \frac{1}{2}[1 + (-)^{m+1}]\quad (m \geqslant 2)$

$$\psi^*_\eta(x)\psi_\eta(y) = b_\tau \int_{Q_p} dt \frac{f_\tau(t)}{|t(1-\tau t^2)|_p^{1/2}} \times \chi\left(\frac{(x-y)^2}{2t} + \frac{\tau t}{2}(x+y)^2\right) \times \eta^*\left(\frac{1+\sqrt{\tau t}}{1-\sqrt{\tau t}}\right). \tag{6.7}$$

The following properties of the wave functions are worth pointing out.

(i) Eigenfunctions have definite *parity*. From the change of variable $u = 1/\tau t$ [see Eq. (A42)] [i.e., $(c,s) \to (-c,-s)$], one finds that

$$\psi^*(x)\psi(-y) = \eta(-1)\psi^*(x)\psi(y).$$

The eigenfunction with the character η as eigenvalue has the parity of η.

(ii) The eigenfunctions can be taken *real valued*. The complex conjugation of Eq. (6.7) followed by the change of variable $u = -t$ [i.e., $(c,s) \to (c,-s)$] gives indeed

$$[\psi^*_\eta(x)\psi_\eta(y)]^* = \psi^*_\eta(x)\psi_\eta(y). \tag{6.8}$$

(iii) Eigenfunctions have *compact support*. This can be seen by computing $P_k(\psi_\eta)$, the total probability for the system to be in the set $p^k Z_p^\times$ of p-adic numbers of norm p^{-k}:

$$P_k(\psi_\eta) = \int_{|x|_p = p^{-k}} dx\, \psi^*_\eta(x)\psi_\eta(x). \tag{6.9}$$

As an example, let us take $p \geqslant 5$, $\tau = \alpha p$. With the results of Appendix A, one finds, for every character of even rank m,

$$P_k(\psi_\eta) \neq 0 \text{ iff } -k = m/2, \tag{6.10}$$

while in the case of odd rank characters, parity introduces an additional distinction:

$$P_k(\psi_\eta) \neq 0 \text{ iff } -k \leqslant (m-1)/2, \text{ even } \eta, \tag{6.11a}$$

$$P_k(\psi_\eta) \neq 0 \text{ iff } -(m-3)/2 \leqslant -k \leqslant (m-1)/2, \text{ odd } \eta. \tag{6.11b}$$

(iv) Eigenfunctions are *locally constant*. We will not prove this fact but in the examples discussed in the remainder of this section it follows immediately [see Eqs. (6.12) and (6.15)].

Obtaining explicit expressions for the wave functions is rather complicated except when the η character is trivial. In this case, Eq. (6.7) gives, for any $\tau, p \neq 2$,

$$\psi_{\eta=1}(x) = \begin{cases} 1, & \text{if } |x|_p \leqslant 1, \\ 0, & \text{if } |x|_p > 1. \end{cases} \tag{6.12}$$

For η characters that are nontrivial, we limit ourselves to the case $p \geqslant 5$, $\tau = \alpha p$ ($\alpha = 1, \varepsilon$) and to even characters of odd rank, namely,

$$\eta(-1) = \eta(1) = 1, \tag{6.13a}$$

$$m = \text{rank } \eta = 2\nu + 1. \tag{6.13b}$$

Under these conditions, we learn from Table II that $(-2ub/p)_{\rm L}$ must equal $+1$ and one can then show that the wave function does not vanish at the origin. Hence we obtain the following integral representation of the wave function:

$$\psi_\eta(x) = \psi_\eta^{-1}(0)\int_{C_\tau} d\mu\, \tau_p^{-1}(cs)\chi\Big(\frac{c}{s}x^2\Big)\eta^*(c+\sqrt{\tau}s), \tag{6.14}$$

where $\psi_\eta^2(0) = 2p^{-\nu}$.

This integral is computed in Appendix E. The calculation is based on a method, due to Odoni,[17] for evaluating Gauss' sums mod p^n $(n>1)$. The final result reads

$$|x|_p = p^{-k} \leqslant p^{-\nu},\quad \psi_\eta(x) = \psi_\eta(0); \tag{6.15a}$$

$$|x|_p = p^{-\nu+1}$$

$$\psi_\eta(x) = \tfrac{1}{2}\psi_\eta(0)p^{-1/2}\varepsilon^*(p) \times \sum_{v=1}^{p-1}\Big(\frac{v}{p}\Big)_{\rm L}\chi\Big(\frac{-bv+v^{-1}p^{-2k}x^2}{p}\Big); \tag{6.15b}$$

$$p^{-\nu+1} < |x|_p \leqslant p^\nu,\quad \psi_\eta(x) = \psi_\eta(0)\cos\Big(\frac{2\pi G(y^*)}{p^{\nu-k}}\Big); \tag{6.15c}$$

$$|x|_p > p^\nu,\quad \psi_\eta(x) = 0. \tag{6.15d}$$

The function $G(y)$ depends on the label b of the character, the label τ of the circle C_τ, as well as on the value x at which the wave function is evaluated. It is defined by

$$G(y) = \sqrt{-bp^{-2k}x^2}\,(\tfrac{1}{2}\varphi_{\tau'}(y) + y^{-1}), \tag{6.16}$$

where

$$\varphi_{\tau'}(y) = 2y\sum_{n=0}^{\infty}\frac{(\tau'y^2)^n}{2n+1} \tag{6.17}$$

is the additive parameter on the circle C^1_τ, and

$$\tau' = \tau p^{2\nu}(x^2/-b). \tag{6.18}$$

Finally

$$y^* = \pm\sqrt{\frac{1}{(1+\tau')}} \tag{6.19}$$

satisfies

$$\frac{dG}{dy}\Big|_{y=y^*} = 0. \tag{6.20}$$

Note that it is not possible to give a more explicit form for $\psi_\eta(x)$ in Eq. (6.15b) since such Gauss' sums mod p cannot, in general, be evaluated in closed form.[18]

The algebraic structure of the wave functions will be examined elsewhere but it is worth pointing out that there is a rather remarkable relation among eigenfunctions corresponding to characters of the same rank, namely,

$$\psi_{\eta^{\sigma(a^2)}}(xa) = (\psi_\eta(x))^{\sigma(a^2)}. \tag{6.21}$$

In this equation $\sigma(a^2)$—with a any p-adic unit—is the following (Galois) map among the complex $p^{2\nu}$th roots of unity:

$$\sigma(a^2){:}\exp\{2\pi i/p^{2\nu}\}\to\exp\{(2\pi i/p^{2\nu})a^2\}.$$

The reader can verify that the wave functions are indeed expressed in terms of such roots. Varying a over the residues modulo $p^{2\nu}$ prime to p, Eq. (6.21) gives all wave functions corresponding to even characters of odd rank $2\nu+1$ in terms of one of them.

VII. CONCLUSION AND OUTLOOK

In this paper we have given a consistent formulation of quantum mechanics on p-adic number fields for simple systems. The essential difference between real and p-adic quantum mechanics stems from a topological property of the corresponding number fields: while $\mathbb{R}$ is connected and thus allows for an infinitesimal as well as a global formulation of quantum mechanics, $\mathbb{Q}_p$ is disconnected and hence only a global (Weyl) formulation seems possible. There is no infinitesimal generator of translation in p-adic space, i.e., no momentum operator. Similarly there is no infinitesimal evolution operator (Hamiltonian). Finite transformations remain, however, well defined and, in particular, there is a "global" evolution operator, which we have explicitly computed. Quantum mechanics takes place in a Hilbert space of complex valued functions of a p-adic variable. We have restricted ourselves to "configuration space" where the wave functions are labeled by a p-adic index, which refers to the "p-adic position."

The quadratic one-dimensional systems we have worked out in this paper have the characteristic feature that their classical evolution is described by a one parameter Abelian subgroup of $SL(2,\mathbb{R})$. We have defined the p-adic analogs of these systems by corresponding subgroups of $SL(2,\mathbb{Q}_p)$. Our study of their quantum mechanical properties relies in an essential way on a precise analysis of p-adic Gaussian integrals.

We have shown that the eigenvalues of the quantum evolution operators are simply characters of these Abelian subgroups. Characters thus provide an elegant and universal solution to the dynamical problem at hand irrespective of the number field one uses. This is a truly remarkable fact.

For the free particle or particle in a constant field we have completely determined the spectrum of the evolution operator and explicitly constructed the corresponding eigenfunctions. The similarity with plane waves or Airy functions is rather striking.

For harmonic oscillators the spectrum has been entirely determined as well and we have illustrated the computation of some wave functions. This required an application of a method due to Odoni for calculating Gauss' sums mod p^n $(n\geqslant 2)$. It is, however, not possible to give a completely closed form to these wave functions because of the presence of Gauss' sums mod p.

In ordinary quantum mechanics, time is not an operator but a parameter. In our formulation of p-adic quantum mechanics, evolution has been identified through group theoretical laws. One may wonder if it is possible to define a p-adic "time" parameter. In our context this would simply mean an *additive* parametrization of the *full* group of motion. For a free particle the answer is trivial, while for harmonic oscillators the results of Sec. IV and Appendix C imply that such a parameter can only be defined on a subgroup.

One of our motivations for studying p-adic quantum mechanics is to develop tools for p-adic strings. The detailed study of characters presented in this paper already suggests various possibilities for string amplitudes that will be examined elsewhere.

ACKNOWLEDGMENTS

It is a pleasure to acknowledge enlightening discussions with Y. Meurice. We also benefited from the collaboration of C. Alacoque in the early stage of this work.

APPENDIX A: p-ADIC INTEGRATION

The purpose of this appendix is to explain in some detail how to compute p-adic "Gaussian integrals."

The Haar measure dx is the (essentially) unique invariant measure on the additive group $\mathbb{Q}_p^+$: for any $a\in\mathbb{Q}_p^+$,

$$d(x+a)=dx. \tag{A1}$$

Its normalization is fixed by taking the measure of $\mathbb{Z}_p$, the set of p-adic integers, as equal to 1:

$$\mu(\mathbb{Z}_p)=\int_{\mathbb{Z}_p}dx=\int_{|x|_p\leqslant 1}dx=1. \tag{A2}$$

It is now straightforward to calculate the measure of any ball

$$U(a;n)=\{x\in\mathbb{Q}_p:\ |x-a|_p\leqslant p^{-n}\}.$$

Consider, for example, $p^{-k}\mathbb{Z}_p=U(0;-k)$, i.e., the set of numbers

$$x=p^{-k}(x_0+x_1p+x_2p^2+\cdots), \tag{A3}$$

with $|x|_p\leqslant p^k$. Let us start with $k=1$, i.e., $p^{-1}\mathbb{Z}_p$. Using Eq. (2.13),

$$p^{-1}\mathbb{Z}_p=\bigcup_{x_0=0}^{p-1}U(x_0p^{-1};0); \tag{A4}$$

hence

$$\mu(p^{-1}\mathbb{Z}_p)=\sum_{x_0=0}^{p-1}\mu(U(x_0p^{-1};0))=\sum_{x_0=0}^{p-1}1=p. \tag{A5}$$

Repeating the reasoning one gets

$$\mu(p^{-k}\mathbb{Z}_p)=p^k. \tag{A6}$$

Let us now take the set of numbers with a given p-adic norm p^k. Clearly

$$\mu(\{|x|_p=p^{+k}\})=\mu(p^{-k}\mathbb{Z}_p)-\mu(p^{-k+1}\mathbb{Z}_p)$$
$$=p^k(1-p^{-1}). \tag{A7}$$

The last two formulas are essentially all one needs for integration over $\mathbb{Q}_p$ or any of its subsets.

Suppose, for example, that one wants to compute

$$\int_{\mathbb{Z}_p}dx\,f(x), \tag{A8}$$

for a "sufficiently well-behaved function" $f:\mathbb{Z}_p\to\mathbb{C}$. In the spirit of Riemann integrals on $\mathbb{R}$, one considers a sequence of partitions of $\mathbb{Z}_p$,

$$\mathscr{P}_N(\mathbb{Z}_p)=\bigcup_{j=0}^{p^N-1}U(j;N), \tag{A9}$$

and computes Eq. (A8) as the limit

$$\int_{\mathbb{Z}_p}dx\,f(x)=\lim_{N\to\infty}\sum_{j=0}^{p^N-1}f(x^{(j)})p^{-N},\quad x^{(j)}\in U(j,N). \tag{A10}$$

In calculating integrals, it is often useful to be able to change variables. Formula (A6) gives us the "Jacobian" for any scale transformation

$$d(ax)=|a|_p\,dx,\quad a\neq 0. \tag{A11}$$

For Gaussian integrals it is also practical to perform quadratic changes of variables:

$$dx^2=\tfrac{1}{2}|2x|_p\,dx. \tag{A12}$$

Concerning this formula, valid for all p, it is worth pointing out that the factor $\frac{1}{2}$ is necessary to prevent "overcounting": the integration on x is over *both* square roots ($\pm x$) of x^2. The following calculation is illustrative. Let us compute the measure of $\mathbb{Z}_p^2$, i.e., the set of squares in $\mathbb{Z}_p$ ($p\neq 2$). Remembering that there are $(p-1)/2$ squares mod p in $\mathbb{F}_p^\times$ and using Eq. (A6) gives

$$\int_{\mathbb{Z}_p^2}dy=\sum_{\alpha=0}^{\infty}\sum_{\substack{y_0=1\\(y_0/p)_L=+1}}^{p-1}\int_{\{p^{2\alpha}(y_0+\cdots)\}}dy$$
$$=\frac{p-1}{2}\sum_{\alpha=0}^{\infty}p^{-2\alpha-1}=\tfrac{1}{2}(1+p^{-1})^{-1}. \tag{A13a}$$

Alternatively, using Eq. (A12),

$$\int_{\mathbb{Z}_p^2}dy=\tfrac{1}{2}\int_{\mathbb{Z}_p}dx|2x|_p=\tfrac{1}{2}\sum_{\alpha=0}^{\infty}p^{-\alpha}\int_{|x|_p=p^{-\alpha}}dx$$
$$=\tfrac{1}{2}(1+p^{-1})^{-1}. \tag{A13b}$$

We can now proceed to evaluate p-adic "Gaussian integrals,"

$$G(a)=\int_{\mathbb{Q}_p}dx\,\chi(ax^2),\quad a\neq 0, \tag{A14}$$

where $\chi(z)=\exp 2\pi iz$ is the usual character on $\mathbb{Q}_p$. We write

$$G(a)=\sum_{m=-\infty}^{+\infty}G_m(a)=\sum_{m=-\infty}^{+\infty}\int_{|x|_p=p^{-m}}dx\,\chi(ax^2), \tag{A15}$$

where $G_m(a)$ is an "incomplete" Gaussian integral. Changing variables in Eq. (A15) and focusing on the case $p\neq 2$, we have, with $t=ax^2$,

$$dt=\tfrac{1}{2}|a|_p|x|_p\,dx, \tag{A16}$$

and $\operatorname{ord} t=k=2m+\operatorname{ord} a$,

$$G_m(a)=\frac{2}{|a|_p}p^m\int_{\substack{|t|_p=p^{-k}\\ t\in\mathrm{Ran}(ax^2)}}dt\,\chi(t). \tag{A17}$$

If $k\geqslant 0$, $t\in Z_p$ and hence $\chi(t)=1$; it follows that

$$G_m(a)=(2/|a|_p)p^{m-k}(1-p^{-1})\tfrac{1}{2}=p^{-m}(1-p^{-1}), \tag{A18}$$

where $2m\geqslant-\operatorname{ord} a$. The factor $\frac{1}{2}$ keeps track of the constraint on the first digit t_0 of $t\in\mathrm{Ran}(ax^2)$, namely, $(t_0/p)_L=(a_0/p)_L$. If $k=-1$, one readily obtains, using the character property of $\chi(t)$,

$$G_m(a)=\frac{2}{|a|_p}p^m \sum_{\substack{t_0=1\\(t_0/p)_L=(a_0/p)_L}}^{p-1} \exp\left\{2\pi i\frac{t_0}{p}\right\}. \quad (A19)$$

Standard number theory gives for the quadratic Gauss' sum[15]

$$\sum_{\substack{t=1\\(t/p)_L=(a/p)_L}}^{p-1} \exp\left\{2\pi i\frac{t}{p}\right\}=\frac{1}{2}\left\{\varepsilon(p)\sqrt{p}\left(\frac{a}{p}\right)_L-1\right\}, \quad (A20)$$

where

$$\varepsilon(p)=i^{((p-1)/2)^2}. \quad (A21)$$

Equation (A19) thus reads

$$G_m(a)=\frac{1}{|a|_p^{1/2}}\left[\varepsilon(p)\left(\frac{a}{p}\right)_L-\frac{1}{\sqrt{p}}\right], \quad (A22)$$

for $2m=-1-\text{ord }a$.

Finally if $k<-1$ we will show that $G_m(a)=0$. The argument is simple and is used again and again in computing integrals on Q_p. It rests on the following identity:

$$\sum_{x=0}^{p-1}\exp\left\{2\pi i\frac{x}{p}\right\}=0. \quad (A23)$$

Noting that the constraint $t\in\text{Ran}(ax^2)$ is a constraint only on the first digit t_0 of t and repeating the previous argument,

$$G_m(a)=\frac{2}{|a|_p}p^m \sum_{\substack{t_0=1\\(t_0/p)_L=(a_0/p)_L}}^{p-1} \exp\left\{2\pi i\frac{t_0}{p^k}\right\} \times\sum_{t_1=0}^{p-1}\cdots\sum_{t_{k-1}=0}^{p-1}\exp\left\{2\pi i\frac{t_{k-1}}{p}\right\}. \quad (A24)$$

The identity (A23) then completes the proof that $G_m(a)=0$, for $k<-1$. Summing up Eq. (A18) and Eq. (A22) gives, for the complete Gaussian integral,

$$G(a)=\sum_{m=-\infty}^{+\infty}G_m(a)=\frac{1}{|2a|_p^{1/2}}\tau_p(a), \quad (A25)$$

where, for $a=p^{\text{ord }a}(a_0+a_1p+a_2p^2+\cdots)$ and $p\neq 2$,

$$\tau_p(a)=1, \quad \text{if ord } a \text{ is even}, \quad (A26a)$$

$$\tau_p(a)=\varepsilon(p)(a_0/p)_L, \quad \text{if ord } a \text{ is odd}. \quad (A26b)$$

Repeating the calculation for $p=2$ gives

$$\tau_2(a)=[1+i(-1)^{a_1}]/\sqrt{2}, \quad \text{if ord } a \text{ is even}, \quad (A27a)$$

$$\tau_2(a)=(1+i)i^{a_1+2a_2}/\sqrt{2}, \quad \text{if ord } a \text{ is odd}. \quad (A27b)$$

The following relations are very often used in the main text:

$$\tau_p(b^2a)=\tau_p(a), \quad (A28)$$

$$\tau_p(a)\tau_p(-a)=1, \quad (A29)$$

$$\tau_p(a)\tau_p(b)=\tau_p(a+b)\tau_p(ab/(a+b)), \quad (A30)$$

$$\tau_p(a)\tau_p(-\tau a)=\tau_p(1)\tau_p(-\tau)\text{sgn}_\tau a. \quad (A31)$$

Equation (A30) is a consequence of the identity

$$\frac{1}{a}(z-at)^2+\frac{1}{b}(z+bt)^2=\left(\frac{1}{a}+\frac{1}{b}\right)z^2+(a+b)t^2. \quad (A32)$$

The last identity [Eq. (A31)] is interesting to derive. One starts from

$$G(a)G(-\tau a)=\int_{Q_p}dx\int_{Q_p}dy\,\chi(a(x^2-\tau y^2))$$
$$=\int_{Q_p(\sqrt{\tau})}dz\,\chi(az\bar{z}). \quad (A33)$$

The integral on the extension field $Q_p(\sqrt{\tau})$ can easily be computed by going to "polar coordinates"

$$z=x+\sqrt{\tau}y=r(c+\sqrt{\tau}s)=r(1+\sqrt{\tau}t)/(1-\sqrt{\tau}t), \quad (A34)$$

with $z\bar{z}=r\bar{r}=\rho$, $\text{sgn}_\tau\,\rho=+1$, and $c+\sqrt{\tau}\,s$ an element of the unit circle C_τ written here in terms of the t parametrization [see Eq. (2.41)]. The "Jacobian" is given by

$$dz=dx\,dy=a_\tau\,d\rho\,d\mu(t), \quad (A35)$$

where $d\mu(t)$ is the invariant measure on the group C_τ normalized to 1 [Eq. (4.39)]. The constant a_τ computed in Ref. 3 is given by [$a_\tau=b_\tau^{-1}$, Eq. (4.40)]

$$a_\tau=\begin{cases}1+p^{-1}, & \text{when } p\neq 2,\ \tau=\varepsilon,\\ 3, & \text{when } p=2,\ \tau=-3,\\ 2, & \text{in any other case.}\end{cases} \quad (A36)$$

Remembering that $\text{sgn}_\tau(x)$ is a character [Eq. (2.37)], one obtains

$$\int_{Q_p(\sqrt{\tau})}dz\,\chi(az\bar{z})$$
$$=a_\tau\int_{Q_p}du\,\chi(au)\left(\frac{1+\text{sgn}_\tau\,u}{2}\right)$$
$$=(a_\tau/2)\{\delta(a)+(\text{sgn}_\tau\,a/|a|_p)\Gamma_\tau(1)\}, \quad (A37)$$

where $\Gamma_\tau(1)$ is the Gel'fand gamma function defined by

$$\Gamma_\tau(1)=\int_{Q_p^\times}dx\,\chi(x)\text{sgn}_\tau(x). \quad (A38)$$

Combining Eq. (A37) with Eq. (A33) and Eq. (A25) gives

$$\frac{\tau_p(a)\tau_p(-\tau a)}{|2a|_p|-\tau|_p^{1/2}}=\frac{a_\tau}{2}\left\{\delta(a)+\frac{\text{sgn}_\tau\,a}{|a|_p}\Gamma_\tau(1)\right\}. \quad (A39)$$

Taking $a=1$, one obtains

$$\Gamma_\tau(1)=(2/a_\tau)\tau_p(1)\tau_p(-\tau)/|2|_p|\tau|_p^{1/2}. \quad (A40)$$

It is amusing to note that this formula determines an unspecified sign in Ref. 3 (pp. 149–150). The identity (A31) is simply Eq. (A39) for $a\neq 0$.

Finally, the reader can now test his skill in p-adic integration by deriving the incomplete shifted Gaussian integral ($p\neq 2$)

$$I_k=\int_{|x|_p<p^k}dx\,\chi(a(x-1)^2),$$

which gives, for $k\geqslant 0$,

$$I_k = \begin{cases} p^k, & \text{if} \quad k \leqslant \frac{1}{2} \text{ ord } a, \\ (1/|a|_p^{1/2})\tau_p(a), & \text{if} \quad k \geqslant \frac{1}{2}(1 + \text{ord } a), \end{cases} \tag{A41a, A41b}$$

and, for $k \leqslant -1$,

$$I_k = \begin{cases} p^k \chi(a), & \text{if} \quad k \leqslant \text{ord } a, \quad \text{(A41c)} \\ 0, & \text{if} \quad k \geqslant \text{ord } a + 1. \quad \text{(A41d)} \end{cases}$$

We conclude this appendix by giving the Jacobian needed for a generic fractional linear change of variable:

$$d\left(\frac{ax+b}{cx+d}\right) = |cx+d|_p^{-2} |ad-bc|_p \, dx. \tag{A42}$$

APPENDIX B: THE PARTICLE IN A CONSTANT FORCE FIELD

For the quadratic Lagrangian

$$L = (m/2)\dot{x}^2 - Fx, \tag{B1}$$

the solution of the classical equations of motion is of the form

$$\begin{pmatrix} x(t) \\ p(t) \end{pmatrix} = \begin{pmatrix} a(t) & b(t) \\ c(t) & d(t) \end{pmatrix} \begin{pmatrix} x(0) \\ p(0) \end{pmatrix} + \begin{pmatrix} e(t) \\ f(t) \end{pmatrix} = M(t) \begin{pmatrix} x(0) \\ p(0) \end{pmatrix} + C(t). \tag{B2}$$

The classical evolution is thus governed by an Abelian one-dimensional subgroup of the semidirect product $SL(2,\mathbb{R}) \circledS T_2$, where T_2 is the two-dimensional translation group. The group composition law reads

$$(M,C)\circ(M',C') = (MM',MC'+C). \tag{B3}$$

The inhomogeneous case is quite similar to the homogeneous one treated in the main text: the Weyl formulation of quantum mechanics remains meaningful when M and C are valued on $\mathbb{Q}_p$. Proceeding as in Sec. III yields

$$\langle y|U|x\rangle = \begin{cases} \dfrac{f(M,C)}{|hb|_p^{1/2}} \chi\left(\dfrac{1}{2hb}[ax^2 + dy^2 - 2xy + 2(bf-de)y + 2xe]\right), & b \neq 0, \quad \text{(B4a)} \\ |a|_p^{1/2} f(M,C) \chi\left(\dfrac{ac}{2h}x^2 + \dfrac{af}{h}x\right) \delta(ax+e-y), & b = 0. \quad \text{(B4b)} \end{cases}$$

The case at hand, Eq. (B1), corresponds to the following parabolic Abelian subgroup of $SL(2,\mathbb{Q}_p) \circledS T_2$:

$$M = \begin{pmatrix} 1 & t/m \\ 0 & 1 \end{pmatrix}, \quad C = \begin{pmatrix} -Ft^2/2m \\ -Ft \end{pmatrix}. \tag{B5}$$

We exploit once again the fact that the group (M,C) is conjugate to a simpler one:

$$(M,C) = (S,0)\circ(\widetilde{M},\widetilde{C})\circ(S^{-1},0) = (S\widetilde{M}S^{-1}, S\widetilde{C}),$$

$$S = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, 0 = \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \widetilde{M} = \begin{pmatrix} 1 & 0 \\ -t/m & 1 \end{pmatrix},$$

$$\widetilde{C} = S^{-1}C = \begin{pmatrix} Ft \\ -Ft^2/2m \end{pmatrix}. \tag{B6}$$

For the evolution operator $U(\widetilde{M},\widetilde{C})$ one then finds, from Eq. (B4b),

$$\langle y|U(\widetilde{M},\widetilde{C})|x\rangle = \chi\left(\frac{-F^2t^3}{6mh}\right) \chi\left(\frac{-t}{2mh}x^2 - \frac{Ft^2}{2mh}x\right) \delta(x+Ft-y). \tag{B7}$$

The phase factor $f(M,C)$ has been chosen so as to obtain a true representation of the group specified by Eq. (B5). The group of elements $(\widetilde{M},\widetilde{C})$ is isomorphic to $\mathbb{Q}_p^+$ and the eigenvalues $\lambda(t)$ of $U(\widetilde{M},\widetilde{C})$ are thus additive characters

$$\lambda_E(t) = \chi(-Et/h), \quad E \in \mathbb{Q}_p. \tag{B8}$$

The corresponding eigenfunctions are given by

$$\varphi_E(x) = N\chi(-x^3/6mFh + Ex/Fh), \tag{B9}$$

where N is a normalization factor. Using the conjugacy relation (B6), one obtains,[8] for the operator $U(M,C)$,

$$\langle y|U(M,C)|x\rangle = \left|\frac{m}{ht}\right|_p^{1/2} \tau_p^{-1}\left(\frac{m}{2ht}\right) \times \chi\left[\frac{m(x-y)^2}{2ht} - \frac{Ft}{2h}(x+y) - \frac{F^2t^3}{24mh}\right] \tag{B10}$$

while the eigenfunction corresponding to the eigenvalue [Eq. (B8)] is

$$U(S,0)|\varphi_E\rangle = |\psi_E\rangle, \tag{B11}$$

which reads, explicitly,

$$\psi_E(x) = |h^2F|_p^{-1/2} \int dy \, \chi\left(-\frac{y^3}{6mhF} + \frac{y}{h}\left(\frac{E}{F} - x\right)\right). \tag{B12}$$

Its normalization is such that

$$\langle \psi_{E'}|\psi_E\rangle = \delta(E-E'). \tag{B13}$$

APPENDIX C: MISCELLANEOUS RESULTS

In this appendix we complete the analysis given in Sec. IV. We will first introduce the p-adic functions log and exp, which provide the appropriate tools for discussing isomorphisms between additive and multiplicative groups. These functions are also of practical use in determining generators of the various cyclic groups discussed in Sec. IV. This appendix also contains a description of the various multiplicative and "τ-elliptic" groups left out of the main text.

We define the p-adic functions[3,4]

$$\log(1+x) = x - \frac{x^2}{2} + \frac{x^3}{3} - \cdots, \tag{C1}$$

$$\exp(y) = 1 + y + \frac{y^2}{2!} + \cdots. \tag{C2}$$

These series converge p-adically iff, $|x|_p < 1$ and $|y|_p < p^{-1/(p-1)}$, respectively. Note that x and y may belong to any finite extension of $\mathbb{Q}_p$.

Provided that the convergence conditions are satisfied one has

$$\log(xy) = \log(x) + \log(y), \tag{C3}$$

$$\exp(x+y) = \exp(x)\cdot\exp(y), \tag{C4}$$

$$\log(\exp(x)) = x, \tag{C5}$$

$$\exp(\log(x)) = x. \tag{C6}$$

1. The groups $\mathbb{Q}_p^\times$ ($p \geqslant 3$)

If $x \in O_1$ (with $\mathbb{Q}_p^\times = \mathbb{Z} \otimes Z_{p-1} \otimes O_1$), $|\log x|_p \leqslant p^{-1}$ and the *multiplicative group* O_1 is thus isomorphic to the *additive group* $p\mathbb{Z}_p$. Similarly for an element x of the subgroup $O_m = \{1 + O(p^m)\}$, $|\log x|_p \leqslant p^{-m}$; hence O_m is isomorphic to $p^m\mathbb{Z}_p$. Then O_1/O_m is cyclic since it is isomorphic to the finite additive group $p\mathbb{Z}_p/p^m\mathbb{Z}_p$. A generator of the latter group is any element of norm p^{-1}; hence a generator of O_1/O_m is any g with $|\log g|_p = p^{-1}$ or, equivalently, $|g-1|_p = p^{-1}$.

The product $Z_{p-1} \otimes O_1/O_m$ is also cyclic, since $p-1$ and p^{m-1} are coprime.

2. The group $\mathbb{Q}_2^\times$

One still has $\mathbb{Q}_2^\times = \mathbb{Z} \otimes \mathbb{Z}_2^\times$, but the analysis of this group cannot proceed as in Sec. IV. Although the elements of $\mathbb{Z}_2^\times$ have a well-defined log with $|\log y|_2 \leqslant 2^{-1}$, the exp function only converges for $|y|_2 \leqslant 2^{-2}$: the isomorphism between $\mathbb{Z}_2^\times$ and some additive subgroup of $\mathbb{Z}_2$ is lost! The reason is that $\log 1 = \log(-1) = 0$; hence log is not injective anymore. The problem is, however, easily solved by recognizing that any element of $\mathbb{Z}_2^\times$ can be uniquely written as

$$y = (-)^j(1 + 0\cdot 2 + y_2\cdot 2^2 + \cdots), \quad j = 0,1, \tag{C7}$$

i.e.,

$$\mathbb{Z}_2^\times = Z_2 \otimes O_2. \tag{C8}$$

Elements of the group O_2, $x = 1 + O(2^2)$ have $|\log x|_2 \leqslant 2^{-2}$ and are isomorphic to the additive group $2^2\mathbb{Z}_2$. The "sheet" of $\mathbb{Z}_2^\times$ that contains -1 has, so to speak, been "factored out."

Repeating the previous argument shows that any element g with $|\log g|_2 = 2^{-2}$ is a generator of the cyclic group O_2/O_m. Note that this time the product $Z_2 \otimes O_2/O_m$ is not cyclic! (The orders of the factors, 2 and 2^{m-2}, are not coprime.) Characters of rank $m > 2$ take the form ($h \in O_m$)

$$\pi(x) = \pi(p^k(-)^j g^n h) = \exp(2\pi i k\rho)\theta((-1)^j g^n),$$
$$\theta((-1)^j g^n) = (-1)^{lj}\chi(an/2^{m-2}),$$
$$\rho \in [0,1[, \quad l = 0,1, \quad 1 \leqslant a \leqslant 2^{m-2}, a \text{ odd}. \tag{C9}$$

The characters of rank 0 are $\pi(x) = \exp(2\pi i k\rho)$ and those of rank 2 are $\pi(x) = (-1)^j \exp(2\pi i k\rho)$. The completeness and orthogonality relations are still given by Eqs. (4.24)–(4.26).

3. Additive parametrization of subgroups of the elliptic group C_τ

We now establish an isomorphism between C_τ^1, Eqs. (4.29) and (4.35), and an additive subgroup of $\mathbb{Z}_p$ (except for $p = 2$ and $p = 3$, $\tau = \pm 3$).

For any element $z = c + \sqrt{\tau}\, s$ of C_τ^1, we define the $\mathbb{Q}_p$ valued function of the parameter t [Eq. (2.41)]:

$$\varphi_\tau(t) = \frac{1}{\sqrt{\tau}}\log z = \frac{1}{\sqrt{\tau}}\log\frac{1+\sqrt{\tau}t}{1-\sqrt{\tau}t} = 2t\sum_{n=0}^{\infty}\frac{(\tau t^2)^n}{2n+1}. \tag{C10}$$

This series converges iff $|\tau t^2|_p < 1$. On the other hand, the reciprocal map $z = \exp(\sqrt{\tau}\varphi_\tau)$, as defined in Eq. (C2), converges iff $|\sqrt{\tau}\varphi_\tau|_p < p^{-1/(p-1)}$.

By construction the function φ_τ is *additive* with respect to the group law

$$\varphi_\tau(t'') = \varphi_\tau\left(\frac{t+t'}{1+\tau t t'}\right) = \varphi_\tau(t) + \varphi_\tau(t'). \tag{C11}$$

From $|\varphi_\tau|_p = |t|_p$ it thus follows that

$$C_{\alpha p}^1 \approx \mathbb{Z}_p, \quad C_{\alpha p}^m \approx p^{m-1}\mathbb{Z}_p, \quad \alpha = 1,\varepsilon, \tag{C12a}$$

$$C_\varepsilon^1 \approx p\mathbb{Z}_p, \quad C_\varepsilon^m \approx p^m\mathbb{Z}_p. \tag{C12b}$$

As a consequence the finite groups C_τ^1/C_τ^m are cyclic of order p^{m-1}. Obviously a generator of $C_\varepsilon^1/C_\varepsilon^m$ is any element $g = (1+\sqrt{\varepsilon u})/(1-\sqrt{\varepsilon u})$ with $|u|_p = p^{-1}$, while for $C_{\alpha p}^1/C_{\alpha p}^m$ the same form for g holds with $|u|_p = 1$.

4. The special cases

For completeness we now describe the τ-elliptic groups for $p = 3$, $\tau = \pm 3$ and for $p = 2$.

Proceeding as in Sec. IV we factorize C_τ as a product of a finite (cyclic) subgroup Z_k and an infinite cyclic group C_τ^1. Then Z_k contains all elements of finite order, while C_τ^1 again contains a sequence of subgroups C_τ^m.

More precisely the following results hold:

$p = 3, \quad \tau = -3,$

$$C_{-3} = Z_6 \otimes C_{-3}^1, \quad Z_6 = \{\pm 1, \pm\tfrac{1}{2} \pm \sqrt{-3}/2\},$$
$$C_{-3}^m = \left\{z = \frac{1+\sqrt{-3}t}{1-\sqrt{-3}t}; \quad |t|_3 \leqslant 3^{-m}\right\}; \tag{C13a}$$

$p = 3, \quad \tau = 3,$

$$C_3 = Z_2 \otimes C_3^1, \quad Z_2 = \{\pm 1\},$$
$$C_3^m = \left\{z = \frac{1+\sqrt{3}t}{1-\sqrt{3}t}; \quad |t|_3 \leqslant 3^{-m+1}\right\}; \tag{C13b}$$

$p = 2, \quad \tau = \pm 2, \pm 6,$

$$C_\tau = Z_2 \otimes C_\tau^1, \quad Z_2 = \{\pm 1\},$$
$$C_\tau^m = \left\{z = \frac{1+\sqrt{\tau}t}{1-\sqrt{\tau}t}; \quad |t|_2 \leqslant 2^{-m+1}\right\}; \tag{C13c}$$

$p = 2, \quad \tau = -1,$

$$C_{-1} = Z_4 \otimes C_{-1}^1, \quad Z_4 = \{\pm 1, \pm\sqrt{-1}\},$$
$$C_{-1}^m = \left\{z = \frac{1+\sqrt{-1}t}{1-\sqrt{-1}t}; \quad |t|_2 \leqslant 2^{-m}\right\}; \tag{C13d}$$

$p=2, \tau=-3,$

$$C_{-3}=Z_6\otimes C^1_{-3},\quad Z_6=\{\pm 1, \pm\tfrac{1}{2}\pm\sqrt{-3}/2\},$$

$$C^m_{-3}=\left\{z=\frac{1+\sqrt{-3}t}{1-\sqrt{-3}t};\quad |t|_2\leqslant 2^{-m}\right\}; \tag{C13e}$$

$p=2,\quad \tau=3,$

$$C_3=Z_2\otimes C_3^1,\quad Z_2=\{\pm 1\},$$

$$C_3^1=\left\{z=\frac{1+\sqrt{3}t}{1-\sqrt{3}t};\right.$$

$$\left.|t|_2=2 \text{ or } |t|_2\leqslant\tfrac{1}{4} \text{ or } t=\pm 1+O(2^3)\right\},$$

$$C_3^2=\left\{z=\frac{1+\sqrt{3}t}{1-\sqrt{3}t};\quad |t|_2=2 \text{ or } |t|_2\leqslant\tfrac{1}{4}\right\},$$

$$C_3^m=\left\{z=\frac{1+\sqrt{3}t}{1-\sqrt{3}t};\quad |t|_2\leqslant 2^{-m+1}\right\}. \tag{C13f}$$

The cyclicity of C^1_τ can be shown either with the help of the log and exp functions ($\tau\neq 3$) or by hand ($\tau=3$). In all cases, the order of the cosets C^1_τ/C^m_τ is p^{m-1}.

We thus have the following list of characters of rank m on $C_\tau=Z_k\otimes C^1_\tau/C^m_\tau$ (ω and g generate Z_k and C^1_τ/C^m_τ, respectively):

rank 0, $\eta(z)=1;$

rank 1,

$$\eta(z=\omega^j h)=\chi(lj/k),\quad 1\leqslant l\leqslant k-1;$$

rank $m\geqslant 2$,

$$\eta(z=\omega^j g^n h)=\chi(lj/k)\chi(bn/p^{m-1}),$$

$$1\leqslant l\leqslant k,\quad 1\leqslant b\leqslant p^{m-1},\quad p\nmid b;$$

where h belongs to C^m_τ.

APPENDIX D: DEGENERACIES OF THE COMPACT HARMONIC OSCILLATOR SPECTRUM ($\tau=\varepsilon$)

In this appendix, we indicate how the integral (6.6) may be computed, by considering the cases $p\neq 2$ $\tau=\varepsilon$. The degeneracy of a character η on the circle C_ε is given by

$$d(\eta)=(1+p^{-1})^{-1}\int_{\mathbb{Q}_p}dx\int_{\mathbb{Q}_p}\frac{dt}{|t(1-\varepsilon t^2)|_p^{1/2}}$$
$$\times f_\varepsilon(t)\chi(2\varepsilon t x^2)\eta^*\left(\frac{1+\sqrt{\varepsilon}t}{1-\sqrt{\varepsilon}t}\right). \tag{D1}$$

We recall that, for an η character of rank m,

$$\eta\left(\frac{1+\sqrt{\varepsilon}t}{1-\sqrt{\varepsilon}t}\right)=\begin{cases}\eta(1)=1, & \text{for} \quad |t|_p\leqslant p^{-m},\\ \eta(-1), & \text{for}\quad |t|_p\geqslant p^{-m},\end{cases}$$

and that the phase factor $f_\varepsilon(t)$ (see Table I) can be recast in the form

$$f_\varepsilon(t)=\begin{cases}\tau_p^{-1}(2t), & |t|_p<1,\\ 1, & |t|_p=1,\\ \tau_p^{-1}(2\varepsilon t), & |t|_p>1.\end{cases}$$

We thus split the integral (D1) in four pieces:

$$\begin{aligned}(1+p^{-1})d(\eta)&=\int dx\int_{|t|_p\leqslant p^{-m}}dt\,|t|_p^{-1/2}\tau_p^{-1}(2t)\chi(2\varepsilon t x^2)\\
&+\int dx\int_{p^{-m+1}\leqslant|t|_p\leqslant p^{-1}}dt\,|t|_p^{-1/2}\tau_p^{-1}(2t)\chi(2\varepsilon t x^2)\eta^*\left(\frac{1+\sqrt{\varepsilon}t}{1-\sqrt{\varepsilon}t}\right)\\
&+\int dx\int_{|t|_p=1}dt\,\chi(2\varepsilon t x^2)\eta^*\left(\frac{1+\sqrt{\varepsilon}t}{1-\sqrt{\varepsilon}t}\right)\\
&+\int dx\int_{|t|_p\geqslant p}dt\,|t|_p^{-3/2}\tau_p^{-1}(2\varepsilon t)\chi(2\varepsilon t x^2)\eta^*\left(\frac{1+\sqrt{\varepsilon}t}{1-\sqrt{\varepsilon}t}\right)\\
&=I_1+I_2+I_3+I_4.\end{aligned} \tag{D2}$$

1. Evaluation of I_1

Some care is needed in evaluating I_1 since the x and t integration cannot be interchanged. The following shortcut leads, however, to the correct answer. Writing

$$|t|_p^{-1/2}\tau_p^{-1}(2t)=|t|_p^{-1/2}\tau_p(-2t)=\int dy\,\chi(-2ty^2),$$

one has

$$I_1=\int_{|t|_p\leqslant p^{-m}}dt\int dx\,dy\,\chi(-2t(y^2-\varepsilon x^2)). \tag{D3}$$

The Gaussian integral on the extension $\mathbb{Q}_p(\sqrt{\varepsilon})$ gives [Eqs. (A37) and (A40)]

$$I_1=\int_{|t|_p\leqslant p^{-m}}dt\,\frac{1}{2}(1+p^{-1})$$
$$\times\left\{\delta(t)+\frac{2}{1+p^{-1}}\frac{1}{|t|_p}\operatorname{sgn}_\varepsilon t\right\}. \tag{D4}$$

Regularizing the second term in Eq. (D4), we write

$$I_1(s)=\int_{|t|_p\leqslant p^{-m}}dt\,\frac{1}{2}(1+p^{-1})$$
$$\times\left\{\delta(t)+\frac{2}{1+p^{-1}}\frac{1}{|t|_p^s}\operatorname{sgn}_\varepsilon t\right\}$$
$$=\frac{1}{2}(1+p^{-1})+(1-p^{-1})(-p^{s-1})^m\frac{1}{1+p^{s-1}},$$

and taking the limit $s \to 1$ we obtain finally

$$I_1 = \tfrac{1}{2}\{1 + p^{-1} + (-)^m(1 - p^{-1})\} = \begin{cases} 1, \text{ for } m \text{ even}, & \text{(D5a)} \\ p^{-1} \text{ for } m \text{ odd}. & \text{(D5b)} \end{cases}$$

2. Evaluation of I_2

Interchanging the x and t integrations and performing the Gaussian integral gives

$$I_2 = \int_{p^{-m+1} < |t|_p < p^{-1}} dt\, |t|_p^{-1} \tau_p^{-1}(2t)\tau_p(2\varepsilon t)\eta^* \left(\frac{1+\sqrt{\varepsilon}t}{1-\sqrt{\varepsilon}t}\right). \tag{D6}$$

The identity $\tau_p^{-1}(2t)\tau_p(2\varepsilon t) = \mathrm{sgn}_\varepsilon t = (-1)^{\mathrm{ord}\, t}$ [Eq. (A31)] allows us to rewrite Eq. (D6) in the form

$$I_2 = \sum_{k=1}^{m-1} (-1)^k p^k \int_{|t|_p = p^{-k}} dt\, \eta^* \left(\frac{1+\sqrt{\varepsilon}t}{1-\sqrt{\varepsilon}t}\right). \tag{D7}$$

From the discussion of Sec. IV, it is clear that an η character of rank m, which essentially truncates its argument mod p^m, only depends on the digits $t_0,\ldots,t_{m-k-1}$ of

$$t = p^k(t_0 + t_1 p + \cdots).$$

Hence

$$\begin{aligned} \int_{|t|_p = p^{-k}} dt\, \eta^* \left(\frac{1+\sqrt{\varepsilon}t}{1-\sqrt{\varepsilon}t}\right) &= p^{-m} \sum_{t \in p^k Z_p^\times / p^m Z_p} \eta^* \left(\frac{1+\sqrt{\varepsilon}t}{1-\sqrt{\varepsilon}t}\right) \\ &= p^{-m} \sum_{\substack{(c,s) \bmod p^m \\ |s|_p = p^{-k} \\ c = 1 + \cdots}} \eta^*(c + \sqrt{\varepsilon}s) \\ &= p^{-m} \left\{ \sum_{\substack{(c,s) \bmod p^m \\ |s|_p < p^{-k} \\ c = 1 + \cdots}} \eta^* - \sum_{\substack{(c,s) \bmod p^m \\ |s|_p < p^{-k-1} \\ c = 1 + \cdots}} \eta^* \right\}. \end{aligned} \tag{D8}$$

The first sum runs over $C_\varepsilon^k \pmod{p^m} = C_\varepsilon^k / C_\varepsilon^m$, which is a (nontrivial) subgroup of $C_\varepsilon^1 / C_\varepsilon^m$ for all k in the interval $1 \leqslant k \leqslant m-1$. Hence it vanishes since summing the values of a character over a nontrivial subgroup gives zero.

The second sum in Eq. (D8) runs over $C_\varepsilon^{k+1}/C_\varepsilon^m$. By the same argument it will vanish unless $k = m-1$. Thus

$$\int_{|t|_p = p^{-k}} dt\, \eta^* \left(\frac{1+\sqrt{\varepsilon}t}{1-\sqrt{\varepsilon}t}\right) = -p^{-m}\delta_{k,m-1}, \tag{D9}$$

and finally

$$I_2 = (-)^{m-1}p^{m-1}(-p^{-m}) = (-)^m/p. \tag{D10}$$

For $m = 0$ or 1, this term is, of course, absent.

3. Evaluation I_3

The Gaussian integral on x yields

$$I_3 = \int_{|t|_p = 1} dt\, \eta^* \left(\frac{1+\sqrt{\varepsilon}t}{1-\sqrt{\varepsilon}t}\right). \tag{D11}$$

According to the factorization given by Eq. (4.35) we write, for $t = t_0 + t_1 \cdot p + \cdots$

$$\frac{1+\sqrt{\varepsilon}t}{1-\sqrt{\varepsilon}t} = \frac{1+\sqrt{\varepsilon}u}{1-\sqrt{\varepsilon}u} \cdot \frac{1+\sqrt{\varepsilon}t'}{1-\sqrt{\varepsilon}t'},$$

where $(1+\sqrt{\varepsilon}u)/(1-\sqrt{\varepsilon}u)$ belongs to $Z_{p+1}(|u|_p = 1)$ and the first digit u_0 of u equals t_0, while $(1+\sqrt{\varepsilon}t')/(1-\sqrt{\varepsilon}t')$ belongs to $C_\varepsilon^1(|t'|_p \leqslant p^{-1})$. Changing variables,

$$t = \frac{u + t'}{1 + \varepsilon u t'}$$

and using Eq. (A42) ($dt = dt'$) leads to ($m \geqslant 1$)

$$I_3 = \sum_{u_0 = 1}^{p-1} \eta^* \left(\frac{1+\sqrt{\varepsilon}u}{1-\sqrt{\varepsilon}u}\right) \int_{|t'|_p \leqslant p^{-1}} dt'\, \eta^* \left(\frac{1+\sqrt{\varepsilon}t'}{1-\sqrt{\varepsilon}t'}\right). \tag{D12}$$

By the same argument as in subsection 2, the integral over t' vanishes except for a character of rank 1, for which its value is $1/p$. Thus

$$I_3 = \delta_{m,1} \sum_{u_0 = 1}^{p-1} \frac{1}{p} \eta^* \left(\frac{1+\sqrt{\varepsilon}u}{1-\sqrt{\varepsilon}u}\right). \tag{D13}$$

Note that the sum over u_0 runs over all elements of the group Z_{p+1} except $\{\pm 1\}$. Hence

$$\begin{aligned} I_3 &= -\frac{1}{p}(\eta^*(1) + \eta^*(-1))\delta_{m,1} \\ &= -\frac{1}{p}(1 + \eta^*(-1))\delta_{m,1}. \end{aligned} \tag{D14}$$

For $m = 0$, there is no I_3 term.

4. Evaluation of I_4

Once again, the Gaussian integral gives

$$I_4 = \int_{|t|_p > p} dt\, |t|_p^{-2} \eta^* \left(\frac{1+\sqrt{\varepsilon}t}{1-\sqrt{\varepsilon}t}\right). \tag{D15}$$

Under the change of variables $t' = 1/\varepsilon t$, the character η^* becomes

$$\eta^* \left(\frac{1+1/\sqrt{\varepsilon}t'}{1-1/\sqrt{\varepsilon}t'}\right) = \eta^*(-1)\eta^* \left(\frac{1+\sqrt{\varepsilon}t'}{1-\sqrt{\varepsilon}t'}\right).$$

Hence, since $dt' = dt/|t|^2$,

$$I_4 = \eta^*(-1) \int_{|t'|_p \leqslant p^{-1}} dt'\, \eta^* \left(\frac{1+\sqrt{\varepsilon}t'}{1-\sqrt{\varepsilon}t'}\right).$$

This integral was calculated in subsection 3, giving the final result

$$I_4 = \begin{cases} 0, & m \geqslant 2, \\ (1/p)\eta^*(-1), & m = 1, \\ 1/p, & m = 0. \end{cases}$$

Adding the four contributions to the degeneracy $d(\eta)$ of an η character gives, at last,

$$d(\eta) = \begin{cases} 1, & \text{if } \eta \text{ is an even rank character}, \\ 0, & \text{if } \eta \text{ is an odd rank character}. \end{cases}$$

Similar calculations give the other entries in Tables II and III.

APPENDIX E: COMPUTATION OF WAVE FUNCTIONS

This Appendix deals with the computation of a class of wave functions $\psi_\eta(x)$ for $p\geqslant 5$, $\tau = p\alpha$, when η is an even character of odd rank, $2\nu+1$. Equation (6.14) can be rewritten as

$$\psi_\eta(x) = \psi_\eta^{-1}(0)\tfrac{1}{2}\int_{Q_p} dt\,\frac{\tau_p^{-1}(2t(1+\tau t^2))}{|t(1-\tau t^2)|_p^{1/2}} \times\chi\Big(\frac{(1+\tau t^2)}{2t}x^2\Big)\eta^*\Big(\frac{1+\sqrt{\tau}t}{1-\sqrt{\tau}t}\Big). \tag{E1}$$

We split this integral in two parts: namely, $|t|_p\leqslant 1$ and $|t|_p>1$ (which correspond to the two sheets of the group C_τ [Eq. (4.29)]). For $|t|_p>1$, we change variables $t\to t'=1/\tau t$. This brings us back in the region $|t'|_p\leqslant 1$. Recalling that

$$\eta^*\Big(\frac{1+1/\sqrt{\tau}t'}{1-1/\sqrt{\tau}t'}\Big) = \eta^*(-1)\eta^*\Big(\frac{1+\sqrt{\tau}t'}{1-\sqrt{\tau}t'}\Big),$$

we obtain, with $v=2t/(1+\tau t^2)$ ($dv=dt$ if $|t|_p\leqslant 1$),

$$\psi_\eta(x) = \psi_\eta^{-1}(0)\int_{|v|_p\leqslant 1} dv|v|_p^{-1/2}\tau_p^{-1}(v) \times\chi\Big(\frac{x^2}{v}\Big)\eta^*\Big(\Big[\frac{1+\sqrt{\tau v}}{1-\sqrt{\tau v}}\Big]^{1/2}\Big). \tag{E2}$$

For definiteness we take the generator u of C_τ^1 (see Table II) as $u=\frac{1}{2}+O(p)$, with $\varphi_\tau(u)=1$, where φ_τ is the additive parameter on C_τ^1 defined by Eq. (6.17). Hence

$$\eta^*\Big(\Big[\frac{1+\sqrt{\tau v}}{1-\sqrt{\tau v}}\Big]^{1/2}\Big) = \chi\Big(-\frac{b\varphi_\tau(v)}{2p^{2\nu}}\Big). \tag{E3}$$

To simplify further note that, for fixed x, all factors in Eq. (E2) depend on a finite number of digits of v. Moreover, for a given value of ord v, if a factor in Eq. (E2) depends on more digits of v than the others, the sum over the *last* of these digits always gives 0. This yields

for $k=\operatorname{ord} x\geqslant\nu$,

$$\psi_\eta(x) = \psi_\eta^{-1}(0)\int_{|v|_p<p^{-2\nu+1}} dv|v|_p^{-1/2}\tau_p^{-1}(v) \times\chi\Big(\frac{-b\varphi_\tau(v)}{2p^{2\nu}}\Big); \tag{E4a}$$

for $-\nu\leqslant k<\nu$,

$$\psi_\eta(x) = \psi_\eta^{-1}(0)\int_{|v|_p=p^{-\nu-k}} dv|v|_p^{-1/2}\tau_p^{-1}(v) \times\chi\Big(\frac{x^2}{v}\Big)\chi\Big(\frac{-b\varphi_\tau(v)}{2p^{2\nu}}\Big); \tag{E4b}$$

for $k<-\nu$, $\psi_\eta(x)=0$. (E4c)

The first integral is

$$\psi_\eta(x) = \psi_\eta^{-1}(0)2p^{-\nu} = \psi_\eta(0), \tag{E5}$$

which gives Eq. (6.15a).

For the integral Eq. (E4b) we change variables (remember that $-b$ is a square!),

$$v = p^\nu\sqrt{x^2/-b}\,y, \tag{E6}$$

and, using the identity

$$\varphi_\tau(av) = a\varphi_{\tau a^2}(v), \tag{E7}$$

we obtain

$$\psi_\eta(x) = \psi_\eta^{-1}(0)p^{-(\nu+k)/2}\int_{|y|_p=1} dy \times\chi\Big(\frac{G(y)}{p^{\nu-k}}\Big)\tau_p^{-1}\Big(p^\nu\sqrt{\frac{x^2}{-b}}\,y\Big). \tag{E8}$$

The function $G(y)$ is given by

$$G(y) = p^{-k}\sqrt{-bx^2}\,(\tfrac{1}{2}\varphi_{\tau'}(y)+y^{-1}), \tag{E9}$$

with

$$\tau' = \tau p^{2\nu}(-x^2/b). \tag{E10}$$

These two equations immediately imply Eq. (6.21). The integral Eq. (E8) is a sum over the residues modulo $p^{\nu-k}$ prime to p since, from Eq. (E9), $|y|_p=1$ implies $|G(y)|_p\leqslant 1$. Thus

$$\psi_\eta(x) = \psi_\eta^{-1}(0)p^{(k-3\nu)/2}\sum_{\substack{y=1\\ p\nmid y}}^{p^{\nu-k}-1}\chi\Big(\frac{G(y)}{p^{\nu-k}}\Big) \times\tau_p^{-1}\Big(p^\nu\sqrt{\frac{x^2}{-b}}\,y\Big). \tag{E11}$$

If $\nu-k=1$, Eq. (E11) is a Gauss sum mod p, which, to the best of our knowledge, cannot be expressed in closed form.[18] This gives Eq. (6.15b). If $\nu-k\neq 1$, we can apply Odoni's method[17] of evaluating the usual Gauss sum mod p^n ($n>1$). The heart of this method is to break up the sum over y in Eq. (E11) as a double sum over u and w,

$$y=u+p^r w,\quad 1\leqslant u<p^r,\quad p\nmid u,\quad 1\leqslant w\leqslant p^{\nu-k-r}, \tag{E12}$$

and then to expand the function $G(y)$ around u,

$$G(u+p^r w) = G(u)+p^r wG'(u)+p^{2r}\frac{w^2}{2}G''(u)+\cdots. \tag{E13}$$

One finds

$$G'(u) = \frac{1}{1-\tau' u^2}-\frac{1}{u^2},\quad\text{with } |G'(u)|_p\leqslant 1, \tag{E14}$$

and

$$G''(u) = \frac{2\tau' u}{(1-\tau' u^2)^2}+\frac{2}{u^3},\quad\text{with } |G''(u)|_p=1. \tag{E15}$$

To proceed let us take $\nu-k$ even ($\geqslant 4$) and Eq. (E12) with $r=(\nu-k)/2$. In this case the factor τ_p^{-1} is equal to 1 and Eq. (E11) becomes

$$\psi_\eta(x) = \psi_\eta^{-1}(0)p^{(k-3\nu)/2} \times\sum_{\substack{u=1\\ p\nmid u}}^{p^r-1}\chi\Big(\frac{G(u)}{p^{\nu-k}}\Big)\sum_{w=1}^{p^r}\chi\Big(\frac{wG'(u)}{p^r}\Big). \tag{E16}$$

The inner sum vanishes except when $G'(u)=0\ (\text{mod}\ p^r)$. This occurs only when u equals y^* mod p^r, with $G'(y^*)=0$, i.e., $y^{*2}=1/(1+\tau')$, and this immediately leads to Eq. (6.15c).

When $\nu-k$ is odd, one takes $r=(\nu-k+1)/2$ and proceeds along similar lines.

[1]See, e.g., C. N. Yang, in *Schrödinger, Centenary Celebration of a Polymath* (Cambridge U.P., Cambridge, 1987).

[2]For the theorem of Koval'skii–Pontryagin, see I. M. Gel'fand, M. I. Graev, and I. I. Pyatetskii–Shapiro, *Representation Theory and Automorphic Functions* (Saunders, London, 1966), p. 125.

[3]See all of Ref. 2.

[4]N. Koblitz, *p-adic Numbers, p-adic Analysis and Zeta-functions* (Springer, Berlin, 1977); K. Mahler, *p-adic Numbers and their Functions* (Cambridge U.P., Cambridge, 1973).

[5]For a review, see R. Rammal, G. Toulouse, and M. A. Virasoro, Rev. Mod. Phys. **58**, 765 (1986).

[6]I. V. Volovich, Class. Quantum Gravit. **4**, L83 (1987); I. Ya Arefeva and I. V. Volovich, "Strings, gravity and *p*-adic space-time," Steklov preprint; I. V. Volovich, Theor. Math. Phys. **71**, 574 (1987).

[7]P. Freund and M. Olson, Nucl. Phys. B **297**, 86 (1988); V. S. Vladimirov and I. V. Volovich, "P-adic quantum mechanics," Steklov preprint, 1988.

[8]C. Alacoque, Ph. Ruelle, E. Thiran, D. Verstegen, and J. Weyers, Phys. Lett. B **211**, 59 (1988).

[9]B. Grossman, Rockfeller Univ. preprint DOE/ER/40325-8-TASK B, 1987; S. Ben-Menahem, "p-adic Iterations," preprint TAUP 1627-88, 1988; E. Thiran, D. Verstegen, and J. Weyers, "*p*-adic dynamics," J. Stat. Phys. **54**, 893 (1989); M. D. Missarov, "Brownian motion with *p*-adic parameter," CPT preprint (Marseille) 88/P 2151, 1988.

[10]J. H. Hannay and M. V. Berry, Physica D **1**, 267 (1980); Y. Nambu, "Field theory of Galois' fields," in E.S. Fradkin Festschrift; Y. Meurice, Argonne National Laboratory preprint ANL-HEP-PR-87-114, 1987.

[11]P. Freund and M. Olson, Phys. Lett. B **199**, 186 (1987); P. Freund and E. Witten, *ibid.* **199**, 191 (1987); L. Brekke, P. Freund, M. Olson, and E. Witten, Nucl. Phys. B **302**, 365 (1988); J. L. Gervais, Phys. Lett. B **201**, 306 (1988); E. Marinari and G. Parisi, *ibid.* **203**, 52 (1988); Z. Hlousek and D. Spector, Cornell University preprint CLNS 88/832, 1988; I. Ya Arefeva, B. G. Dragovic, and I. V. Volovich, CERN preprint TH 5076/88, 1988; Phys. Lett. B **209**, 445 (1988); P. Frampton and Y. Okada, Phys. Rev. Lett. **60**, 484 (1988); Phys. Rev. D **37**, 3077 (1988); I. V. Volovich, Lett. Math. Phys. **16**, 61 (1988); B. L. Spokoiny, Phys. Lett. B **207**, 401 (1988); G. Parisi, Mod. Phys. Lett. A **3**, 369 (1988); Z. Ryzak, Phys. Lett. B **208**, 411 (1988); R. B. Zhang, *ibid.* **209**, 229 (1988); P. Frampton, Y. Okada, and M. R. Ubriaco, *ibid.* **213**, 260 (1988); Z. Hlousek and D. Spector, Phys. Lett. B **214**, 19 (1988).

[12]J. P. Serre, in *Proceedings of the International Summer School on Modular Functions*, Antwerp, 1972, *Lecture Notes in Mathematics*, Vol. 350 (Springer, Berlin, 1973). See, also, the references in S. Lang, *Introduction to Modular Forms* (Springer, Berlin, 1976).

[13]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1950).

[14]S. Lang, $SL_2(\mathbb{R})$ (Addison–Wesley, Reading, MA, 1975).

[15]Z. I. Borevitch and I. R. Chafarevitch, *Théorie des nombres* (Gauthier-Villars, Paris, 1967).

[16]For a Weyl formulation on finite fields $\mathbf{F}_p$, see Y. Meurice, lectures given at the Université Catholique de Louvain, 1987.

[17]R. Odoni, Bull. London Math. Soc. **5**, 325 (1973).

[18]B. Berndt and R. Evans, Bull. Am. Math. Soc. **5**, 107 (1981).

Dirac operators with a spherically symmetric δ-shell interaction

J. Dittrich, P. Exner, and P. Šeba
Nuclear Physics Institute, Czechoslovak Academy of Sciences, 250 68 Řež near Prague, Czechoslovakia and Laboratory of Theoretical Physics, Joint Institute for Nuclear Research, 141 980 Dubna, Union of Soviet Socialist Republics

(Received 28 April 1989; accepted for publication 12 July 1989)

Dirac operators with a contact interaction supported by a sphere are studied restricting attention to the operators that are rotationally and space-reflection symmetric. The partial wave operators are constructed using the self-adjoint extension theory, a particular attention being paid to those among them that can be interpreted as δ-function shells of scalar and vector nature. The class of interactions for which the sphere becomes impenetrable is specified and spectral properties of the obtained Hamiltonians are discussed.

I. INTRODUCTION

Dirac Hamiltonians with external potentials have attracted a lot of attention recently.[1,2] Of course, they require a fixed inertial frame and represent an approximative description of the true relativistic two-particle dynamics only, but nevertheless they can provide us with various useful and physically interesting models. Unfortunately, the number of situations when a Dirac-operator model is exactly solvable is very low compared to the nonrelativistic quantum mechanics.[3]

In the nonrelativistic case, many new solvable models have appeared recently as a result of extensive investigation of point and contact interaction phenomena—cf. the monograph[4] for summary and further references. One of these models concerns the three-dimensional Schrödinger operator with the interaction formally given by the δ-shell potential:

$$g\delta(r-R), \quad R=\text{const}, \tag{1.1}$$

cf. Refs. 5–12 and Refs. 13–15 for some generalizations. The aim of the present paper is to investigate the Dirac operator with this sort of interaction, and to add, thereby, a new item to the short list of exactly solvable problems of relativistic quantum mechanics. We are going to construct all rotationally and space reflection invariant contact interactions supported by the sphere, and to specify those among them that correspond to a mixture of electrostatic and Lorentz scalar δ-shell potentials with coupling constants of g_v and g_s, respectively:

$$g_v\delta(r-R)+g_s\beta\delta(r-R). \tag{1.2}$$

In contrast to the Schrödinger case, such a shell can confine a particle within it at finite values of the coupling constants provided its scalar component is strong enough: We will show that it happens if

$$g_v^2-g_s^2+4=0. \tag{1.3}$$

Other properties of the corresponding Dirac operators, in particular their spectra, will also be discussed. In a sequel to this paper, we are going to discuss Dirac operators with a δ-shell plus Coulomb potential, the nonrelativistic limit, and the approximation of the δ-shell interaction by short-range potentials.

II. PARTIAL WAVE DECOMPOSITION

Our construction starts from the Dirac Hamiltonian on the Hilbert space $\mathscr{H}=L^2(\mathbb{R}^3)\otimes\mathbb{C}^4$ defined by

$$H_D:\ H_D\psi=-i\boldsymbol{\alpha}\nabla\psi+\beta m\psi \tag{2.1}$$

with the domain

$$D(H_D)=H^{1,2}(\mathbb{R}^3)\otimes\mathbb{C}^4.$$

The Dirac matrices are taken as

$$\boldsymbol{\alpha}=\begin{pmatrix}0 & \boldsymbol{\sigma}\\ \boldsymbol{\sigma} & 0\end{pmatrix},\quad \beta=\begin{pmatrix}I & 0\\ 0 & -I\end{pmatrix}.$$

For the definition of other quantities related to the Dirac equation (spherical spinors, etc.), we follow the convention of Ref. 16. The operator H_D is self-adjoint and $C_0^\infty(\mathbb{R}^3)\otimes\mathbb{C}^4$ is its core. Moreover, $C_0^\infty(\mathbb{R}^3\setminus\{0\})\otimes\mathbb{C}^4$ is also a core of H_D as can be seen from its density in $D(H_D)$ in the $H^{1,2}$ norm. It illustrates the known fact that there is no nontrivial point interaction for a three-dimensional Dirac operator.[17]

Our construction of the δ-shell interaction proceeds in a usual way: One restricts the starting operator to a set of functions with supports disjoint with the support of the interaction, and constructs self-adjoint extensions of the obtained operator. Since, in our case, the support is the sphere $S_R=\{\mathbf{x}\in\mathbb{R}^3\colon |\mathbf{x}|=R\}$, where R is a given positive number, we are interested in the operator

$$H_1:=H_D|C_0^\infty(\mathbb{R}^3\setminus S_R)\otimes\mathbb{C}^4. \tag{2.2}$$

For technical reasons, it is useful to consider also its restriction

$$H_0:=H_D|C_0^\infty(\mathbb{R}^3\setminus(S_R\cup\{0\}))\otimes\mathbb{C}^4. \tag{2.3}$$

Since $\overline{H}_0=\overline{H}_1$, as one can check in the above-mentioned way, the two operators have identical families of self-adjoint extensions.

The operators (2.2) and (2.3) have infinite deficiency indices and, hence, a vast family of extensions. In this paper, we restrict our attention to those extensions that are *rotationally invariant*; it will give us a possibility to reduce the problem to the analysis of ordinary differential operators. It does not mean, however, that other self-adjoint extensions are not physically attractive. On the contrary, one should expect the existence of interesting extensions that are rotationally noninvariant and, at the same time, specified by lo-

 0022-2488/89/122875-08$02.50

cal boundary conditions. These problems will be discussed elsewhere. In addition to the rotational symmetry requirement, we shall consider only *reflection-symmetric* extensions, i.e., our group of symmetry will be $SU(2) \otimes \mathbb{Z}_2$. We use the Hermitian operator of space reflection $(U(P)\psi)(\mathbf{x}) = \beta\psi(-\mathbf{x})$. The anti-Hermitian choice of $iU(P)$ leads to the same Eqs. (2.4), and, therefore, to the same results.

These requirements mean that there is a single-valued unitary representation U of $SU(2) \otimes \mathbb{Z}_2$ such that an arbitrary self-adjoint extension H of H_0 from the considered class fulfills

$$U(R)HU(R)^{-1} = H \tag{2.4}$$

for any $R \in SU(2) \otimes \mathbb{Z}_2$. One can decompose the state Hilbert space into orthogonal sum of subspaces referring to the total angular momentum j, its third component μ, and the parity $(-1)^l$ as

$$\mathscr{H} = \bigoplus_{j=1/2}^{\infty} \bigoplus_{l=j-1/2}^{j+1/2} \bigoplus_{\mu=-j}^{j} \mathscr{H}_{jl\mu}, \tag{2.5a}$$

with

$$\mathscr{H}_{jl\mu} = \left\{ \psi \in \mathscr{H}:\ \psi(r\mathbf{n}) = \begin{pmatrix} f(r)\Omega_{jl\mu}(\mathbf{n}) \\ g(r)\Omega_{jl'\mu}(\mathbf{n}) \end{pmatrix}; \right.$$

$$\left. f,g \in L^2(\mathbb{R}_+, r^2\,dr) \right\}, \tag{2.5b}$$

where $\Omega_{jl\mu}$ are the spherical spinors[16] and $l' = j \mp 1/2$ for $l = j \pm 1/2$. It follows from (2.4) that H commutes with all functions of the operators $U(R)$, in particular, with the projection corresponding to the representation of $SU(2) \otimes \mathbb{Z}_2$ with a given j, μ and parity. Hence, we have the decomposition

$$H = \bigoplus_{j=1/2}^{\infty} \bigoplus_{l=j-1/2}^{j+1/2} \bigoplus_{\mu=-j}^{j} H_{jl\mu} \tag{2.6}$$

where the "component operators" $H_{jl\mu}$ are self-adjoint with the domains $D(H_{jl\mu}) = D(H) \cap \mathscr{H}_{jl\mu}$.

In each subspace $\mathscr{H}_{jl\mu}$ one can separate the radial part of the operator $H_{jl\mu}$. To this aim, we introduce the isomorphisms

$$U_{jl\mu}:\ \mathscr{H}_{jl\mu} \to \hat{\mathscr{H}} = L^2(\mathbb{R}_+) \otimes \mathbb{C}^2$$

by

$$(U_{jl\mu}\psi)(r) = \begin{pmatrix} rf(r) \\ (-1)^{j-l-1/2} rg(r) \end{pmatrix}, \tag{2.7}$$

where f,g are related to ψ as in (2.5b). One can easily check the inclusion

$$H_0 \subset \bigoplus_{j=1/2}^{\infty} \bigoplus_{l=j-1/2}^{j+1/2} \bigoplus_{\mu=-j}^{j} H_{jl\mu}^{(0)} \tag{2.8}$$

where the operators $H_{jl\mu}^{(0)}$ are equal to $U_{jl\mu}^{-1} \hat{H}_{jl}^{(0)} U_{jl\mu}$, the last named operator being defined by

$$\hat{H}_{jl}^{(0)} := \begin{pmatrix} m; & -d/dr + \kappa_{jl}/r \\ d/dr + \kappa_{jl}/r; & -m \end{pmatrix} \tag{2.9a}$$

on $D(\hat{H}_{jl}^{(0)}) = C_0^\infty((0,R) \cup (R,\infty)) \otimes \mathbb{C}^2$ with

$$\kappa_{jl} = (-1)^{j-l+1/2}(j+1/2) \tag{2.9b}$$

(we note that there the quantities with carets refer to the two-component space $\hat{\mathscr{H}}$). Now we want to prove that a decomposition analogous to (2.8) holds for any rotation-invariant self-adjoint operators on $\mathscr{H}$.

Proposition 2.1: For a self-adjoint operator H in $\mathscr{H}$, the equality (2.4) holds for all $R \in SU(2) \otimes \mathbb{Z}_2$ iff

$$H = \bigoplus_{j=1/2}^{\infty} \bigoplus_{l=j-1/2}^{j+1/2} \bigoplus_{\mu=-j}^{j} U_{jl\mu}^{-1} \hat{H}_{jl} U_{jl\mu}, \tag{2.10}$$

where $\hat{H}_{jl}$ is a self-adjoint operator in $\hat{\mathscr{H}}$ independent of μ.

Proof: The sufficient condition is trivial. The requirement (2.4) implies the decomposition (2.6), and one defines naturally $\hat{H}_{jl\mu} = U_{jl\mu} H_{jl\mu} U_{jl\mu}^{-1}$, so it is only necessary to check its independence of μ.

Let us show first that $D(\hat{H}_{jl\mu}) = D(\hat{H}_{jl\mu'})$. The operator $U_{jl\mu}^{-1}$ maps $D(\hat{H}_{jl\mu})$ onto $D(H_{jl\mu})$ and the vector

$$U(R)\begin{pmatrix} r^{-1}f\Omega_{jl\mu} \\ (-1)^{j-l-1/2} r^{-1} g\Omega_{jl'\mu} \end{pmatrix}$$

$$= \sum_{\mu'=-j}^{j} \mathscr{D}_{\mu'\mu}^{(j)}(R) \begin{pmatrix} r^{-1}f\Omega_{jl\mu'} \\ (-1)^{j-l-1/2} r^{-1} g\Omega_{jl'\mu'} \end{pmatrix}$$

belongs to $D(H)$ for any $\binom{f}{g} \in D(\hat{H}_{jl\mu})$ and $R \in SU(2)$ according to (2.4). Since $\mathscr{H}_{jl\mu} \perp \mathscr{H}_{jl\mu'}$ for $\mu \neq \mu'$, each term on the rhs must belong to the corresponding $D(H_{jl\mu'})$. For any given μ,μ', one always finds R such that $\mathscr{D}_{\mu'\mu}^{(j)}(R) \neq 0$ due to the Burnside theorem.[18] Then, applying $U_{jl\mu'}$, we find that

$$\begin{pmatrix} f \\ g \end{pmatrix} \in D(\hat{H}_{jl\mu'})$$

for all $\mu' = -j,...,j$. The index $\mu = -j,...,j$ can be chosen arbitrarily so the equality of the domains is obtained.

Consider, now, a vector $\psi \in D(H)$ referring to fixed j,l, i.e.,

$$\psi(r\,\mathbf{n}) = \sum_{\mu=-j}^{j} \begin{pmatrix} r^{-1} f_{jl\mu}(r)\Omega_{jl\mu}(\mathbf{n}) \\ (-1)^{j-l-1/2} r^{-1} g_{jl'\mu}(r)\Omega_{jl'\mu}(\mathbf{n}) \end{pmatrix},$$

with

$$\hat{\psi}_{jl\mu} \equiv \begin{pmatrix} f_{jl\mu} \\ g_{jl'\mu} \end{pmatrix} \in \hat{D}_{jl} \equiv D(\hat{H}_{jl\mu})$$

for $\mu = -j,...,j$. Denoting

$$\hat{H}_{jl\mu}\begin{pmatrix} f \\ g \end{pmatrix} = \begin{pmatrix} \hat{H}_{jl\mu}^{(1)}(f,g) \\ \hat{H}_{jl\mu}^{(2)}(f,g) \end{pmatrix},$$

we can calculate easily

$$U(R)HU(R)^{-1}\psi$$

$$= \sum_{\mu=-j}^{j} \sum_{\nu,\sigma} \mathscr{D}_{\mu\nu}^{(j)}(R)\mathscr{D}_{\nu\sigma}^{(j)}(R^{-1})$$

$$\times \begin{pmatrix} r^{-1}\hat{H}_{jl\nu}^{(1)}(f_{jl\sigma}, g_{jl'\sigma})\Omega_{jl\mu} \\ (-1)^{j-l-1/2} r^{-1} \hat{H}_{jl\nu}^{(2)}(f_{jl\sigma}, g_{jl'\sigma})\Omega_{jl'\mu} \end{pmatrix}.$$

Since this should equal to $H\psi$, we get the relation

$$\sum_{\nu,\sigma} \mathscr{D}_{\mu\nu}^{(j)}(R)\mathscr{D}_{\nu\sigma}^{(j)}(R^{-1})\hat{H}_{jl\nu}\hat{\psi}_{jl\sigma} = \hat{H}_{jl\mu}\hat{\psi}_{jl\mu}$$

valid for any $\hat{\psi}_{jl\rho} \in \hat{D}_{jl}$, $\rho = -j,...,j$. Multiplying it by $\mathscr{D}_{\rho\mu}^{(j)}(R^{-1})$ and summing over μ, we obtain

$$\sum_{\sigma} \mathscr{D}^{(j)}_{\rho\sigma}(R^{-1})[\hat{H}_{jl\rho}\hat{\psi}_{jl\sigma} - \hat{H}_{jl\sigma}\hat{\psi}_{jl\sigma}] = 0$$

for each $R \in SU(2)$. Using the Burnside theorem again, we get

$$\hat{H}_{jl\rho}\hat{\psi}_{jl\sigma} = \hat{H}_{jl\sigma}\hat{\psi}_{jl\sigma}$$

for all $\rho,\sigma = -j,...,j$. ■

It follows now easily from the proved assertion and the decomposition (2.8) that in order to construct all rotationally and space-reflection invariant δ-shell interactions for the operator (2.1), one has to construct all self-adjoint extensions $\hat{H}_{jl}$ of $\hat{H}^{(0)}_{jl}$ in each partial wave subspace and insert them into the formula (2.10).

III. SELF-ADJOINT EXTENSIONS OF THE RADIAL OPERATORS

We have reduced the problem to analysis of ordinary differential operators corresponding to the formal expression

$$\tau = \begin{pmatrix} 0, & -1 \\ 1, & 0 \end{pmatrix} \frac{d}{dr} + \begin{pmatrix} m, & \varkappa_{jl}/r \\ \varkappa_{jl/r}, & -m \end{pmatrix}, \tag{3.1}$$

which can be handled by standard methods (e.g., that of Ref. 19, Chap. XIII), because the coefficient of the derivative is a constant and nonsingular matrix. The adjoint operator $\hat{H}^{(0)*}_{jl}$ to (2.9) acts as (3.1) on the domain $D(\hat{H}^{(0)*}_{jl})$ consisting of the functions $\hat{\psi} \in \hat{\mathscr{H}}$ which are absolutely continuous in $(0,R)$ and (R,∞) with $\tau\hat{\psi} \in \hat{\mathscr{H}}$. Since τ is real, the deficiency indices are equal and self-adjoint extensions of the operator (2.9) exist.

The deficiency indices fulfill $d \leqslant 4$, because they correspond to solutions of a two-component first-order differential equation in the intervals $(0,R)$ and (R,∞). In order to find them explicitly, consider the equation

$$(\tau - i)\varphi = 0, \tag{3.2}$$

whose solutions are obtained easily by analytic continuation of the well-known stationary solutions of the free radial Dirac equation to imaginary values of energy:

$$\varphi_z(r) = \begin{pmatrix} r^{1/2}Z_\nu(i(1+m^2)^{1/2}r) \\ (-1)^{j-l+1/2}((1+im)/(1-im))^{1/2}r^{1/2}Z_{\nu'}(i(1+m^2)^{1/2}r) \end{pmatrix}, \tag{3.3}$$

where $\nu = l + 1/2$, $\nu' = l' + 1/2$, and the cylindrical function Z_ν stands for J_ν or $H^{(1)}_\nu$. The first choice yields a solution that is square integrable in $(0,R)$ but not in (R,∞), while the reverse is true for the second case. Extending the square integrable solutions by zero in the other interval, we get $d = 2$.

One could use this explicit form of deficiency vectors for construction of the self-adjoint extensions via the von Neumann formulas, but this is not very practical. Instead, we are going to characterize the extensions by suitable boundary conditions. Since $H_D | C^\infty_0(\mathbb{R}^3 \setminus \{0\}) \otimes \mathbb{C}^4$ is essentially self-adjoint, there are no nontrivial boundary values at 0 and ∞.

Proposition 3.1: There is a complete set of four independent boundary values on $D(\hat{H}^{(0)*}_{jl})$, namely,

$$\psi \mapsto \psi(R_\pm) \equiv \lim_{r \to R_\pm} \psi(r)$$

(recall that ψ is a two-component vector).

Proof: Consider $\psi = \binom{f}{g} \in D(\hat{H}^{(0)*}_{jl})$. The functions f,g are absolutely continuous inside $(0,R)$ and (R,∞) and square integrable on $\mathbb{R}_+$ with $\tau\psi \in L^2(\mathbb{R}_+) \otimes \mathbb{C}^2$. Then, for instance, f' is square integrable in a left neighborhood of R and

$$|f(r) - f(s)| \leqslant |r-s|^{1/2} \left| \int_r^s |f'(t)|^2 \, dt \right|^{1/2},$$

so $\lim_{r \to R_-} f(r)$ exists and in the same way one checks the existence of the other three limits. Since we have $4 = 2d$ linearly independent boundary values, they form a complete set[19] for $\hat{H}^{(0)}_{jl}$. ■

Self-adjoint extensions of $\hat{H}^{(0)}_{jl}$ are restrictions of $\hat{H}^{(0)*}_{jl}$ to a subspace of $D(\hat{H}^{(0)*}_{jl})$ specified by a symmetric set of two linearly independent boundary conditions. We define the boundary form

$$F(\psi,\varphi;r) := \psi^+(r)\tau_0\varphi(r),$$

where $\tau_0 = \binom{0, \; -1}{1, \; 0}$. Integration by parts yield for any $\varphi,\psi \in D(\hat{H}^{(0)*}_{jl})$ the equality

$$\int_r^s \psi^+ \tau\varphi \, dt - \int_r^s (\tau\psi)^+ \varphi \, dt = F(\psi,\varphi;s) - F(\psi,\varphi;r)$$

provided R is not contained in the interval (r,s). Since the integrals on the lhs converge in any subinterval of $\mathbb{R}_+$, one can establish existence of the limits of $F(\psi,\varphi;\cdot)$ at the points $0, R_-, R_+, \infty$ similarly as in Proposition 3.1. Furthermore,

$$\lim_{r \to 0_+} F(\psi,\varphi;r) = \lim_{r \to \infty} F(\psi,\varphi;r) = 0,$$

for any $\varphi,\psi \in D(\hat{H}^{(0)*}_{jl})$ since, otherwise, one could define an additional independent boundary value in contradiction to Proposition 3.1. Hence we get

$$(\psi,\hat{H}^{(0)*}_{jl}\varphi) - (\hat{H}^{(0)*}_{jl}\psi,\varphi) = F(\psi,\varphi;R_-) - F(\psi,\varphi;R_+), \tag{3.4}$$

for any $\varphi,\psi \in D(\hat{H}^{(0)*}_{jl})$ and we have to choose those boundary conditions for which the rhs of (3.4) vanishes.

Theorem 3.2: Any self-adjoint extension $\hat{H}_{jl}$ of $\hat{H}^{(0)}_{jl}$ in $\hat{\mathscr{H}}$ act as $\hat{H}_{jl}\psi = \tau\psi$ for $\psi \in D(\hat{H}_{jl})$ where τ is given by (3.1) and $D(\hat{H}_{jl})$ consists of the functions $\psi \in L^2(\mathbb{R}_+) \otimes \mathbb{C}^2$, which are absolutely continuous inside $(0,R)$ and (R,∞), $\tau\psi \in L^2(\mathbb{R}_+) \otimes \mathbb{C}^2$, and satisfy the following boundary conditions:

$$\psi(R_-) = e^{i\alpha}A\psi(R_+) \tag{3.5a}$$

where $\alpha\in[0,\pi)$ and A is a real 2×2 matrix with $\det A = 1$, or

$$\begin{pmatrix} c_1, & c_2 \\ 0, & 0 \end{pmatrix}\psi(R_-) + \begin{pmatrix} 0, & 0 \\ d_1, & d_2 \end{pmatrix}\psi(R_+) = 0, \tag{3.5b}$$

where c_1,c_2,d_1,d_2 are real and both matrices are nonzero. Conversely, any operator of this form is a self-adjoint extention of $\widehat{H}_{jl}^{(0)}$ in $\mathscr{H}$.

Proof: It remains to check that (3.5) are all symmetric pairs of linearly independent boundary conditions. According to proposition 3.1, the general form of such boundary conditions is

$$C\psi(R_-) + D\psi(R_+) = 0, \tag{3.6}$$

where C,D are 2×2 matrices such that the 2×4 matrix (C,D) has rank 2. The symmetry conditions according to (3.4) read

$$\psi(R_-)^+\tau_0\varphi(R_-) - \psi(R_+)^+\tau_0\varphi(R_+) = 0, \tag{3.7}$$

for any $\varphi,\psi\in D(\widehat{H}_{jl}^{(0)*})$ satisfying (3.6). We distinguish the following three cases.

(i) C is nonsingular, then (3.6) can be written as $\psi(R_-) = B\psi(R_+)$ and substitution into (3.7) gives

$$\psi(R_+)^+(B^+\tau_0 B - \tau_0)\varphi(R_+) = 0.$$

Since this equation should hold for any $\psi(R_+)$, $\varphi(R_+)$, we get

$$B^+\tau_0 B = \tau_0. \tag{3.8}$$

So, $|\det B| = 1$, and B is nonsingular. A simple algebra shows that (3.8) is equivalent to (3.5a) with $B = \exp(i\alpha)A$.

(ii) D is nonsingular, then (3.6) can be written as $\psi(R_+) = \widetilde{B}\psi(R_-)$, where $\widetilde{B}$ is nonsingular due to (3.7), so this case reduces to the previous one.

(iii) Both C,D are of rank 1 but (C,D) has rank 2. Multiplying (3.6) by a suitable nonsingular matrix, we can write it as

$$\begin{pmatrix} c_1, & c_2 \\ 0, & 0 \end{pmatrix}\psi(R_-) + D_1\psi(R_+) = 0, \tag{3.9}$$

where at least one of the numbers c_1,c_2 is nonzero. Since D_1 is again a rank-1 matrix, one can write it as

$$D_1 = \begin{pmatrix} \lambda d_1, & \lambda d_2 \\ d_1, & d_2 \end{pmatrix}$$

with at least one of the numbers d_1,d_2 nonzero (the other possibility when only the first row is nonzero is excluded because the combined 2×4 matrix should be of rank 2). It is easy to see that (3.9) is in that case equivalent to (3.5b) or to

$$\begin{pmatrix} c_1, & c_2 \\ 0, & 0 \end{pmatrix}\psi(R_-) = \begin{pmatrix} 0, & 0 \\ d_1, & d_2 \end{pmatrix}\psi(R_+) = 0, \tag{3.5b$'$}$$

i.e., that the boundary conditions decouple in this case. The coefficients c_1,c_2,d_1,d_2 might be still complex. The condition (3.5b$'$), however, means that the two-dimensional complex vector $\psi(R_-)$ is for any $\psi\in D(\widehat{H}_{jl})$ orthogonal to $\binom{\bar{c}_1}{\bar{c}_2}$, i.e.,

$$\psi(R_-) = \alpha_-(\psi)\begin{pmatrix} c_2 \\ -c_1 \end{pmatrix}$$

and the corresponding expression holds for $\psi(R_+)$ in terms of d_1,d_2. Substituting it to the expression (3.7) and realizing that $\alpha_\pm(\psi),\alpha_\pm(\phi)$ may be arbitrary, we get $\operatorname{Im}\bar{c}_1c_2 = \operatorname{Im}\bar{d}_1d_2 = 0$. Thus c_1,c_2 and d_1,d_2 must have the same complex phases and can be chosen real. ■

Remark 3.3: In addition to the stated symmetry requirements, one may want the constructed Hamiltonians to be time-reversal invariant. The corresponding antiunitary operator T can be defined as in the free-particle case[16]

$$T\psi = \begin{pmatrix} \sigma_2, & 0 \\ 0 & \sigma_2 \end{pmatrix}K\psi,$$

where K means the complex conjugation. After the partial-wave decomposition, we see that H is time-reversal invariant iff $D(\widehat{H}_{jl})$ is invariant with respect to the complex conjugation for all j,l. The just proved theorem shows that this is the case when $\widehat{H}_{jl}$ are specified by the boundary conditions (3.5b) or by (3.5a) with $\alpha = 0$.

IV. δ SHELLS

As we have mentioned in the Introduction, we are interested primarily in the potentials (1.2), i.e., a combination of the scalar external field $g_s\delta(r-R)$ and the vector field described in the given reference frame by $(g_v\delta(r-R),\mathbf{0})$ with real coupling constants. In the radial Hamiltonians $\widehat{H}_{jl}$, this interaction corresponds to the formal potential

$$\begin{pmatrix} g_v + g_s, & 0 \\ 0, & g_v - g_s \end{pmatrix}\delta(r-R) \tag{4.1}$$

with g_v,g_s independent of j,l. More generally, one can consider the potential

$$G\delta(r-R), \tag{4.2}$$

where G is a 2×2 matrix. Our aim is now to specify the self-adjoint extensions $\widehat{H}_{jl}$ that can be associated with the formal Dirac operator with the potential (4.2). Suppose that ψ satisfies the equation

$$[\tau + G\delta(r-R)]\psi = E\psi$$

and the limits $\psi(R_\pm)$ exist. Integrating over $(R-\epsilon, R+\epsilon)$ and taking the limit $\epsilon\to 0_+$, we get

$$(1 - \tfrac{1}{2}\tau_0 G)\psi(R_+) - (1 + \tfrac{1}{2}\tau_0 G)\psi(R_-) = 0 \tag{4.3}$$

provided we have chosen the relation

$$\int_{R-\epsilon}^{R+\epsilon} \delta(r-R)\psi(r)dr = \frac{1}{2}\,(\psi(R_+) + \psi(R_-)) \tag{4.4}$$

as a definition of the lhs. Of course, only those matrices G are acceptable for which the boundary condition (4.3) is compatible with (3.5). As one expects, the following assertion is true.

Proposition 4.1: Boundary conditions (4.3) define a self-adjoint extension of $\widehat{H}_{jl}^{(0)}$ iff $G^+ = G$.

Proof: The matrix $(1 - \tau_0 G/2, 1 + \tau_0 G/2)$ has rank 2 since the sum of the submatrices is nonsingular. It remains to check that (4.3) implies (3.7) iff $G^+ = G$. We start with the necessary condition and distinguish four cases denoting $B = \tfrac{1}{2}\tau_0 G$.

(i) $1 - B$ is nonsingular, then, (4.3) is equivalent to

$\psi(R_+) = (1-B)^{-1}(1+B)\psi(R_-)$; substituting it into (3.7), we get, after a simple algebra, $G^+ = G$.

(ii) $1+B$ is nonsingular, then, the same procedure with the interchange $G \to -G$ can be used.

(iii) $1-B=0$ or $1+B=0$; in that case, $G = \mp 2\tau_0 \neq G^+$, but the condition (4.3) reads $\psi(R_\mp) = 0$, while $\psi(R_\pm)$ is arbitrary. Hence the lhs of (3.7) equals $\mp \psi(R_\pm)^+ \tau_0 \psi(R_\pm)$ so these boundary conditions cannot define a self-adjoint extension.

(iv) Both $1+B$ and $1-B$ have rank 1. As in the proof of the Theorem 3.2, one can find a nonsingular matrix V that converts them into nonzero matrices of the following form:

$$V(1+B) = \begin{pmatrix} c_1, & c_2 \\ 0, & 0 \end{pmatrix}, \quad V(1-B) = \begin{pmatrix} 0, & 0 \\ d_1, & d_2 \end{pmatrix};$$

one can express V and VB from there. Furthermore, V is nonsingular, so one can calculate B and

$$G = -2\tau_0 B = \frac{2}{c_2 d_1 - c_1 d_2} \begin{pmatrix} 2c_1 d_1, & c_1 d_2 + c_2 d_1 \\ c_1 d_2 + c_2 d_1, & 2c_2 d_2 \end{pmatrix}. \quad (4.5)$$

Since we can choose V so that the numbers c_1, c_2, d_1, d_2 are real, G is real symmetric, and, therefore, Hermitian.

On the contrary, assume $G^+ = G$ and let us prove that (4.3) defines a self-adjoint extension. For the cases (i) and (ii), we have done it already; the case (iii) does not occur. It remains to complete the proof for the case (iv). Any Hermitian G is of the form

$$G = \begin{pmatrix} a, & b \\ \bar{b}, & c \end{pmatrix},$$

with a,c real. It gives

$$1+B = \begin{pmatrix} 1-\bar{b}/2, & -c/2 \\ a/2, & 1+b/2 \end{pmatrix},$$

$$1-B = \begin{pmatrix} 1+\bar{b}/2, & c/2 \\ -a/2, & 1-b/2 \end{pmatrix}.$$

Since $\det(1+B) = \det(1-B) = 0$, b must be real and $b^2 = 4 + ac$. For any a,b,c satisfying this restriction, one can choose c_1, c_2, d_1, d_2 so that G is expressed in the form (4.5), e.g., by taking $c_1 = d_1 = 1$ for $a \neq 0$ and $c_1 = 0$ or $d_1 = 0$ for $a = 0$. ∎

Let us return now to the physically interesting case (4.1). The corresponding matrix G is Hermitian and the boundary conditions (4.3) read

$$\begin{pmatrix} 1, & (g_v - g_s)/2 \\ -(g_v + g_s)/2, & 1 \end{pmatrix} \psi(R_+) - \begin{pmatrix} 1, & (g_s - g_v)/2 \\ (g_v + g_s)/2, & 1 \end{pmatrix} \psi(R_-) = 0. \quad (4.6)$$

It is clear that they can be cast into the form (3.5a) iff $g_v^2 - g_s^2 + 4 \neq 0$; otherwise they belong to the type (3.5b). Remark 3.3 shows the corresponding operators, as well as the more general Hamiltonians referring to the δ-shell interaction (4.2) with real G, are time-reversal invariant. The interactions (4.2) do not cover, of course, the class of all self-adjoint extensions $\hat{H}_{jl}$ described by Theorem 3.2; a possible interpretation of the remaining ones is discussed in the Appendix.

V. CONFINEMENT

In some cases, the contact interaction on the sphere may separate the two spatial regions fully, i.e., the particle under consideration is either confined in the ball $B_R = \{\mathbf{x} \in \mathbb{R}^3 : |\mathbf{x}| \leq R\}$ or lives outside B_R and cannot enter it. In other words, the sphere S_R is impenetrable for the particles.

Let us denote $\mathscr{H}_R = \{\psi \in \mathscr{H} : \operatorname{supp} \psi \subset B_R\}$; we are interested in the situation when $\mathscr{H}_R$ is invariant under $\exp(-iHt)$ for all $t \in \mathbb{R}$ or equivalently $E_R H \subset H E_R$, where E_R is the projection onto $\mathscr{H}_R$ in $\mathscr{H}$

$$(E_R \psi)(\mathbf{x}) = \Theta(R - |\mathbf{x}|)\psi(\mathbf{x}).$$

In the spherically symmetric case, it is further equivalent to

$$\hat{E}_R D(\hat{H}_{jl}) \subset D(\hat{H}_{jl}),$$

for all j,l, where $\hat{E}_R$ is the projection onto $L^2(0,R) \otimes \mathbb{C}^2$ in $\hat{\mathscr{H}}$. Combining the last requirements with Theorem 3.2, we arrive at the following conclusion.

Proposition 5.1: Let H be a rotationally and space-reflection symmetric Dirac operator with contact interaction on the sphere S_R, then S_R is impenetrable for the particles iff the corresponding partial-wave operators $\hat{H}_{jl}$ are defined by the boundary conditions (3.5b) for all j,l.

As an example, consider again the physically interesting case of the interaction (4.1) corresponding to the boundary conditions (4.6). The observation made at the end of the previous section shows that the sphere S_R is impenetrable in this case iff

$$g_v^2 - g_s^2 + 4 = 0. \quad (5.1)$$

Notice that presence of the scalar component is essential here,

$$|g_s| = (g_v^2 + 4)^{1/2} \geq 2.$$

In particular, a purely scalar δ shell confines the particles iff $g_s = \pm 2$. We remark also that the relation (5.1) has been found recently (on a heuristic level) as the impenetrability condition for a δ-shaped separable potential in one-dimensional Dirac operator.[20]

VI. SPECTRAL PROPERTIES

A. Point spectrum

In order to solve the eigenvalue problem $\hat{H}_{jl}\psi = \lambda\psi$, one has to find $\psi = \binom{f}{g} \in D(\hat{H}_{jl})$ so that the equations

$$-g' + (\varkappa/r)g + mf = \lambda f, \quad (6.1a)$$

$$f' + (\varkappa/r)f - mg = \lambda g, \quad (6.1b)$$

are fulfilled in $(0,R)$ and (R,∞) together with the appropriate boundary conditions coupling the solutions at the point R; for simplicity, we write $\varkappa \equiv \varkappa_{jl}$.

Proposition 6.1: For any of the boundary conditions (3.5), the operator $\hat{H}_{jl}$ has at most two eigenvalues (with account of multiplicity) in $[-m,m]$.

Proof: One has only to modify slightly the argument leading to Corollary 1 to Proposition 8.19 in Ref. 21. Denote by A_1, A_2 two extensions $\hat{H}_{jl}$, where the first corresponds to the free Dirac Hamiltonian, and suppose there are more than two eigenvalues of A_2 in $[-m,m]$. Since A_2 is a self-adjoint extension of an operator with deficiency indices (2,2), there has to exist a nonzero vector $\psi\in \operatorname{Ran} E_{A_2}([-m,m]) \cap D(\overline{\widehat{H}_{jl}^{(0)}})$. We have

$$\|A_2\psi\|^2=\int_{\mathbb{R}}\lambda^2 d(\psi,E_\lambda^{(2)}\psi)\leqslant m^2\|\psi\|^2,$$

where we have denoted by $\{E_\lambda^{(j)}\}$ the spectral decompositions of A_j. At the same time, the spectrum of A_1 is contained in $(-\infty,-m]\cup[m,\infty)$ and the end points $\pm m$ are not its eigenvalues so

$$\|A_1\psi\|^2=\int_{(-\infty,-m]\cup[m,\infty)}\lambda^2 d(\psi,E_\lambda^{(1)}\psi);$$
$$> m^2\|\psi\|^2,$$

but $A_1\psi=A_2\psi$ since $\psi\in D(\overline{\widehat{H}_{jl}^{(0)}})$ so we arrive at a contradiction. ■

The points $\lambda=\pm m$ can be eigenvalues of $\widehat{H}_{jl}$ for particular boundary conditions. For instance, consider $\lambda=-m$ and $l=j-\frac{1}{2}$, i.e., $\varkappa=-(l+1)$. Equations (6.1) then have the following square integrable solutions:

$$f(r)=ar^{-\varkappa},\quad g(r)=[2ma/(1-2\varkappa)]r^{1-\varkappa},$$
$$\text{for } r\in(0,R),$$
$$f(r)=0,\quad g(r)=br^{\varkappa},\quad \text{for } r\in(R,\infty).$$

Substituting them into the boundary conditions (3.5), one can find the cases when $\lambda=-m$ is an eigenvalue. In particular, for the boundary conditions (4.6), this is true if

$$g_v^2-g_s^2-4+[8mR/(1-2\varkappa)](g_v-g_s)=0.$$

Similarly one can handle the remaining cases.

Let us turn now to eigenvalues $\lambda\in(-m,m)$ for fixed boundary conditions (3.5), which we shall write for simplicity in the form (3.6),

$$C\psi(R_-)+D\psi(R_+)=0.$$

It is clear from (6.1) that the functions f,g are continuously differentiable to any order in $(0,R)$ and (R,∞). Expressing g from (6.1b) and substituting into (6.1a), we get the Bessel equation whose solutions in $(0,R)$ and (R,∞) are of the form (3.3) with $i\mp m$ replaced by $\lambda\mp m$. Substituting them into the boundary conditions, we get the following eigenvalue equation:

$$\det(C\varphi^{(-)}(\lambda),D\varphi^{(+)}(\lambda))=0, \tag{6.2a}$$

where

$$\varphi^{(-)}(\lambda)=\begin{pmatrix} J_\nu(i(m^2-\lambda^2)^{1/2}R) \\ (-1)^{j-l+1/2}i((m-\lambda)/(m+\lambda))^{1/2}J_{\nu'}(i(m^2-\lambda^2)^{1/2}R)\end{pmatrix}, \tag{6.2b}$$

$$\varphi^{(+)}(\lambda)=\begin{pmatrix} H_\nu^{(1)}(i(m^2-\lambda^2)^{1/2}R) \\ (-1)^{j-l+1/2}i((m-\lambda)/(m+\lambda))^{1/2}H_{\nu'}^{(1)}(i(m^2-\lambda^2)^{1/2}R)\end{pmatrix}, \tag{6.2c}$$

with $\nu=l+\frac{1}{2}$, $\nu'=l'+\frac{1}{2}$, and $l'=j\mp\frac{1}{2}$ for $l=j\pm\frac{1}{2}$. According to Proposition 6.1, it has at most two solutions, or even less if some of the points $\lambda=\pm m$ is an eigenvalue.

Similarly, one can proceed for $|\lambda|>m$. There are no nonzero square integrable solutions in (R,∞) in this case and, therefore, $\widehat{H}_{jl}$ referring to the boundary conditions (3.5a) has no eigenvalues of that type. On the other hand, the boundary conditions (3.5b) yield the eigenvalue equation

$$c_1J_\nu((\lambda^2-m^2)^{1/2}R)+c_2(-1)^{j-l+1/2}\times[(\lambda^2-m^2)^{1/2}/(\lambda+m)]J_{\nu'}((\lambda^2-m^2)^{1/2}R)=0 \tag{6.3}$$

It is clear that it has for any real c_1,c_2 two infinite sequences of solutions accumulating at $\lambda=\pm\infty$ only.

B. Continuous spectrum

The spectrum of the free Dirac operator is known[22] to be purely (and absolutely) continuous and equal to $(-\infty,-m]\cup[m,\infty)$. We are going to show that the same is true for the continuous spectrum of the operators with the δ-shell interaction.

The resolvents of the self-adjoint extensions $\widehat{H}_{jl}$ with fixed j,l differ mutually by a finite-rank operator (this fact follows immediately from the Krein resolvent formula[4]) and have, therefore, the same continuous spectrum

$$\sigma_c(\widehat{H}_{jl})=\sigma_{\text{ess}}(\widehat{H}_{jl})=\sigma_{\text{ess}}(\widehat{H}_{jl}^{(D)})$$
$$=(-\infty,-m]\cup[m,\infty),$$

for all j,l, where $\widehat{H}_{jl}^{(D)}$ denotes the partial-wave "component" of H_D. The essential spectrum of $\widehat{H}_{jl}^{(D)}$ can be easily computed just solving Eqs. (6.1) in $\mathbb{R}_+$ for each $\lambda\in(-\infty,-m]\cup[m,\infty)$ and taking a suitable sequence of cutoff functions. Moreover, the spectrum of $\widehat{H}_{jl}^{(D)}$ is absolutely continuous for all j,l; this follows immediately from the partial-wave decomposition of H_D, which has an absolutely continuous spectrum. It remains in such a way to check that the singular continuous spectrum

$$\sigma_{\text{sc}}(\widehat{H}_{jl})=\varnothing,$$

for all j,l and all self-adjoint extensions $\widehat{H}_{jl}$. To this purpose, we use once more the Krein resolvent formula that yields the following relation for the resolvent of $\widehat{H}_{jl}$:

$$(\widehat{H}_{jl}-z)^{-1}=(\widehat{H}_{jl}^{(D)}-z)^{-1}+\sum_{m,n=1}^{k}\mu_{m,n}^{(jl)}(z)|g_m(z)\rangle\langle g_n(\bar{z})|, \tag{6.4}$$

where $k\leqslant 2$ is the deficiency index of the maximal common part of $\hat{H}_{jl}$ and $\hat{H}_{jl}^{(D)}$, the matrix $\mu^{(jl)}(z)$ represents a solution to the equation

$$[\mu^{(jl)}(z)]_{mn}^{-1} = [\mu^{(jl)}(z_0)]_{mn}^{-1} - (z - z_0)(g_m(\bar{z}), g_n(z_0)), \tag{6.5}$$

and the vectors $g_m(z)$ solve

$$g_m(z) = g_m(z_0) + (z - z_0)(\hat{H}_{jl}^{(D)} - z)^{-1} g_m(z_0) \tag{6.6}$$

being, therefore, analytic in $\rho(\hat{H}_{jl}^{(D)})$.

Let us consider the matrix element

$$\begin{aligned}(\phi,(\hat{H}_{jl}^{(D)} - z)^{-1}\psi) &= (\mathscr{F} U_{jl\mu}^{-1}\phi, \mathscr{F}(H_D - z)^{-1}\mathscr{F}^{-1}\mathscr{F} U_{jl\mu}^{-1}\psi) \\ &= \int_0^\infty \frac{F(p)}{p^2 + m^2 - z^2}\,dp + z\int_0^\infty \frac{G(p)}{p^2 + m^2 - z^2}\,dp,\end{aligned} \tag{6.7}$$

where $z \in \mathbb{C}\setminus\mathbb{R}$, $\phi,\psi \in \hat{\mathscr{H}}$, $\mu = -j,...,j$, and $\mathscr{F}: \mathscr{H} \to \mathscr{H}$ is the Fourier transformation to the momentum representation. If $\phi = \binom{f}{g}, \psi = \binom{u}{v}$ with $f,g,u,v \in L^2(\mathbb{R}_+) \cap L^1(\mathbb{R}_+, r\,dr)$, then

$$\begin{aligned}F(p) = p\,\overline{\tilde{f}(p)}\,[(-1)^{j-l+1/2} p\tilde{v}(p) + m\tilde{u}(p)] \\ + p\,\overline{\tilde{g}(p)}\,[(-1)^{j-l+1/2} p\tilde{u}(p) - m\tilde{v}(p)],\end{aligned} \tag{6.8a}$$

$$G(p) = p[\,\overline{\tilde{f}(p)}\,\tilde{u}(p) + \overline{\tilde{g}(p)}\,\tilde{v}(p)], \tag{6.8b}$$

where

$$\tilde{f}(p) = \int_0^\infty r^{1/2} f(r) J_{l+1/2}(pr)\,dr, \tag{6.9a}$$

$$\tilde{g}(p) = \int_0^\infty r^{1/2} g(r) J_{l'+1/2}(pr)\,dr, \tag{6.9b}$$

and similarly for $\tilde{u}$, $\tilde{v}$. Under our assumptions, $\tilde{f}$, $\tilde{g}$, $\tilde{u}$, $\tilde{v} \in C^1(\mathbb{R}_+)$. Then also $F,G \in C^1(\mathbb{R}_+)$ and for $x \in (-\infty, -m) \cup (m,\infty)$,

$$\begin{aligned}\lim_{\substack{z \to x \\ \operatorname{Im} z \gtrless 0}} \int_0^\infty \frac{F(p)}{p^2 + m^2 - z^2}\,dp &= \mathscr{P}\int_0^\infty \frac{F(p)}{p^2 + m^2 - x^2}\,dp \\ &\quad \pm i(\pi/2)(x^2 - m^2)^{-1/2} F \\ &\quad \times ((x^2 - m^2)^{1/2})\,\operatorname{sgn}(x),\end{aligned} \tag{6.10}$$

where the principal value of integral is a finite continuous function of x and $\pm$ correspond to sgn Im z. The proof of the last formula is based on the first-order Taylor expansion of F around the point $p = [(\operatorname{Re} z)^2 - m^2]^{1/2}$ in the interval $[\frac{1}{2}(x^2 - m^2)^{1/2}, 2(x^2 - m^2)^{1/2}]$. Similar calculations were used, e.g., in Ref. 23. The formula (6.10) holds also with F replaced by G.

For $z_0 = i$, $g_n(z_0)$ belong to the deficiency subspaces and are proportional to the corresponding functions (3.3) in intervals $(0,R)$ and (R,∞). Consequently, functions (6.9) corresponding to $g_n(z_0)$ are holomorphic in $\{p \in \mathbb{C}: \operatorname{Re} p > 0, |\operatorname{Im} p| < (1 + m^2)^{1/2}\}$. The edge-of-wedge theorem then shows via Eqs. (6.6)–(6.10) that $(g_m(\bar{z}), g_n(z_0))$ can be holomorphically continued from the upper half-plane on an open set containing $(-\infty,-m) \cup (m,\infty)$. The same is true for the matrix elements (6.5) and $\det[\mu^{(jl)}(z)]^{-1}$ whose continuation, therefore, has at most a discrete set of zeros. Hence $\mu^{(jl)}(z)$ can be meromorphically continued from the upper half-plane on an open set containing $(-\infty,-m) \cup (m,\infty)$.

Let us now take an interval $(a,b) \subset (-\infty,-m] \cup [m,\infty)$, which does not contain a singular point of the continued matrix function $\mu^{(jl)}$. The formulas (6.4)–(6.10) show that

$$z \mapsto (\phi,(\hat{H}_{jl} - z)^{-1}\phi) \tag{6.11}$$

can be continuously (even holomorphically) continued from the upper half-plane into (a,b) for all $\phi \in C_0^\infty(\mathbb{R}_+) \otimes \mathbb{C}^2$. Then, the known criterion[24] shows that $(a,b) \cap \sigma_{sc}(\hat{H}_{jl}) = \emptyset$. Since the poles of continued matrix $\mu^{(jl)}$ are isolated, it follows that $\sigma_{sc}(\hat{H}_{jl}) = \emptyset$. Summarizing the above results we get the following theorem.

Theorem 6.2: For any of the boundary conditions (3.5), the operator $\hat{H}_{jl}$ has at most two eigenvalues (with the account of multiplicities) in $[-m,m]$. For (3.5a), there are no eigenvalues in $(-\infty,-m) \cup (m,\infty)$ while, for (3.5b), there are two infinite sequences of eigenvalues accumulating only at $\lambda = \pm\infty$. Furthermore, one has

$$\sigma_c(\hat{H}_{jl}) = \sigma_{ess}(\hat{H}_{jl}) = \sigma_{ac}(\hat{H}_{jl}) = (-\infty,-m] \cup [m,\infty)$$

and $\sigma_{sc}(\hat{H}_{jl}) = \emptyset$.

Note added in proof: The results of Sec. VI can be obtained alternatively, without using resolvents, by the method of Behncke.[25] We are grateful to Professor Behncke for bringing this paper to our attention.

APPENDIX: ASYMMETRIC δ-SHELLS

The δ-shells do not exhaust the class of extensions covered by Theorem 3.2. Though it might not be physically interesting, we are going to demonstrate that most of the remaining extensions correspond to "asymmetric" δ-shells with (4.2) replaced by $G\delta_a(r - R)$, where G is again a 2×2 matrix and δ_a is defined by

$$\int_{R-\epsilon}^{R+\epsilon} \delta_a(r - R)\psi(r)\,dr = a\psi(R_+) + (1 - a)\psi(R_-), \tag{A1}$$

for $\psi \in D(\hat{H}_{jl}^{(0)*})$, where a is a complex number. Condition (4.3) is now replaced by

$$(1 - aB)\psi(R_+) - (1 + bB)\psi(R_-) = 0 \tag{A2}$$

where we have denoted $B = \tau_0 G$ and $b = 1 - a$. Let us denote further $\hat{H}_{jl}^{(G,a)}$ the restriction of $\hat{H}_{jl}^{(0)*}$ to the subset of its domain specified by the boundary conditions (A2).

Proposition A1: $\hat{H}_{jl}^{(G,a)}$ is a self-adjoint extension of $\hat{H}_{jl}^{(0)}$ iff

$$G - G^+ = (1 - 2\operatorname{Re} a) G^+ \tau_0 G. \tag{A3}$$

Proof: The relation (A2) represents two linearly independent boundary conditions since rank $(1 - aB, 1 + bB)$ $=$ rank$(1 - aB, B) =$ rank$(1,B) = 2$. It remains to check that it is symmetric iff (A3) is valid. We distinguish again several cases.

(i) $1 - aB$ is nonsingular. Then $\psi(R_+) = (1 - aB)^{-1}(1 + bB)\psi(R_-)$ so the requirement gives

$$\tau_0 - (1+\bar{b}B^+)(1-\bar{a}B^+)^{-1}\tau_0(1-aB)^{-1}(1+bB) = 0 \tag{A4}$$

multiplying this relation by $(1-\bar{a}B^+)$ and $(1-aB)$ from the left and right, respectively, we arrive after a short calculation at the relation (A3).

(ii) An analogous argument can be used if $1+bB$ is nonsingular.

(iii) Suppose that $\mathrm{rank}(1-aB) = \mathrm{rank}(1+bB) = 1$. Then there is a nonsingular V such that

$$V(1+bB) = \begin{pmatrix} c_1, & c_2 \\ 0, & 0 \end{pmatrix}, \quad V(1-aB) = \begin{pmatrix} 0, & 0 \\ -d_1, & -d_2 \end{pmatrix}. \tag{A5}$$

By a suitable choice of V, we can have one of the following possibilities:

(a) $c_1 = d_1 = 1$,

(b) $c_1 = d_2 = 1, \quad d_1 = 0$,

(c) $c_2 = d_1 = 1, \quad c_1 = 0$.

One can calculate the matrices VB and V from here obtaining, in particular, $\det V = a(1-a)(c_2d_1 - c_1d_2)$. Since it is nonzero, one has $a \neq 0$, 1 and $c_2 \neq d_2$ in the case (a). Furthermore, one can calculate

$$G = -\tau_0 B = \frac{1}{\det V} \begin{pmatrix} c_1d_1, & (1-a)c_2d_1 + ac_1d_2 \\ (1-a)c_1d_2 + ac_2d_1, & c_2d_2 \end{pmatrix}. \tag{A6}$$

The operator $\hat{H}_{jl}^{(G,a)}$ is self-adjoint iff (A2) is equivalent to (3.5). Thus one has to check that matrix (A6) fulfills the condition (A3) iff all the coefficients c_1, c_2, d_1, d_2 are real. For each of the possibilities (a)–(c) this can be done by a straightforward algebra. ■

The operators $\hat{H}_{jl}^{(G,a)}$ with $a \in (0,1)$ cover almost all extensions given by Theorem 3.2.

Proposition A2: The set of the self-adjoint operators $\hat{H}_{jl}^{(G,a)}$ as well as its subset corresponding to $a \in (0,1)$ coincides with the set of self-adjoint extensions $\hat{H}_{jl}$ from Theorem 3.2 with the exception of those given by the boundary conditions (3.5b) with $c_1 = d_1, c_2 = d_2$.

Proof: First, we check that the condition (A2) with $\det(1-aB) \neq 0$ is equivalent to (3.5a). Using (A4), one can check that $1+bB$ is also nonsingular, and, therefore, (A2) is equivalent to $\psi(R_-) = A_1\psi(R_+)$ for a nonsingular A_1. Since the (A2) defines self-adjoint operator by definition, it must hold $A_1 = e^{i\alpha}A$ for some α, A. Conversely, consider (3.5a) with some $A_1 = e^{i\alpha}A$. Since A_1 is nonsingular and det is a continuous function, $\det(a1 + bA_1) \neq 0$ for all $|a|$ small enough. We choose such an a and set $B = (1-A_1)(a1+bA_1)^{-1}$. Then, $1+bB = (a1+bA_1)^{-1}$ and $A_1 = (1+bB)^{-1}(1-aB)$, so we arrive back at (A2). It is clear that there are many pairs of a,B corresponding to a given A_1.

Next, one has to check that (A2) with $\det(1-aB) = 0$ is equivalent to (3.5b) with $c_1 \neq d_1$ or $c_2 \neq d_2$. As in the previous proof, there is a nonsingular V such that the relation (A5) holds. From here, one can calculate VB and V, and also $\det V = a(a-1)(c_1d_2 - c_2d_1)$. The last relation shows that it cannot hold for $c_1 = d_1$ and $c_2 = d_2$. Conversely, consider (3.5b) with $c_1d_2 - c_2d_1 \neq 0$ [i.e., just those conditions (3.5b) that cannot be rewritten with $c_1 = d_1$, $c_2 = d_2$]. Choosing $a \neq 0$, 1 and

$$V = \begin{pmatrix} ac_1, & ac_2 \\ (a-1)d_1, & (a-1)d_2 \end{pmatrix},$$

we can define

$$B = V^{-1}\begin{pmatrix} c_1, & c_2 \\ d_1, & d_2 \end{pmatrix}$$

and $G = -\tau_0 B$ so we arrive back at (A5) and (3.5b) implies (A2). ■

Finally, let us remark that the remaining extensions of Theorem 3.2 can be described as "asymmetric" δ shells with the parameter a being a 2×2 matrix (which must multiply G from the right in all formulas).

[1]S. Abe and T. Fujita, Nucl. Phys. A **475**, 657 (1987).

[2]D. D. Brayshaw, Phys. Rev. D **36**, 1465 (1987).

[3]F. Cooper, A. Khane, R. Musto, and A. Wipf, "Supersymmetry and the Dirac equation," preprint, Los Alamos National Laboratory LA-UR-88-169.

[4]S. Albeverio, F. Gesztesy, H. Holden, and R. Hoegh-Krohn, *Solvable Models in Quantum Mechanics* (Springer, Berlin, 1988).

[5]J. P. Antoine, F. Gesztesy, and J. Shabani, J. Phys. A **20**, 3687 (1987).

[6]J. Shabani, J. Math. Phys. **29**, 660 (1988).

[7]L. Dabrowski and J. Shabani, J. Math. Phys. **29**, 2241 (1988).

[8]E. Gutkin, Duke. Math. J. **49**, 1 (1982).

[9]S. M. Blinder, Phys. Rev. A **18**, 853 (1978).

[10]J. Brasche, P. Exner, J. Kuperin, and P. Šeba, "Contact interactions supported by curves and surfaces" (to appear).

[11]C. Dullemond and E. van Beveren, Ann. Phys. **105**, 318 (1977).

[12]V. D. Mur and V. S. Popov, Teor. Mat. Fiz. **65**, 238 (1985) (in Russian).

[13]B. O. Kerbikov, Yad. Fiz. **41**, 725 (1985) (in Russian).

[14]R. M. More, Phys. Rev. A **4**, 1782 (1971).

[15]R. M. More and E. Gerjuoy, Phys. Rev. A **7**, 1288 (1973).

[16]V. B. Beresteckij, E. M. Lifschitz, and L. P. Pitaevskij, *Quantum Electrodynamics* (Nauka, Moscow, 1980) (in Russian).

[17]E. C. Svendsen, J. Math. Anal. Appl. **80**, 551 (1981).

[18]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Dover, New York, 1931), Sec. III.10.

[19]N. Dunford and J. T. Schwartz, *Linear Operators II: Spectral Theory* (Interscience, New York, 1963).

[20]M. G. Calkin, D. Kiang, and Y. Nogami, Phys. Rev. C **38**, 1076 (1988).

[21]J. Weidmann, *Linear Operators in Hilbert Spaces* (Springer, Berlin, 1980).

[22]T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1966).

[23]J. Dittrich and P. Exner, Czech. J. Phys. B **37**, 1028 (1987).

[24]M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV* (Academic, New York, 1978).

[25]N. Behncke, Proc. Math. Soc. **72**, 82 (1978).

Coherent and thermal coherent state

A. Mann and M. Revzen
Department of Physics, Technion, Haifa 32000, Israel
K. Nakamura
Division of Natural Science, Meiji University, Izumi, 1-9-1 Eifuku, Suginami-ku, Tokyo, Japan
H. Umezawa and Y. Yamanaka
Theoretical Physics Institute, Department of Physics, University of Alberta, Edmonton, Alberta T6G 2J1, Canada

(Received 22 June 1989; accepted for publication 9 August 1989)

The characterization of coherent states as the quantum states that split into two uncorrelated beams is considered. The characterization leads to the study of coherent states at finite temperature—thermal coherent states (TCS's). These TCS's are defined within the formalism of thermo field dynamics (TFD). TFD allows a generalization of the uncertainty relation that accounts for both thermal and quantum fluctuations. The TCS is shown to be a minimal state for the generalized uncertainty relation.

I. INTRODUCTION

Coherent states were introduced in physics by Schrödinger[1] in 1926 in connection with the simple harmonic motion of a particle in quantum mechanics. Schrödinger showed that if the particle is in what we now call a coherent state (CS) then its motion simulates most closely the motion of the corresponding (i.e., with the same harmonic oscillator Hamiltonian) classical particle. Further, it turned out that the CS is a minimal uncertainty state, i.e., the particle that is in a CS has its dispersion in momentum multiplied by its dispersion in its coordinates equal to $\frac{1}{4}$ ($\hbar = 1$), which is its minimal possible value.

In more recent times, CS's are used extensively in studies of quantum optics.[2,3] It is in this connection that the state acquired its name: coherent state (CS).[3] Here too when the quantum radiation field is in a CS it simulates most closely the classical radiation. Thus the quantum field that is produced by a classical current (i.e., a current whose value is a prescribed function of space and time) is in a CS.[3] The CS proved useful in the study of interacting bosons[4] and CS was considered for potentials other than the harmonic oscillator,[5] and in field theory one now studies a fermionic CS.[6,7]

Particularly illuminating characterization of the CS was unveiled by Aharonov *et al.* (AFLP).[8] These authors showed that the CS possesses yet another (though related) classical property. We discuss this at length in Sec. II. There we show that a modification to AFLP formulas (that must be instated when finite systems are considered) allows us to connect this work with the study of Emch and Hegerfeldt (EH).[9] We discuss this connection in Sec. III. The work of EH extended the idea of a CS to the thermal coherent state (TCS), i.e., to states possessing some features of the CS at finite temperature. This topic, the TCS, is our main concern in this paper and we study the TCS via the language and technique that was developed for the formulation of thermal physics: the so-called thermo field dynamics (TFD).[10,11] We contend that the TFD formalism is a convenient one for the study of the TCS. The equilibrium TFD was shown[12] to be equivalent to the C^*-algebra approach to statistical physics[13]; it was in the language of the latter that the TCS was defined by EH.

We show that TFD allows a natural extension of the formalism of the CS to finite temperatures and leads to a new version of the TCS, as was first proposed by Barnett and Knight (BK).[14,15] In Sec. IV we review the main results of TFD that are relevant to us. On physical grounds we expect that as the temperature is raised from zero the product of the dispersion in the momentum and coordinates should rise because of thermal fluctuations. We show in Sec. V that TFD provides a relation between quantum and thermal fluctuations called the generalized uncertainty relation (GUR), from which an uncertainty inequality including thermal effects is derived. In Sec. VI we study the TCS of TFD given by BK,[14] putting an emphasis on the GUR. When the dispersion product in the TCS is evaluated in the GUR, the minimal value is attained. There the characteristic function of the TCS is also calculated and we compare the TCS of TFD with that of EHs. Section VII is devoted to remarks and conclusions.

II. SPLIT SOURCE—CHARACTERIZATION OF COHERENT STATE

In this section we shall review the analysis of Ref. 8 (AFLP). We shall introduce a modification to their formulas, which will be useful to us later on. AFLP consider the problem of splitting a beam from a source A into two channels, B and C. The source A defines a state [their Eq. (12)],

$$|\psi_A\rangle = f_A(a_A^\dagger)|0\rangle_A. \tag{2.1}$$

Here $|0\rangle_A$ is the "*no quanta*" state of the source A. Note that $a_A^\dagger$ is the creation operator for a quantum of the source while f_A is some function to be determined. The beam of A is split into two beams, or modes, B and C, which are orthogonal; for example, they move in different directions. We now search for the case in which the split beam is exactly a product wavefunction. In this case the state of the split beam is given by

$$|\psi\rangle_2 = f_B(a_B^\dagger) f_C(a_C^\dagger)|0\rangle_{BC}. \tag{2.2}$$

 0022-2488/89/122883-08$02.50

Now the "no quanta" mode for the channels B and C, which are mutually orthogonal, may be written as

$$|0\rangle_{BC} = |0\rangle_B \otimes |0\rangle_C \tag{2.3}$$

[their Eq. (9)]. At this point AFLP proceed to relate the above-mentioned vacua, as they appear in Eqs. (2.2) and (2.3) above, by [their Eq. (17)]

$$|0\rangle_A = |0\rangle_B \otimes |0\rangle_C. \tag{2.4}$$

It is our contention that this is too narrow. The more general equation is

$$|0\rangle_A \otimes |0\rangle_R = |0\rangle_B \otimes |0\rangle_C. \tag{2.5}$$

Here $|0\rangle_R$ is the vacuum orthogonal to $|0\rangle_A$ and, with it, spans the vector space of the rhs of Eq. (2.5). We hasten to add that this modification of the equation does not affect the argument or results of Ref. 8 becasue this orthogonal state was taken to be unexcited in AFLP study. To see that this state should be in the equation we note that both sides of the equation pertain to the same space and since the rhs involves two orthogonal states—so should the lhs. (Although this is not strictly that simple for infinite spaces.) We now proceed with the proof of Ref. 8, which, as was stated above, is unaffected by the above modification. Nonetheless, it is useful for us to go through with that part of the argument and add the properties of this orthogonal state as it will play a more prominent role later.

With AFLP[8] we now wish to identify the quantum state that possesses the property of being split into a product wavefunction. The splitting itself is expressed [Eq. (13) of AFLP]:

$$a_A^\dagger = \mu a_B^\dagger + \nu a_C^\dagger, \tag{2.6}$$

which describes the case of each quantum of the source A splitting with amplitudes μ and ν to the channels B and C, respectively. The creation operator to the orthogonal (albeit empty) state is

$$a_R^\dagger = -\mu^* a_B^\dagger + \nu^* a_C^\dagger \tag{2.7}$$

with

$$|\nu|^2 + |\mu|^2 = 1. \tag{2.8}$$

The operator equation that is to be solved is the very same as the one considered by AFLP, i.e., [their Eq. (14)],

$$f_A(\mu a_B^\dagger + \nu a_C^\dagger) = f_B(a_B^\dagger) f_C(a_C^\dagger), \tag{2.9}$$

because the orthogonal state is vacuum, i.e., $f_R(a_R^\dagger) = 1$. The solution of (2.9) is discussed in Ref. 8, where it is shown to be

$$F_A(a_A^\dagger) = \exp(\alpha a_A^\dagger), \tag{2.10a}$$

$$F_B(a_B^\dagger) = \exp(\alpha\mu a_B^\dagger), \tag{2.10b}$$

$$F_C(a_C^\dagger) = \exp(\alpha\nu a_C^\dagger), \tag{2.10c}$$

where

$$F_i(z) = f_i(z)/f_i(0), \tag{2.11}$$

for $i = A, R, B$, and C. In (2.11) and hereafter, $f(0)$ stands for the coefficient of the zeroth power of $a^\dagger$ in the expansion of $f(a^\dagger)$, and we simply say that $f(0)$ is obtained by setting $a^\dagger = 0$ in $f(a^\dagger)$. Equations (2.10) imply that the unique normalized quantum state that can be split as stipulated is the Glauber coherent state, $|\alpha\rangle$, which is an eigenfunction of the annihilation operator a with eigenvalue α:

$$|\alpha\rangle = \exp(-|\alpha|^2/2)\exp(\alpha a^\dagger)|0\rangle. \tag{2.12}$$

Now if the wavefunction of the source A is given by some complex α, viz., $|\psi_A\rangle = |\alpha\rangle_A$, then the orthogonal state is given by $\alpha_R = 0$; while the other two channels are specified by $\alpha_B = \mu\alpha$, $\alpha_C = \nu\alpha$. This completes our review of the results of Ref. 8 with the formal modification as a result of the orthogonal state.

The important of the role of the orthogonal state can now be seen as follows. Consider two *independent* sources A and R—the combined wavefunction is the product

$$|\psi_A\rangle|\psi_R\rangle = f_A(a_A^\dagger) f_R(a_R^\dagger)|0\rangle_A \otimes |0\rangle_R. \tag{2.13}$$

Here we have taken the two modes to be orthogonal. We now ask a more general question than the one considered above: What are the quantum states A and R that can be split into two different channels B and C such that the latter are also a product wavefunction, i.e.,

$$|\psi_B\rangle|\psi_C\rangle = f_B(a_B^\dagger) f_C(a_C^\dagger)|0\rangle_B \otimes |0\rangle_C. \tag{2.14}$$

Physically this would mean that the splitting be such that it could not be delineated from a case where the two beams were simply two independent sources. In this more general case than the one considered in AFLP, we have to solve the operator equation

$$f_A(\mu a_B^\dagger + \nu a_C^\dagger) f_R(-\nu^* a_B^\dagger + \mu^* a_C^\dagger) = f_B(a_B^\dagger) f_C(a_C^\dagger). \tag{2.15}$$

The technique of tackling this equation parallels that used in solving Eq. (2.9) above. We note that we have, from Eq. (2.15),

$$f_A(0) f_R(0) = f_B(0) f_C(0), \tag{2.16}$$

with none of the $f_i(0) = 0$. See the remark just below (2.11). Thus we may divide Eq. (2.15) to obtain the equation for the normalized functions (of operators)

$$F_A(\mu a_B^\dagger + \nu a_C^\dagger) F_R(-\nu^* a_B^\dagger + \mu^* a_C^\dagger) = F_B(a_B^\dagger) F_C(a_C^\dagger), \tag{2.17}$$

with $F_i(0) = 1$ for $i = A, R, B$, and C. Setting in turn $a_B^\dagger = 0$ and $a_C^\dagger = 0$, respectively, we obtain

$$\begin{aligned} &F_A^{-1}(\mu a_B^\dagger) F_A^{-1}(\nu a_C^\dagger) F_A(\mu a_B^\dagger + \nu a_C^\dagger) \\ &\quad = F_R^{-1}(-\nu^* a_B^\dagger + \mu^* a_C^\dagger) F_R(\mu^* a_C^\dagger) F_R(-\nu^* a_B^\dagger). \end{aligned} \tag{2.18}$$

We discuss the solution of this equation in a separate publication. We now wish to consider a particular case where

$$F_A(a_A^\dagger) = \exp(\alpha a_A^\dagger), \tag{2.19}$$

i.e., when the normalized wavefunction of the A channel is a coherent state, thus the A wavefunction is given by

$$|\psi_A\rangle = \exp(-|\alpha|^2/2)\exp(\alpha a_A^\dagger)|0\rangle_A. \tag{2.20}$$

It is seen directly (e.g., by substitution) that in this case we have (with arbitrary α')

$$F_R(a_R^\dagger) = \exp(\alpha' a_R^\dagger), \tag{2.21}$$

i.e., the normalized orthogonal state is also a coherent state,

$$|\psi_R\rangle = \exp(-|\alpha'|^2/2)\exp(\alpha' a_R^\dagger)|0\rangle_R, \tag{2.22}$$

and from Eq. (2.17) it then follows that

$$|\psi_B\rangle = \exp(-|\alpha_B|^2/2)\exp(\alpha_B a_B^\dagger)|0\rangle_B, \quad (2.23a)$$

$$|\psi_C\rangle = \exp(-|\alpha_C|^2/2)\exp(\alpha_C a_C^\dagger)|0\rangle_C, \quad (2.23b)$$

with

$$\alpha_B = \mu\alpha - \nu^*\alpha', \quad (2.24a)$$

$$\alpha_C = \nu\alpha - \mu^*\alpha', \quad (2.24b)$$

and thus

$$|\alpha_B|^2 + |\alpha_C|^2 = |\alpha|^2 + |\alpha'|^2. \quad (2.25)$$

III. CENTER OF MASS OF TWO PARTICLES

In this section we shall establish a relation between the results of Ref. 8 (AFLP) as summarized above and some results of Emch and Hegerfeldt (EH).[9] EH consider two independent particles, which we label B and C, whose combined wavefunction is a product [cf. their discussion below Eq. (2.4) in Ref. 9]. We shall write the single particle wavefunctions in field theoretical notation, e.g., the wavefunction of particle B is

$$|\psi_B\rangle = f_B(a_B^\dagger)|0\rangle_B. \quad (3.1)$$

This we can always do because of the completeness, in this case, of the harmonic oscillator wavefunctions. In (3.1)

$$a_B^\dagger = (1/\sqrt{2})\left[\sqrt{\lambda_B}Q_B - i(P_B/\sqrt{\lambda_B})\right]; \quad \lambda_B = m_B\omega_B, \quad (3.2)$$

with $\hbar = 1$, and Q and P, respectively, the coordinate and momentum operators for the particle B. Note that m_B and ω_B are, respectively, the mass and frequency of the B oscillator in terms of which we expand $|\psi_B\rangle$. Here f_B is the wavefunction for the B particle in this notation. Hence the two particles wavefunction is given by

$$|\psi_B\rangle|\psi_C\rangle = f_B(a_B^\dagger)f_C(a_C^\dagger)|0\rangle_B \otimes |0\rangle_C. \quad (3.3)$$

It is understood that the B and C modes are orthogonal. EH impose an additional constraint: they consider the case where the center of mass of the two particles is in a pure state, and moreover it is a coherent state. Clearly under these conditions, where the center of mass, which we call A, is in a pure state [so is the relative coordinate (R)]. Thus EH consider

$$f_B(a_B^\dagger)f_C(a_C^\dagger)|0\rangle_B \otimes |0\rangle_C = f_A(a_A^\dagger)f_R(a_R^\dagger)|0\rangle_A \otimes |0\rangle_R. \quad (3.4)$$

We note in passing that this equation is quite similar in appearance to the one considered by AFLP. We shall return to this point below. In (3.4) EH stipulate that the center of mass is in a coherent state,

$$f_A(a_A^\dagger) = [\exp(-|\alpha|^2)/2)]\exp(\alpha a_A^\dagger) \quad (3.5)$$

for some α. Since the A mode describes the center of mass we have, via (3.2), relations between $a_A^\dagger$ with $a_R^\dagger$ on one hand and $a_C^\dagger$ with $a_B^\dagger$ on the other. These relations follow from

$$m_A P_R = m_B P_C - m_C P_B, \quad Q_R = Q_C - Q_B,$$
$$P_A = P_B + P_C, \quad m_A Q_A = m_B Q_B + m_C Q_c, \quad (3.6)$$
$$m_A = m_B + m_C, \quad m_R m_A = m_B m_C.$$

It is convenient to define

$$\mu_i = m_i/m_A. \quad (3.7)$$

Converting the relations among the Q's and P's with (3.7) to those among the $a^\dagger$'s and a's we obtain

$$\begin{aligned} a_A^\dagger = {} & \tfrac{1}{2}[\mu_B\sqrt{\lambda_A/\lambda_B} + \sqrt{\lambda_B/\lambda_A}]a_B^\dagger \\ & + \tfrac{1}{2}[\mu_B\sqrt{\lambda_A/\lambda_B} - \sqrt{\lambda_B/\lambda_A}]a_B \\ & + \tfrac{1}{2}[\mu_C\sqrt{\lambda_A/\lambda_C} + \sqrt{\lambda_C/\lambda_A}]a_C^\dagger \\ & + \tfrac{1}{2}[\mu_C\sqrt{\lambda_A/\lambda_C} - \sqrt{\lambda_C/\lambda_A}]a_C, \end{aligned} \quad (3.8a)$$

$$\begin{aligned} a_R^\dagger = {} & -\tfrac{1}{2}[\sqrt{\lambda_R/\lambda_B} + \mu_C\sqrt{\lambda_B/\lambda_R}]a_B^\dagger \\ & - \tfrac{1}{2}[\sqrt{\lambda_R/\lambda_B} - \mu_C\sqrt{\lambda_B/\lambda_R}]a_B \\ & + \tfrac{1}{2}[\sqrt{\lambda_R/\lambda_C} + \mu_B\sqrt{\lambda_C/\lambda_R}]a_C^\dagger \\ & + \tfrac{1}{2}[\sqrt{\lambda_R/\lambda_C} - \mu_B\sqrt{\lambda_C/\lambda_R}]a_C. \end{aligned} \quad (3.8b)$$

This is seen to be a Bogoliubov transformation, i.e., a linear mixing of $a^\dagger$'s and a's. A more detailed analysis of the EH paper would require this to be discussed. However, the essence of our point can be brought in when the simpler case holds—i.e., when $a_A^\dagger$ and $a_R^\dagger$ depend only on $a_B^\dagger$ and $a_C^\dagger$. This implies [recall (3.2)]

$$\omega_A = \omega_B = \omega_C = \omega_R. \quad (3.9)$$

With (3.9) we are led to

$$a_B^\dagger = \sqrt{\mu_B}a_A^\dagger - \sqrt{\mu_C}a_R^\dagger, \quad (3.10a)$$

$$a_C^\dagger = \sqrt{\mu_C}a_A^\dagger + \sqrt{\mu_B}a_R^\dagger. \quad (3.10b)$$

Substituting this into (3.4) leads to the operator equation:

$$f_B(\sqrt{\mu_B}a_A^\dagger - \sqrt{\mu_C}a_R^\dagger)f_C(\sqrt{\mu_C}a_A^\dagger + \sqrt{\mu_B}a_R^\dagger) = f_A(a_A^\dagger)f_R(a_R^\dagger), \quad (3.11)$$

with $f_A(a_A^\dagger)$ given by (3.5). We can solve this equation using the technique developed in Ref. 8 and reviewed in the previous section. The essential point is that we may treat the operator as an independent complex variable. The argument is briefly as follows. First we note that none of the functions can vanish anywhere (except at infinity). Then we consider $F_j(a_j^\dagger) = f_j(a_j^\dagger)/f_j(0)$. In terms of these normalized (at the origin) F's the equation is

$$F_B(\sqrt{\mu_B}a_A^\dagger - \sqrt{\mu_C}a_R^\dagger)F_C(\sqrt{\mu_C}a_A^\dagger + \sqrt{\mu_B}a_R^\dagger) = \exp(\alpha a_A^\dagger)F_R(a_R^\dagger), \quad (3.12)$$

where we substituted for $F_A(a_A^\dagger)$ its known value, $\exp(\alpha a_A^\dagger)$. A solution to this equation is all $f_j(a_j^\dagger)|0\rangle$ ($j = B$, C, and R) are coherent states, i.e.,

$$\begin{aligned} F_B(a_B^\dagger) &= \exp(\alpha_B a_B^\dagger), \\ F_C(a_C^\dagger) &= \exp(\alpha_C a_C^\dagger), \\ F_R(a_R^\dagger) &= \exp(\alpha_R a_R^\dagger), \end{aligned} \quad (3.13)$$

with

$$\alpha = \sqrt{\mu_B}\alpha_B + \sqrt{\mu_C}\alpha_C, \quad (3.14a)$$

$$\alpha_R = -\sqrt{\mu_C}\alpha_B + \sqrt{\mu_B}\alpha_C. \quad (3.14b)$$

[In a separate publication the properties of functions that factorize in two distinct ways (here, B and C are factorized as well as A and R) are considered in detail. There we show that (3.13) is, in fact, the only solution.]

We now compare the results of Refs. 8 and 9 as follows: The basic equations that characterize the coherent state are the same. AFLP consider the case where the orthogonal state is empty [i.e., $f_R(a_R^\dagger) = 1$] while EH stipulate $f_R(a_R^\dagger)$ via (3.4). In both cases the conclusions are that all the states are coherent states (CS's). [We may consider $f_R(a_R^\dagger) = 1$ as CS with $\alpha_R = 0$.] We summarize the results as stating that if the system can be described as a product of two wavefunctions in two different ways and if one of the states is a coherent state then—all are. We see that, as was stressed originally by AFLP, coherent states (CS's) are characterized by their possession of the classical property of allowing the split states to have only "local properties." For example, the states B and C do not have global correlations as a result of their origin from A: Measurements of B and C separately determine these states completely. The consideration of EH, though similar, led them to consider thermal coherent states (TCS's). These we shall consider in Sec. VI, within the theory of thermo field dynamics (TFD), after a brief introduction to TFD that is given in the next section, and derive an important relation in Sec. V.

IV. THERMO FIELD DYNAMICS

In this section we list some formulas and results from thermo field dynamics (TFD)[10,11] that we shall use later.

TFD is a formulation of thermal physics that was initiated by one of the authors (H. U.). Our discussion will be confined to equilibrium only so most of the results listed below can be found in Refs. 10 and 11.

TFD gives the expectation value of an arbitrary dynamical quantity, an operator A, as an expectation value in a thermal vacuum,

$$\overline{A} = \langle 0(\beta)|A|0(\beta)\rangle. \tag{4.1}$$

Here $|0(\beta)\rangle$ is the thermal vacuum state characterized by the inverse temperature β and the Hamiltonian of the problem. In this way the various thermal vacua are, for a fixed Hamiltonian, characterized by various possible values of β, much like the ground states of a magnetic system are characterized by the various possible orientations for the magnetization. The price for this concise "extension of zero temperature field theory to encompass thermal degrees of freedom"[10] is that we have to work in an expanded space of states. Thus to every state $|n\rangle$ of the original Hilbert space $\mathscr{H}$ we now associate $|\tilde{n}\rangle \in \tilde{\mathscr{H}}$, where $\tilde{\mathscr{H}}$ is the "tildian" space. To every physical operator A acting in $\mathscr{H}$ we associate a tildian operator $\tilde{A}$ in $\tilde{\mathscr{H}}$. The operations of associating $\tilde{A}$ to A, i.e., to its tilde conjugate, are spelled out in Refs. 10 and 11. Thus we consider operators that act in $\mathscr{H} \otimes \tilde{\mathscr{H}}$. An important role is played by operators that are "self-tildian," i.e., that remain invariant under the tilde conjugation operation. Such an operator is the one that carries the vacuum (ground) state from $\beta \to \infty$ (i.e., $T = 0$) to finite β,

$$|0(\beta)\rangle = U(\beta)|0(\infty)\rangle. \tag{4.2}$$

This operator, for free bosons, is given by[10,11]

$$U(\beta) = \exp(-iG_B), \tag{4.3}$$

$$G_B = i\theta(\beta)[a^\dagger \tilde{a}^\dagger - \tilde{a}a]. \tag{4.4}$$

(Here we confine ourselves to one mode only—for notational simplicity.) In these formulas $\tilde{a}(\tilde{a}^\dagger)$ is the annihilation (creation) operator for the mode in question for the tilde particle. The operator $U(\beta)$ is unitary, self-tildian, and acts in the combined space

$$\hat{\mathscr{H}} = \mathscr{H} \otimes \tilde{\mathscr{H}}. \tag{4.5}$$

The angle $\theta(\beta)$ is given by

$$\cosh\theta(\beta) = c(\beta) = [1 - \exp(-\beta\omega)]^{-1/2}, \tag{4.6a}$$

$$\sinh\theta(\beta) = s(\beta) = [\exp(\beta\omega) - 1]^{-1/2}, \tag{4.6b}$$

from which we have

$$c(\beta)^2 - s(\beta)^2 = 1. \tag{4.7}$$

Here ω is the energy of the mode with $\hbar = 1$. The operators $(a, a^\dagger, \tilde{a}, \tilde{a}^\dagger)$ are related to other operators $(\xi, \xi^\dagger, \tilde{\xi}, \tilde{\xi}^\dagger)$ as follows:

$$\xi \equiv U(\beta) a U(\beta)^{-1} \tag{4.8a}$$

$$= c(\beta)a - s(\beta)\tilde{a}^\dagger, \tag{4.8b}$$

$$\tilde{\xi}^\dagger \equiv U(\beta)\tilde{a}^\dagger U(\beta)^{-1} \tag{4.8c}$$

$$= -s(\beta)a + c(\beta)\tilde{a}^\dagger, \tag{4.8d}$$

$$\xi^\dagger \equiv U(\beta)a^\dagger U(\beta)^{-1} \tag{4.9a}$$

$$= c(\beta)a^\dagger - s(\beta)\tilde{a}, \tag{4.9b}$$

$$\tilde{\xi} \equiv U(\beta)\tilde{a}U(\beta)^{-1} \tag{4.9c}$$

$$= -s(\beta)a^\dagger + c(\beta)\tilde{a}. \tag{4.9d}$$

Equations (4.8b), (4.8d), (4.9b), and (4.9d) are called the thermal Bogoliubov transformation. It is obvious from (4.2) and (4.7) that

$$\xi|0(\beta)\rangle = \tilde{\xi}|0(\beta)\rangle = 0. \tag{4.10}$$

The arguments and (4.2), (4.3), (4.8a), (4.8c), (4.9a), and (4.9c) are formal in quantum field theory because there infinite degrees of freedom are involved and the operator U does not exist. Even in such a case, Eqs. (4.8b), (4.8d), (4.9b), (4.9d), and (4.10) are still true and, as a matter of fact, these were the starting points in construction of the TFD formulation.[10]

The dynamical law in TFD is described by the total Hamiltonian $\hat{H}$,[10]

$$\hat{H} = H - \tilde{H}, \tag{4.11}$$

where H is the ordinary Hamiltonian and $\tilde{H}$ is obtained from it from replacing $(a, a^\dagger)$ with $(\tilde{a}, \tilde{a}^\dagger)$ together with complex conjugations of c numbers.

V. RELATION BETWEEN QUANTUM AND THERMAL FLUCTUATIONS

At zero temperature the commutation relation for the momentum P and the coordinate Q leads to the uncertainty relation

$$\langle (P - \langle P\rangle)^2\rangle \langle (Q - \langle Q\rangle)^2\rangle \geqslant \hbar^2/4. \tag{5.1}$$

We explicitly preserve $\hbar$ for clarity in this section. This inequality is saturated, i.e., attains its lowest value, $\hbar^2/4$, when the quantities are evaluated in a CS.

In this section we derive a relation among the fluctuations and correlations of momenta $(P,\widetilde{P})$ and coordinates $(Q,\widetilde{Q})$—both the dynamical (nontildian) and thermal (tildian) operators—and the corresponding zero-temperature operators. This relation reduces to the uncertainty relation (5.1) as the temperature is reduced to zero ($\beta\rightarrow\infty$). We term this relation the generalized uncertainty relation (GUR).[17] The GUR can be put into an inequality of a simpler form by imposing some limitations on the state vectors. In this sense, the GUR inequality is not as general as that implied by the ordinary uncertainty relation. Nonetheless, it does apply to a wide class of thermal states that includes all equilibrium states. We will also see that the GUR inequality characterizes TCS's introduced in the next section as states for which the GUR is saturated.

The GUR has been derived and discussed in a separate paper.[17] Here its derivation is sketched only briefly below.

The momenta and coordinate operators are related to the oscillator operators by [cf. (3.2)]

$$P=i(\sqrt{\lambda/2\hbar})(a^\dagger-a),\quad Q=(\sqrt{\hbar/2\lambda})(a^\dagger+a); \tag{5.2a}$$

$$\widetilde{P}=-i(\sqrt{\lambda/2\hbar})(\tilde{a}^\dagger-\tilde{a}),\quad \widetilde{Q}=(\sqrt{\hbar/2\lambda})(\tilde{a}^\dagger+\tilde{a}); \tag{5.2b}$$

where

$$[Q,P]=i\hbar, \tag{5.3a}$$

$$[\widetilde{Q},\widetilde{P}]=-i\hbar. \tag{5.3b}$$

These equations can be inverted to give a, $\tilde{a}$, $a^\dagger$, and $\tilde{a}^\dagger$ in terms of P, $\widetilde{P}$, Q, and $\widetilde{Q}$, respectively. We get from (4.8) and (4.9)

$$UQU^{-1}=cQ+s\widetilde{Q},\quad UPU^{-1}=cP+s\widetilde{P}, \tag{5.4a}$$

$$U\widetilde{Q}U^{-1}=c\widetilde{Q}+sQ,\quad U\widetilde{P}U^{-1}=c\widetilde{P}+sP. \tag{5.4b}$$

The momentum operators at finite β could in general involve a different λ, e.g.,

$$P_\xi=i(\sqrt{\lambda'/2\hbar})(\xi^\dagger-\xi), \tag{5.5}$$

where $\lambda'\neq\lambda$ in general, with corresponding relations for the other operators. The relations between the ξ operators (i.e., those at finite β) and the zero temperature ones are

$$P=(\sqrt{\lambda/\lambda'})(cP_\xi+s\widetilde{P}_\xi),\quad Q=(\sqrt{\lambda'/\lambda})(cQ_\xi+s\widetilde{Q}_\xi), \tag{5.6a}$$

$$\widetilde{P}=(\sqrt{\lambda/\lambda'})(c\widetilde{P}_\xi+sP_\xi),\quad \widetilde{Q}=(\sqrt{\lambda'/\lambda})(c\widetilde{Q}_\xi+sQ_\xi), \tag{5.6b}$$

with the commutation relations

$$[Q_\xi,P_\xi]=i\hbar, \tag{5.7a}$$

$$[\widetilde{Q}_\xi,\widetilde{P}_\xi]=-i\hbar. \tag{5.7b}$$

Let us now define

$$\Delta P=P-\langle P\rangle \tag{5.8}$$

with the corresponding definitions for ΔQ, $\Delta\widetilde{P}$, $\Delta\widetilde{Q}$ and ΔP_ξ, ΔQ_ξ, $\Delta\widetilde{P}_\xi$, and $\Delta\widetilde{Q}_\xi$. Now by direct substitution of (5.6) with (4.7) one gets the following equality:

$$\begin{aligned}\tfrac{1}{2}[\langle(\Delta P)^2\rangle\langle(\Delta Q)^2\rangle&+\langle(\Delta\widetilde{P})^2\rangle\langle(\Delta\widetilde{Q})^2\rangle\\&-2\langle\Delta P\,\Delta\widetilde{P}\rangle\langle\Delta Q\,\Delta\widetilde{Q}\rangle]\\=\tfrac{1}{2}[\langle(\Delta P_\xi)^2\rangle&\langle(\Delta Q_\xi)^2\rangle\\&+\langle(\Delta\widetilde{P}_\xi)^2\rangle\langle(\Delta\widetilde{Q}_\xi)^2\rangle\\&-2\langle\Delta P_\xi\,\Delta\widetilde{P}_\xi\rangle\langle\Delta Q_\xi\Delta\widetilde{Q}_\xi\rangle].\end{aligned} \tag{5.9}$$

In (5.8) and (5.9) the expectation value is taken with the arbitrary thermal state $|T\rangle$. The factor $\frac{1}{2}$ is introduced for convenience—we want the averaged, tildian–nontildian symmetrized, expression. Equation (5.9) is the most general form of the GUR.[17]

To obtain our final inequality we restrict the considered states $|T\rangle$ to those that are subject to the following two conditions: (i) $|T\rangle$ is a product of $\xi^\dagger$ and $\tilde{\xi}^\dagger$ acting on $|0(\beta)\rangle$, so that

$$|T\rangle=|T_\xi\rangle\otimes|T_{\tilde{\xi}}\rangle; \tag{5.10}$$

(ii) $|T\rangle$ is tildian invariant,

$$|T\rangle^{\sim}=|T\rangle,\quad \langle T|^{\sim}=\langle T|. \tag{5.11}$$

The thermal vacuum $|0(\beta)\rangle$ is the simplest example that satisfies all of these conditions. Because of (ii), since the tilde conjugation of the matrix elements implies its complex conjugation, we obtain

$$\langle P\rangle=\langle\widetilde{P}\rangle,\quad \langle P^2\rangle=\langle\widetilde{P}^2\rangle \tag{5.12}$$

and so on. With (5.10) we have

$$\langle\Delta P_\xi\,\Delta\widetilde{P}_\xi\rangle=\langle\Delta\widetilde{P}_\xi\rangle\langle\Delta\widetilde{P}_\xi\rangle=0,\ \text{etc.} \tag{5.13}$$

With substitutions of (5.12) and (5.13) into (5.9), (5.9) now reads simply as

$$\begin{aligned}\langle(\Delta P)^2\rangle\langle(\Delta Q)^2\rangle&-\langle\Delta P\,\Delta\widetilde{P}\rangle\langle\Delta Q\,\Delta\widetilde{Q}\rangle\\&=\langle(\Delta P_\xi)^2\rangle\langle(\Delta Q_\xi)^2\rangle\geqslant\hbar^2/4.\end{aligned} \tag{5.14}$$

In deriving the last inequality in (5.14), we used the fact that the calculations of $\langle(\Delta P_\xi)^2\rangle\langle(\Delta Q_\xi)^2\rangle$ are identical with those of $\langle(\Delta P)^2\rangle\langle(\Delta Q)^2\rangle$ at zero temperature. Equation (5.14) is the GUR inequality. It is (i) and (ii) on $|T\rangle$ that make this GUR inequality less general than the uncertainty relation (5.1). As it was pointed out above, $|0(\beta)\rangle$ satisfies (i) and (ii). At the zero-temperature limit, $\langle\Delta P\,\Delta\widetilde{P}\rangle$ and $\langle\Delta Q\,\Delta\widetilde{Q}\rangle$ in (5.14) factorize as

$$\langle\Delta P\,\Delta\widetilde{P}\rangle=\langle\Delta P\rangle\langle\Delta\widetilde{P}\rangle=0 \tag{5.15}$$

so that (5.14) simply reduces to (5.1).

The importance of (5.14) [or (5.9), in general] lies in the fact that since the total fluctuation and the quantum fluctuation are given by $\langle(\Delta P)^2\rangle\langle(Q)^2\rangle$ and $\langle(\Delta P_\xi)^2\rangle\langle(\Delta Q_\xi)^2\rangle$, respectively, the pure thermal fluctuation appears to be $\langle\Delta P\,\Delta\widetilde{P}\rangle\langle\Delta Q\,\Delta\widetilde{Q}\rangle$.

As it will be shown in the following section, the GUR inequality will in effect be used to extend the usual CS to the TCS.

VI. THERMAL COHERENT STATE AND ITS CHARACTERISTIC FUNCTION

Extending the Glauber CS in (2.12) to finite temperature, Barnett and Knight introduced the thermal coherent state by the relation[14–16]

$$\xi\,|\alpha(\beta)\rangle = \alpha(\beta)|\alpha(\beta)\rangle. \tag{6.1}$$

There are two natural candidates for this state

$$|\alpha;\rangle \equiv U(\beta)D(\alpha,a)|0(\infty)\rangle = D(\alpha,\xi)|0(\beta)\rangle, \tag{6.2}$$

$$\|\alpha;\beta\rangle\rangle \equiv D(\alpha,a)\,U(\beta)|0(\infty)\rangle. \tag{6.3}$$

Here U is found in (4.3) and the displacement operator is defined by

$$D(\alpha,a) \equiv \exp[\alpha a^\dagger - \alpha^* a + \tilde{\alpha}\tilde{a}^\dagger - \tilde{\alpha}^*\tilde{a}]. \tag{6.4}$$

With (4.2) and (4.8) the first definition (6.2) leads to eigenvalues α, $\tilde{\alpha}$ that are independent of β,

$$\xi\,|\alpha;\beta\rangle = \alpha|\alpha;\beta\rangle, \tag{6.5a}$$

$$\tilde{\xi}\,|\alpha;\beta\rangle = \tilde{\alpha}|\alpha;\beta\rangle. \tag{6.5b}$$

The second definition (6.3) can be written as

$$\|\alpha;\beta\rangle\rangle = D(\alpha c - \tilde{\alpha}^* s,\xi)|0(\beta)\rangle \tag{6.6}$$

and, therefore, can be expressed in terms of the first definition

$$\|\alpha;\beta\rangle\rangle = |\alpha c - \tilde{\alpha}^* s;\beta\rangle. \tag{6.7}$$

Thus the second definition leads to β-dependent eigenvalues,

$$\xi\,\|\alpha\beta\rangle\rangle = [\alpha c(\beta) - \tilde{\alpha}^* s(\beta)]\|\alpha;\beta\rangle\rangle, \tag{6.8a}$$

$$\tilde{\xi}\,\|\alpha;\beta\rangle\rangle = [\tilde{\alpha} c(\beta) - \alpha^* s(\beta)]\|\alpha;\beta\rangle\rangle. \tag{6.8b}$$

Since we know the relationship (6.7) between these two definitions, we concentrate ourselves on the first definition without sacrificing the generality—see the remark below Eq. (6.13) and in the Appendix.

Let us require that the TCS is a state with the minimum uncertainty in the GUR inequality (5.14). Then we immediately find that (6.2) satisfies this requirement. It is easy to show that the state (6.2) satisfies the conditions (i) and (ii) with a constraint

$$\tilde{\alpha} = \alpha^*, \tag{6.9}$$

which is due to (ii).

Now we evaluate the characteristic function (CF) of the TCS,

$$\langle \mathrm{CF}\rangle_\alpha \equiv \langle\alpha;\beta|\exp\{-i(qQ + pP + \tilde{q}\tilde{Q} + \tilde{p}\tilde{P})\}|\alpha;\beta\rangle, \tag{6.10}$$

where the capital letters represent the operators given by (5.2) while the small letters stand for c numbers. Direct calculation (see the Appendix) gives us

$$\langle \mathrm{CF}\rangle_\alpha = \exp[-(\Theta/4)\{\lambda(p^2+\tilde{p}^2) + (q^2+\tilde{q}^2)/\lambda\} + \Lambda(\lambda p\tilde{p} + q\tilde{q}/\lambda) - iR] \tag{6.11}$$

with $\hbar = 1$ and

$$\Theta = \coth(\beta\omega/2), \tag{6.12a}$$

$$\Lambda = \tfrac{1}{2}\operatorname{cosech}(\beta\omega/2), \tag{6.12b}$$

$$R = q\langle Q\rangle_\alpha + p\langle P\rangle_\alpha + \tilde{q}\langle\tilde{Q}\rangle_\alpha + \tilde{p}\langle\tilde{P}\rangle_\alpha, \tag{6.12c}$$

where

$$\langle Q\rangle_\alpha \equiv \langle\alpha;\beta|Q|\alpha;\beta\rangle = (1/\sqrt{2\lambda})[(\alpha^*+\alpha)c(\beta) + (\tilde{\alpha}+\tilde{\alpha}^*)s(\beta)], \tag{6.13a}$$

$$\langle P\rangle_\alpha = i(\sqrt{\lambda/2})[(\alpha^*-\alpha)c(\beta) + (\tilde{\alpha}-\tilde{\alpha}^*)s(\beta)]. \tag{6.13b}$$

[Remember (6.9).] The corresponding values for the tilde operators are obtained upon exchanging $c(\beta)$ with $-s(\beta)$ and vice versa. It is obvious from (6.10) that (6.13) are the expectation values of Q, P, and their tilde conjugates.

We can make the same analysis by using the TCS of the second type [(6.3) and (6.8)], for which the expectation values will be denoted by the double bracket $\langle\langle\cdots\rangle\rangle_\alpha$. Thus we calculate $\langle\langle \mathrm{CF}\rangle\rangle_\alpha$. The result can be put in the form of (6.11). The values of Θ and Λ remain the same as those in (6.12a) and (6.12b). However, the expectation values $\langle\langle Q\rangle\rangle_\alpha$, $\langle\langle P\rangle\rangle_\alpha$, and their tilde conjugates become independent of β—see (A9) in the Appendix. It is seen that the two definitions are simply related. We prefer to use the first definition (6.2). We note that the TCS (6.2) saturates the GUR inequality (5.14).

The CF leads to the dispersion

$$\langle(\Delta P)^2\rangle_\alpha\langle(\Delta Q)^2\rangle_\alpha = \Theta^2/4, \tag{6.14}$$

$$\langle\Delta P\,\Delta\tilde{P}\rangle_\alpha\langle\Delta Q\,\Delta\tilde{Q}\rangle_\alpha = \Lambda^2. \tag{6.15}$$

In the limit of zero temperature, the thermal (tildian) degrees of freedom decouple from the dynamical (nontildian) ones:

$$\Theta^2/4 \to 1/4, \quad \text{as } \beta\to\infty. \tag{6.16a}$$

$$\Lambda \to 0, \tag{6.16b}$$

In the limit of high temperature we obtain

$$\lim_{\beta\to 0}(\Theta^2/4\Lambda^2) = 1. \tag{6.17}$$

For arbitrary β, in accordance with (5.14), we have

$$\Theta^2/4 - \Lambda^2 = 1/4. \tag{6.18}$$

Note that the rhs of (6.18) is temperature independent while Θ and Λ depend on β.

We now consider the TCS as defined by Emch and Hegerfeldt[9] (EH) from a TFD viewpoint. EH define a state via its CF. Their definition of the CF for the TCS did not consider the tilde operators. Further, having $\langle Q\rangle$ and $\langle P\rangle$, but not $\langle\tilde{Q}\rangle$ and $\langle\tilde{P}\rangle$, only partially defines α and $\tilde{\alpha}$ in (6.4). We will not provide an in-depth comparison between EH and the TCS of TFD. However, we would like to mention that the concept of the TCS in TFD is somewhat more stringent than EH's. To illustrate this we shall discuss briefly, within our formalism of one of their results. Given a product state of two particles, and their center of mass is known to be in a TCS. We seek solution to the operator equation (here the operators are temperature dependent)

$$f_A(a_R^\dagger)f_R(a_A^\dagger)|0(\beta_A)\rangle_A \otimes |0(\beta_R)\rangle_R = f_B(a_B^\dagger)f_C(a_C^\dagger)|0(\beta_B)\rangle_B \otimes |0(\beta_C)\rangle_C. \tag{6.19}$$

Here the two (independent) particles are labeled by B and C while the center of mass coordinates are denoted by A, and R are the relative coordinates. In writing (6.19) we used the information that if the particles state is a product (because they are independent) and A is known to be in a TCS, then the lhs of (6.19) is also a product. Further, we shall assume here (we prove it in a separate publication) that only a pure state (in TFD all states are pure states) solves this equation. What we wish to prove is that the relation (6.19) implies the equality of all four temperatures. [We assume $\omega_A = \omega_B = \omega_C = \omega_R$—differences in ω involve a Bogoliubov transformation that complicates the discussion—see (3.9) in Sec. III.] This proof is a particular case of the EH theorem [their Eq. (3.55)]. In giving this proof here we illustrate the use of the thermal transformation operator $U(\beta)$ (4.3). The proof is now simple: Since A represents the center of mass and R the relative coordinates we have from (3.10) with (3.9) that [$\mu_i = m_i/(m_B + m_C)$; $i = B,C$]

$$
\begin{aligned}
&|0(\beta_B)\rangle_B \otimes |0(\beta_C)\rangle_C \\
&\quad = \exp[\theta(\beta_B)(a_B^\dagger \tilde{a}_B^\dagger - \tilde{a}_B a_B) \\
&\qquad + \theta(\beta_C)(a_C^\dagger \tilde{a}_C^\dagger - \tilde{a}_C a_C)]\,|0(\infty)\rangle_B \otimes |0(\infty)\rangle_C,
\end{aligned} \tag{6.20}
$$

$$
|0(\infty)\rangle_B \otimes |0(\infty)\rangle_C = |0(\infty)\rangle_A \otimes |0(\infty)\rangle_R. \tag{6.21}
$$

Equation (6.21) is the usual relation when the actual ground state is not altered. Substituting for $a_B^\dagger$, $\tilde{a}_B^\dagger$, etc., in terms of $a_A^\dagger$, $\tilde{a}_A^\dagger$, etc. [cf. (3.10) in Sec. III], we obtain directly (we display only the creation operators—similar terms are present with annihilation operators) for the rhs of (6.20),

$$
\begin{aligned}
&\exp[\theta(\beta_B)\{\mu_B a_A^\dagger \tilde{a}_A^\dagger + \mu_C a_R^\dagger \tilde{a}_R^\dagger \\
&\quad - \sqrt{\mu_C \mu_B}\,(a_A^\dagger \tilde{a}_R^\dagger + \tilde{a}_A^\dagger a_R^\dagger)\} + \theta(\beta_C)\{\mu_C a_R^\dagger \tilde{a}_R^\dagger \\
&\quad + \mu_B a_A^\dagger \tilde{a}_A^\dagger + \sqrt{\mu_C \mu_B}\,(a_A^\dagger \tilde{a}_R^\dagger + \tilde{a}_A^\dagger a_R^\dagger)\}].
\end{aligned} \tag{6.22}
$$

This evidently reduces to a product of two thermal transformations iff $\beta_B = \beta_C$ and then obviously all the temperatures are equal. The demonstration of the above is, we believe, more transparent with TCS explicitly written rather than being implied via the CF. That the TCS in TFD is more stringent than TCS in EH formalism can also be seen by using the CF. In TFD there appear also tildian operators and the compatibility equations [Eqs. (3.47) of EH] have to be extended to include Λ, (6.12b), which can then be satisfied only at equal temperatures.

VII. SUMMARY AND CONCLUSIONS

This paper contains studies of two distinct but related aspects of coherent states (CS's). The first involves zero temperature CS's. Here, historically, the quality of the CS as a minimal uncertainty state was emphasized.[1] In this paper, however, we review the particular characteristic of the CS that was discovered by Aharonov *et al.* (AFLP).[8] There the emphasis is on the particular ("classical") property of CS, which allows it to be split into two independent states with no mutual correlation. It was pointed out that, at least in a finite volume, this split is in fact due to the CS wavefunction being factorizable in (at least) two different ways. We pointed out that this very property is the basis of the Emch and Hegerfeldt (EH) study[9] that pertains to CS. To illuminate the connection between these two works,[8,9] which was not stressed hitherto, the study of EH was rephrased in the language of field theory, and that of AFLP was corrected (when finite volume is considered) for a missing empty state. Both these modifications are somewhat forced to allow the connection between these works to be exhibited. The second topic considered in this paper is the extension of the study to finite temperatures. EH, in their formulation of the problem, were led to what they termed the thermal coherent state (TCS).

In this connection the theory of thermo field dynamics (TFD)[10] was shown to be a natural vehicle for the presentation and interpretation of the TCS. The TFD both the thermal ("tildian") and dynamical ("nontildian") degrees of freedom are treated on equal footing. Furthermore, TFD provides us with a simple relation between quantum and thermal fluctuations. In TFD fluctuations resulting from thermal or quantum origin are handled on an equivalent basis; we used this to show that TFD has a natural extension of the uncertainty relation at zero temperature. Recall that the latter puts a lower limit on the product of fluctuations in two noncommuting dynamical quantities. (We consider only P and Q in this paper). Now, basing ourselves on TFD, we have formulated the corresponding inequality for finite temperatures, which we termed the generalized uncertainty relation (GUR).[17] Then it turns out that the TCS by Barnett and Knight,[14,15] which is naturally introduced in TFD, is characterized as the state for which the GUR inequality is saturated. The characteristic function of the TCS was shown to be related to the one proposed by EH. The difference between the two is that the TCS of TFD contains information relating to the thermal degrees of freedom.

In a capsule: This paper establishes the relation between the work of AFLP and EH; the TCS of EH is generalized and is formulated within TFD; TFD is used to deduce an inequality for product of fluctuating noncommuting dynamical quantities, which generalizes the uncertainty relation to finite temperatures, and the TCS is shown to saturate this inequality.

In conclusion: The use of TFD that handles thermal and dynamical fluctuations on equal footing was shown to be almost indispensable for the study of thermal coherent states. In view of the wide use of coherent states it is hoped that thermal coherent states will prove useful.

ACKNOWLEDGMENTS

One of the authors (M. R.) acknowledges the kind hospitality of the Department of Physics at the University of Alberta.

This work was supported by NSERC, Canada, the Dean of Science at the University of Alberta, and by the Fund for Promotion of Research in the Technion.

APPENDIX: THE CHARACTERISTIC FUNCTIONS FOR THE TWO THERMAL COHERENT STATES

The three basic equations that we require are

$$
\begin{aligned}
D(\gamma,a)D(\alpha,a) &= \exp\{2i\,\mathrm{Im}(\gamma\alpha^* + \tilde{\gamma}\tilde{\alpha}^*)\} \\
&\quad \times D(\alpha,a)D(\gamma,a),
\end{aligned} \tag{A1}
$$

$$
U^\dagger(\beta)D(\gamma,a)U(\beta) = D(\gamma c - \tilde{\gamma}^* s, \xi), \tag{A2}
$$

$$
\langle 0(\infty)|D(\gamma,a)|0(\infty)\rangle = \exp[-(|\gamma|^2 + |\tilde{\gamma}|^2)/2]. \tag{A3}
$$

Here $D(\alpha,a)$ is given by (6.4); c and s by (4.6); and Im means imaginary part. The Weyl function,

$$\exp[-iqQ-ipP-i\tilde{q}\tilde{Q}-i\tilde{p}\tilde{P}]=D(\gamma,a), \tag{A4}$$

$$\gamma=p\sqrt{\gamma/2}-iq/\sqrt{2\gamma},\quad \tilde{\gamma}=-[\tilde{p}\sqrt{\gamma/2}+i\tilde{q}/\sqrt{2\gamma}]. \tag{A5}$$

To get (A5) we used (5.2).

We wish to evaluate the two expressions:

$$\langle CF\rangle_\alpha=\langle 0(\infty)|D^\dagger(\alpha,a)U^\dagger(\beta)\times D(\gamma,a)U(\beta)D(\alpha,a)|0(\infty)\rangle, \tag{A6}$$

$$\langle\langle CF\rangle\rangle_\alpha=\langle 0(\infty)|U^\dagger(\beta)D^\dagger(\alpha,a)\times D(\gamma,a)D(\alpha,a)U(\beta)|0(\infty)\rangle, \tag{A7}$$

where the first and second expressions are our choices for the TCS in (6.2) and (6.3), respectively. Using the basic formulas (A1)–(A3) yields

$$\langle CF\rangle_\alpha=\exp[-\tfrac{1}{2}\{(c^2+s^2)(|\gamma|^2+|\tilde{\gamma}|^2)\}-cs\{\gamma\tilde{\gamma}+\gamma^*\tilde{\gamma}^*\}+2i\,\mathrm{Im}\{(\gamma c-\tilde{\gamma}^*s)\alpha^*+(\tilde{\gamma}c-\gamma^*s)\tilde{\alpha}^*\}]. \tag{A8}$$

Using (A4)–(A6), (5.5) and (5.6) give (6.11). The $\langle\langle CF\rangle\rangle_\alpha$ is obtained from $\langle CF\rangle_\alpha$ by replacing α and $\tilde{\alpha}$ with $\alpha c-\tilde{\alpha}^*s$ and $\tilde{\alpha}c-\alpha^*s$, respectively. The difference between the two cases is the dependence of the expectation values on α and β (through c and s). For the first case $\langle Q\rangle$ and $\langle P\rangle$ are given by (6.13) whereas the expectation values for the second case are

$$\langle\langle Q\rangle\rangle_\alpha=(\alpha^*+\alpha)/\sqrt{2\lambda};\quad \langle\langle P\rangle\rangle_\alpha=i(\alpha^*-\alpha)\sqrt{\lambda/2}. \tag{A9}$$

In both cases the values of the tilde operators can be deduced from the above upon replacing $\alpha\leftrightarrow\tilde{\alpha}^*$ and an overall sign for the $\langle P\rangle$'s.

[1]E. Schrödinger, Naturwiss. **14**, 664 (1926).
[2]I. R. Senitzki, Phys. Rev. **95**, 904 (1954).
[3]R. J. Glauber, Phys. Rev. **130**, 2529 (1963).
[4]A. Casher, D. Lurie, and M. Revzen, J. Math. Phys. **9**, 1312 (1968).
[5]A. M. Perelomov, Commun. Math. Phys. **21**, 41 (1971).
[6]F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[7]Y. Ohnuki and T. Kashiwa, Prog. Theor. Phys. **60**, 548 (1978).
[8]Y. Aharonov, D. Fallkoff, E. Lerner, and H. Pendelton, Ann. Phys. (NY) **39**, 498 (1966); referred to as AFLP.
[9]G. G. Emch and G. C. Hegerfeldt, J. Math. Phys. **27**, 2731 (1986); referred to as EH.
[10]The most original form of TFD and its applications in the form of Feynman diagram were presented in L. Leplae, F. Mancini, and H. Umezawa, Phys. Rep. C **10**, 151 (1974). The theory was formulated in a systematic form in the paper of Ref. 11. The review of TFD up to 1982 is found in H. Umezawa, H. Matsumoto, and M. Tachiki, *Thermo Field Dynamics and Condensed States* (North-Holland, Amsterdam, 1982). The more recent one is N. P. Landsman and Ch. G. van Weert, Phys. Rep. **145**, 143 (1987).
[11]Y. Takahashi and H. Umezawa, Collect. Phenom. **25**, 55 (1975).
[12]I. Ojima, Ann. Phys. (NY) **139**, 1 (1981).
[13]G. G. Emch, *Mathematical and Conceptual Foundations of 20th Century Physics*, Mathematical Studies (North-Holland, Amsterdam, 1984), Vol. 100.
[14]S. M. Barnett and P. L. Knight, J. Opt. Soc. Am. B **2**, 467 (1985).
[15]T. Garavaglia, Phys. Rev. A **38**, 4365 (1988).
[16]A. Mann and M. Revzen, Phys. Lett. A **134**, 273 (1989).
[17]A. Mann, M. Revzen, H. Umezawa, and Y. Yamanaka, "Relation between quantum and thermal fluctuations," Phys. Lett. A (in press).

$o(4,2)\times o(4,2)$ group theoretical expression of the interelectronic Coulomb potential

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

(Received 1 March 1989; accepted for publication 19 July 1989)

The interelectronic Coulomb interaction between two electrons is expressed in terms of the o(4,2) generators of each electron. The formulation allows an expansion in terms of scaled hydrogenic (also called Sturmian) states of each electron with respect to a common center if the total orbital angular momentum is different from zero. The formulation is exact in the limit where a dimensionless parameter goes to infinity. Numerical evaluation of matrix elements of the interelectronic potential between hydrogenic configurations illustrates the convergence with respect to this parameter.

I. INTRODUCTION

The purpose of this paper is to express the Coulomb potential between two electrons in terms of scaled hydrogenic states of each electron with respect to a common center [see Eqs. (9) and (10)].

It is well known[1] that the Lie algebra o(4,2) is realized by the following 15 monoelectronic generators:

$$\mathbf{l}=\mathbf{r}\wedge\mathbf{p},\quad \mathbf{a}=\tfrac{1}{2}\mathbf{r}p^2-\mathbf{p}(\mathbf{r}\cdot\mathbf{p})-\tfrac{1}{2}\mathbf{r},$$
$$\mathbf{b}=\tfrac{1}{2}\mathbf{r}p^2-\mathbf{p}(\mathbf{r}\cdot\mathbf{p})+\tfrac{1}{2}\mathbf{r},\quad \mathbf{g}=r\mathbf{p},$$
$$t_1=\tfrac{1}{2}(rp^2-r),\quad t_2=\mathbf{r}\cdot\mathbf{p}-i=rp_r,$$
$$t_3=\tfrac{1}{2}(rp^2+r).$$

As usual, r corresponds to the position of a particle with respect to a center and p corresponds to its impulsion. These generators are Hermitic with respect to the so-called $1/r$ scalar product[1] defined by

$$\langle e|f\rangle\equiv\int d^3r\left(\tfrac{1}{r}\right)e^*(\mathbf{r})f(\mathbf{r}).$$

The interelectronic Coulomb potential V between two electrons can easily be expressed in terms of the generators corresponding to each electron:

$$V\equiv 1/|\mathbf{r}'-\mathbf{r}|=1/|\mathbf{b}'-\mathbf{a}'-\mathbf{b}+\mathbf{a}|=1/|\mathbf{A}-\mathbf{B}|,\quad(1)$$

where

$$\mathbf{A}\equiv\mathbf{a}-\mathbf{a}',\quad \mathbf{B}\equiv\mathbf{b}-\mathbf{b}'.$$

II. EXPRESSION IN TERMS OF SCALED HYDROGENIC STATES

Introducing the operator

$$T_2\equiv t_2+t_2'$$

one obtains

$$[T_2,\mathbf{A}]=i\mathbf{B},\quad [T_2,\mathbf{B}]=i\mathbf{A}.$$

It follows that

$$\exp(i\beta T_2)A^2\exp(-i\beta T_2)/\cosh^2(\beta)$$
$$=(\mathbf{A}-\mathbf{B})^2+(1-\tanh(\beta))[-(\mathbf{A}-\mathbf{B})^2+A^2-B^2+(1-\tanh(\beta))B^2].\quad(2)$$

One therefore expects that the interelectronic interaction could be written as

$$V=\lim_{\beta\to\infty}V(\beta),\quad(3)$$

$$V(\beta)\equiv\cosh(\beta)\exp(i\beta T_2)(1/|A|)\exp(-i\beta T_2).\quad(4)$$

It remains to define $1/|A|$ more precisely. This can be done by considering the unitary irreducible representation (UIR) of o(4,2) associated with the present realization.

A UIR of o(4,2) is obtained[1] if the 15 generators introduced above act on the scaled hydrogenic states $|(n,l,m)\rangle$ defined in terms of the usual hydrogenic states $|n,l,m\rangle$ by

$$|(n,l,m)\rangle\equiv n\exp(i\log(n)t_2)|n,l,m\rangle.\quad(5)$$

The action of each of the 15 o(4,2) generators on these states can be found in Ref. 1. The space of the UIR is spanned by all the $|(n,l,m)\rangle$. These states are orthonormal with respect to the $1/r$ scalar product:

$$\langle(n_j,l_j,m_j)|(n_k,l_k,m_k)\rangle=\delta_{m_jm_k}\delta_{l_jl_k}\delta_{n_jn_k}.$$

Spin is never taken into account in this paper and the $1/r$ scalar product is always used. The addition of spin degree of freedom, together with the antisymmetrization procedure, can be made without difficulty. Decisive progress in the study of two electron atoms has been made[2,3] by considering the eigenstates of the square of the difference of the Runge–Lenz vectors of each electron. The operators a and a' introduced in Sec. I can be obtained from these Runge–Lenz vectors via a scaling transformation.[1] Therefore, the results obtained in Refs. 2 and 3 can directly be translated to the present situation and are now briefly outlined. Using the relations[1]

$$[\tfrac{1}{2}(l\pm a)_j,\tfrac{1}{2}(l\mp a)_k]=0,$$
$$[\tfrac{1}{2}(l\pm a)_j,\tfrac{1}{2}(l\pm a)_k]=i\epsilon_{jkl}\tfrac{1}{2}(l\pm a)_l,$$
$$(\tfrac{1}{2}(\mathbf{l}\pm\mathbf{a}))^2|(n,l,m)\rangle=((n^2-1)/4)|(n,l,m)\rangle,$$

it can be shown that the eigenvectors of the operator A^2 are

$$|(n,n',J_1,J_2,L,M)\rangle$$
$$\equiv\sum_{l,l'}|(n,l,n',l',L,M)\rangle\times[(2l+1)(2l'+1)(2J_1+1)(2J_2+1)]^{1/2}\times(-1)^{l'}\begin{Bmatrix}\tfrac{1}{2}(n-1)&\tfrac{1}{2}(n'-1)&J_1\\\tfrac{1}{2}(n-1)&\tfrac{1}{2}(n'-1)&J_2\\l&l'&L\end{Bmatrix}.\quad(6)$$

 0022-2488/89/122891-04$02.50

The quantum numbers L, M in Eq. (6) are associated with the total orbital angular momentum and its projection over some axis. In Eq. (6) the bracketed coefficients are $9-j$ symbols[4] and the two electron states on the rhs are defined in terms of one electron state by means of the SU(2) Clebsch–Gordan coefficients:

$$|(n,l,n',l',L,M)\rangle \equiv \sum_{m,m'} |(n,l,m)\rangle |(n',l',m')\rangle \times \langle l,m;l',m'|L,M\rangle .$$

For $J_1 = J_2$, the vectors defined in Eq. (6) are eigenvectors of the total parity operator P with the eigenvalue π equal to $(-1)^L$. In order to obtain eigenstates of the parity P for any J_1 and J_2 values one has to consider the change of basis:

$$|(n,n',(J_1J_2),L,\pm,M)\rangle \equiv (1/2)^{1/2}[|(n,n',J_1,J_2,L,M)\rangle \pm |(n,n',J_2,J_1,L,M)\rangle].$$

Then, using symmetries properties of the $9j$ coefficients one obtains

$$\begin{aligned}&|(n,n',(J_1J_2),L^{\pi},M)\rangle \\ &\equiv \sum_{l,l'} |(n,l,n',l',L,M)\rangle \\ &\times[(2l+1)(2l'+1)(2J_1+1)(2J_2+1)]^{1/2} \\ &\times(-1)^{l'}\begin{Bmatrix}\frac{1}{2}(n-1) & \frac{1}{2}(n'-1) & J_1\\ \frac{1}{2}(n-1) & \frac{1}{2}(n'-1) & J_2\\ l & l' & L\end{Bmatrix} \\ &\times C(l+l',\pi,J_1-J_2),\end{aligned} \tag{7}$$

where C is defined by

$$C(l+l',\pi,J_1-J_2) \equiv \delta_{\pi(-1)^{l+l'}} \times \begin{cases} 1, & \text{if } J_1-J_2=0,\\ 2^{1/2}, & \text{if } J_1-J_2\neq 0,\end{cases}$$

and δ is a Kronecker symbol. The two electron states $|(n,n',(J_1J_2),L^{\pi},M)\rangle$ for all possible values of $n,n',J_1\geqslant J_2,L,\pi,M$ provide a complete orthonormal basis and satisfy the eigenvalue equation

$$\begin{aligned}A^2|(n,n',(J_1J_2),L^{\pi},M)\rangle &= [2(J_1(J_1+1)+J_2(J_2+1))-L(L+1)] \\ &\times|(n,n',(J_1J_2),L^{\pi},M)\rangle.\end{aligned} \tag{8}$$

An examination of relation (8) shows that A^2 can have zero as an eigenvalue only for the case $L=0$. The maximum value of L for J_1, J_2 fixed is indeed J_1+J_2 since it follows from the occurrence of the $9j$ symbol in Eq. (7). Then the eigenvalue is minimum, equal to $(J_1-J_2)^2+J_1+J_2$, and can be zero only for $J_1=J_2=0$. However, the case $J_1=J_2=0$ can only be realized if $L=0$. Thus A^2 has an inverse in the subspace where L is different from zero. Therefore, one obtains the following expression for the interelectronic potential V acting in a two-electron space of fixed total parity, fixed total orbital angular momentum $L\neq 0$, and fixed total projection M:

$$V^{L\neq 0,M,\pi} = \lim_{\beta\to\infty} V^{L\neq 0,M,\pi}(\beta), \tag{9}$$

$$\begin{aligned}&V^{L\neq 0,M,\pi}(\beta) \\ &\equiv \cosh(\beta)\sum_{n,n'}\sum_{J_1\geqslant J_2}[2(J_1(J_1+1) \\ &+J_2(J_2+1))-L(L+1)]^{-1/2} \\ &\times\exp(i\beta T_2)|(n,n',(J_1J_2),L^{\pi},M)\rangle \\ &\times\langle(n,n',(J_1J_2),L^{\pi},M)|\exp(-i\beta T_2).\end{aligned} \tag{10}$$

Equation (10) has two basic advantages. First, it expresses the interelectronic interaction in terms of properly normalizable stationary states corresponding to two independent (i.e., noninteracting) particles. These stationary states correspond to the model case, where two particles of charge $-n$ and $-n'$, respectively, move without repulsive interaction in the field of the same positive charge equal to $\exp(\beta)$. Second, Eq. (10) involves only discrete summation and avoids the problem of handling the continuum explicitly. This is of particular interest when dealing with autoionizing atomic states.

The summations over n and n' in Eq. (10) go to infinity by step of unity. The range of variation for J_1, $J_2(J_1\geqslant J_2)$ is then determined by triangular conditions implicit in the $9j$ coefficients. If the order of summation is reversed, i.e., if summation over n and n' is performed first with the range of variation depending on J_1, J_2, it is seen that $V(\beta)$ acts as a multiple of the identity operator in the subspaces spanned by the $\exp(i\beta T_2)\,|(n,n',J_1,J_2,L,M)\rangle$ vectors for fixed values of the eigenvalues of A^2 [see Eq. (2)].

III. MATRIX ELEMENTS BETWEEN HYDROGENIC CONFIGURATIONS

In order to illustrate the convergence of $V(\beta)$ [Eq. (10)] with respect to the parameter β, a convenient procedure is to consider the matrix elements of V between hydrogenic configurations. These matrix elements can be expressed in terms of the so-called Slater integrals.[4] Approximate o(4,2) expressions for these matrix elements have been proposed in Ref. 5. Recalling that the $1/r$ scalar product is used throughout this work, one has to consider the following M-independent matrix elements:

$$\begin{aligned}X^L(a,b,c,d) &\equiv \langle n_a,l_a,n_b,l_b,L,M|r'rV|n_c,l_c,n_d,l_d,L,M\rangle \\ &=(-1)^{l_b+l_d+L} \\ &\times[(2l_a+1)(2l_b+1)(2l_c+1)(2l_d+1)]^{1/2} \\ &\times\sum_k \begin{Bmatrix}L & l_b & l_a\\ k & l_c & l_d\end{Bmatrix}\begin{pmatrix}l_c & k & l_a\\ 0 & 0 & 0\end{pmatrix}\begin{pmatrix}l_d & k & l_b\\ 0 & 0 & 0\end{pmatrix} \\ &\times R(k,ab,cd).\end{aligned} \tag{11}$$

The $6j$ and $3j$ symbols originate from angular integration and the Slater integral R is defined in terms of radial hydrogenic wavefunctions by

$$\begin{aligned}&R(k,ab,cd) \\ &=\int r^2\,dr\int r'^2\,dr'\left(\frac{(r\langle)^k}{(r\rangle)^{k+1}}\right)R_{n_a,l_a}(r) \\ &\times R_{n_b,l_b}(r')R_{n_c,l_c}(r)R_{n_d,l_d}(r')\end{aligned}$$

These radial integrals can be evaluated analytically and some numerical values have been tabulated.[6] Using Eq. (5), together with the relations

$$\exp(i\theta t_2) r \exp(-i\theta t_2) = \exp(\theta) r, \qquad (12)$$

$$\begin{aligned} r|(n,l,m)\rangle &= -\tfrac{1}{2}[(n+l)(n-l-1)]^{1/2} \\ &\quad \times |(n-1,l,m)\rangle + n|(n,l,m)\rangle \\ &\quad -\tfrac{1}{2}[(n-l)(n+l+1)]^{1/2}|(n+1,l,m)\rangle, \end{aligned} \qquad (13)$$

$$\langle (n_a,l_a,m_a)|\exp(-i\theta t_2)|(n_b,l_b,m_b)\rangle = \delta_{l_a l_b}\delta_{m_a m_b}\delta^{-l_a-1}_{n_a,n_b}(\theta), \qquad (14)$$

where the first two delta symbols are Kronecker symbols and the latter, with upper index $-l_a-1$, denotes SU(1,1) representation functions[7–9] as defined in Ref. 10, one finally obtains

$$\begin{aligned} X^L(\beta,a,b,c,d) &\equiv \langle n_a,l_a,n_b,l_b,L,M|rr'V(\beta)|n_c,l_c,n_d,l_d,L,M\rangle \\ &= (-1)^{l_b+l_d}[(2l_a+1)(2l_b+1)(2l_c+1)(2l_d+1)]^{1/2}(n_c n_d)^{-1} \\ &\quad \times \cosh(\beta)\sum_{n,n'}[f(n,n_a,\beta_a)f(n',n_b,\beta_b) + n_b f(n,n_a,\beta_a) + n_a f(n',n_b,\beta_b) + n_a n_b] \\ &\quad \times \delta^{-l_a-1}_{n,n_a}(\beta_a)\delta^{-l_b-1}_{n',n_b}(\beta_b)\delta^{-l_c-1}_{n,n_c}(\beta_c)\delta^{-l_d-1}_{n',n_d}(\beta_d)F(n,n',l_a,l_b,l_c,l_d,L), \end{aligned} \qquad (15)$$

where

$$f(n,n_a,\beta_a) \equiv (n - n_a\cosh(\beta_a))/\sinh(\beta_a),$$

$$\begin{aligned} F(n,n',l_a,l_b,l_c,l_d,L) &\equiv \sum_{J_1>J_2}[2(J_1(J_1+1)+J_2(J_2+1)) - L(L+1)]^{-1/2} \\ &\quad \times (2J_1+1)(2J_2+1)\begin{Bmatrix} \tfrac{1}{2}(n-1) & \tfrac{1}{2}(n'-1) & J_1 \\ \tfrac{1}{2}(n-1) & \tfrac{1}{2}(n'-1) & J_2 \\ l_a & l_b & L \end{Bmatrix}\begin{Bmatrix} \tfrac{1}{2}(n-1) & \tfrac{1}{2}(n'-1) & J_1 \\ \tfrac{1}{2}(n-1) & \tfrac{1}{2}(n'-1) & J_2 \\ l_c & l_d & L \end{Bmatrix} \\ &\quad \times C(l_a+l_b,\pi,J_1-J_2)C(l_c+l_d,\pi,J_1-J_2), \quad \beta_a \equiv \beta + \log(n_a), \end{aligned}$$

etc. for the indices b, c, d.

In the derivation of Eq. (15) use has been made of a three-term recursion relation satisfied by the SU(1,1) representation functions[10]

$$\begin{aligned} 2\delta^{\gamma}_{\mu',\mu}(\beta)f(\mu,\mu',\beta) &= [(\gamma-\mu'+1)(-\mu'-\gamma)]^{1/2}\delta^{\gamma}_{\mu'-1,\mu}(\beta) \\ &\quad + [(\gamma+\mu'+1)(\mu'-\gamma)]^{1/2}\delta^{\gamma}_{\mu'+1,\mu}(\beta). \end{aligned}$$

The phase convention implicit in this relation is compatible with the one implicit in Eq. (13).

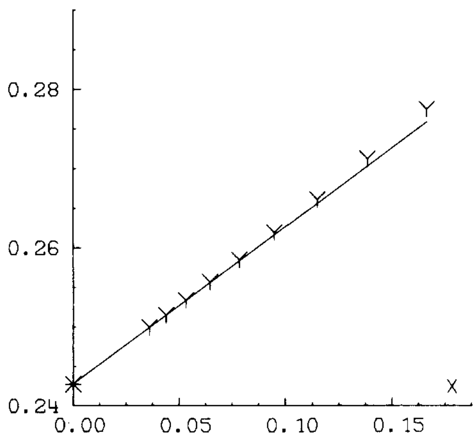


FIG. 1. The direct matrix element [Eqs. (15) and (16)] as a function of $X = 1 - \tanh(\beta)$. The straight line goes through the two points with the largest β values (see text). The asterisk corresponds to the exact Coulombic value.

Numerical applications have been made for two matrix elements $X_L(\beta,a,b,c,d)$ corresponding to the lowest hydrogenic configurations with $L = 1$:

$$n_a = n_c = 1, \quad l_a = l_c = 0 \quad n_b = n_d = 2, \quad l_b = l_d = 1, \qquad (16)$$

$$n_a = n_d = 1, \quad l_a = l_d = 0, \quad n_b = n_c = 2, \quad l_b = l_c = 1. \qquad (17)$$

The first matrix element will be called direct and the second will be called exchange. It should be noted that the number of values for n, n' which must be taken effectively into account in Eq. (15) increases very rapidly with β. The reason

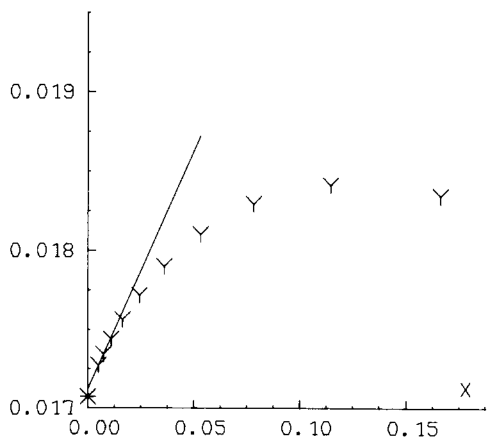


FIG. 2. Same as in Fig. 1, but for the exchange matrix element [Eqs. (15) and (17)] (see text).

is that the classical domain for SU(1,1) representation functions δ [see Eq. (14)] increases exponentially with β.[10] In practical calculations the convergence with respect to the summation over n, n' has been tested by using the unitarity condition

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

The exact Coulombic values deduced from Eq. (11) and Ref. 6 are 0.2427984... and 0.01707057..., respectively. For $\beta = 2$ one obtains 0.2500, 0.0179, respectively. It is more instructive to consider the β dependence. The above direct and exchange matrix elements [Eqs. (15)–(17)] have been reported as a function of $X = 1 - \tanh(\beta)$ in Figs. 1 and 2, respectively. It is seen in Fig. 1 that an almost linear dependence is rapidly obtained as X decreases. The straight line in Fig. 1 goes through the two points of smallest X values corresponding to $\beta = 2$ and 1.9. A linear extrapolation yields the value 0.2428. It is seen in Fig. 2 that the β dependence is not so simple for the exchange case. A linear extrapolation from the two points corresponding to the greatest values of β ($\beta = 3,2.8$) yields the value 0.0171. A graphical or more sophisticated extrapolation procedure would further improve the agreement. It should be noted that an accurate calculation for $\beta = 3$ becomes difficult and time consuming since n values up to 150 have been involved. As used here the above method is certainly not the best for computing the matrix elements defined by Eq. (11): The above calculations only illustrate the behavior with respect to β. Numerical studies corresponding to the case of finite β values may be helpful as a model for three-body problems; however, they become difficult as β increases.

IV. CONCLUSION

The interest of the present formulation is mainly that it provides a continuous set of potentials $V(\beta)$ which can be expanded in terms of properly normalizable states [Eq. (10)]. If restricted to subspaces of $L \neq 0$, this set converges toward the Coulomb potential as β goes to infinity. These potentials can be expressed in terms of $o(4,2) \times o(4,2)$ generators corresponding to the noninvariance algebra of two noninteracting electrons in the field of the same positive charge. This provides a new starting point for a better understanding of approximate symmetries in atomic structures. It should be of particular interest for the study of diexcited states of two-electron atomic systems.

ACKNOWLEDGMENTS

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

[1] B. G. Adams, J. Cizek, and J. Paldus, Adv. Quant. Chem. **19**, 1 (1988) and references therein.
[2] C. Wulfman, Chem. Phys. Lett. **23**, 370 (1973).
[3] O. Sinanoglu and D. R. Herrick, J. Chem. Phys. **62**, 886 (1975).
[4] A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U. P., Princeton, 1974).
[5] C. Wulfman and S. Kumei, Chem. Phys. Lett. **23**, 367 (1963).
[6] P. H. Butler, P. E. H. Minchin, and B. G. Wybourne, Atom. Data **3**, 153 (1971).
[7] V. Bargmann, Ann. Math. **48**, 568 (1947).
[8] W. J. Holman and L. C. Biedenharn, Ann. Phys. (NY) **39**, 1 (1966).
[9] W. J. Holman and L. C. Biedenharn, Ann. Phys. (NY) **47**, 205 (1968).
[10] E. de Prunelé, J. Math. Phys. **29**, 2523 (1988).

Gravitational analogs of the Aharonov–Bohm effect

V. B. Bezerra[a)]
Center for Theoretical Physics, Laboratory for Nuclear Science and Department of Physics, Massachusetts Institute of Technology, Cambridge, Massachusetts 02139

(Received 21 February 1989; accepted for publication 28 June 1989)

A quantum scalar particle is considered in the following background gravitational fields due to (a) a tubular matter source with axial interior magnetic field and vanishing exterior magnetic field; (b) slowly moving mass currents (weak approximation); and (c) a spinning cosmic string. It is shown that in the flat space-time around these sources, the energy spectrum and wave function of the particle depend on the amount of matter and magnetic field (tubular matter source case), on the velocity of the moving mass currents, and on the angular momentum in the spinning cosmic string case. These represent gravitational analogs of the Aharonov–Bohm effect in electrodynamics and are due to global (topological) features of the background space-times under consideration.

I. INTRODUCTION

In classical electromagnetism, a charged particle is influenced only by the electric and magnetic fields at the location of the particle. At the quantum level the behavior of a charged particle is modified by the action of an external magnetic field[1] confined to a region from which the charged particle is excluded. This nonlocal (global) phenomenon of the enclosed magnetic flux on the charged particle is the well-known Aharonov–Bohm effect,[1] which can be understood as a manifestation of gauge theories. As a consequence of this phenomenon the field strength underdescribes electromagnetism in situations where global aspects (nontrivial topology) are present, and an intrinsic and complete description of electrodynamics in a space-time region is given[2] only in terms of a nonintegrable phase factor

$$\exp\left\{i\left(\frac{e}{\hbar c}\right)\oint_C A_\mu\, dx^\mu\right\}$$

evaluated over all trajectories lying in a space-time region of nontrivial topology accessible to the quantum system. So, there is a linking between the description of the theory in terms of the phase factor and the topological feature of the background space-time.

In a metric theory of gravitation, a gravitational field is frequently related to a nonvanishing Riemann curvature tensor. The gravitational analog of the well-known electromagnetic Aharonov–Bohm effect is the following: particles constrained to move in a region where the Riemann curvature tensor vanishes may exhibit a gravitational effect arising from a region of nonzero curvature from which they are excluded, or, in a more general sense, particles constrained to move in a region where the Riemann curvature does not vanish but does not depend on certain parameters such as velocity (in the example of a weak gravitational field arising from moving mass currents) may exhibit a gravitational effect associated with this parameter. In the case of bound states of a quantum system interacting with a gravitational field the analog of the Aharonov–Bohm effect is the following: the energy spectrum of a quantum particle living in a region where the Riemann curvature tensor vanishes suffers an alteration that arises from the region of nonzero curvature from which it is excluded, or the energy spectrum as well as the wave function can depend on a parameter that does not contribute to the curvature in the region where the particle lives.

Effects analogous to the electromagnetic Aharonov–Bohm exist in classical theories like the Sagnac effect in general relativity[3] which consists of a phase shift between two beams of light traversing in opposite directions the same path around a rotating mass distribution. In the gravitational case numerous analogies at the classical level have been discussed,[4] as well in the case of gravitational interaction on a quantum mechanical system.[5]

In this paper we study the influence of external gravitational fields on the bound states of a quantum scalar particle. We consider the influence of the external gravitational field due to a tubular matter source with an axial interior magnetic field and vanishing exterior magnetic field corresponding to a particular model given by Safko and Witten,[6] which is a solution of the combined Einstein–Maxwell field with cylindrical symmetry. Also, we will consider the gravitational field associated with slowly moving mass currents[7] and to a spinning cosmic string.[8,9]

All of the effects presented in this paper are of topological origin; therefore we can use the phase factor mentioned before, with an appropriate connection, in order to understand these effects. In the cases under consideration, because of the nontrivial topology of the underlying space-time manifolds, which introduces a discontinuity in the connection, it is not necessarily true that the phase factors corresponding to a curve encircling the sources, in the different cases, are equal to the identity.

In the Safko–Witten model[6] the space-time is locally flat and the Riemann curvature tensor vanishes everywhere outside the tube of matter, but we have a gravitational effect on the particle that lives in this flat region which comes from the region where the curvature does not vanish.

In the moving mass current case there is no influence of

a) Permanent address: Departamento de Física, Centro de Ciencias Exatas e da Natureza, Universidade Federal da Paraíba, Cidade Universitária, 58000 João Pessoa, Pb-Brazil.

 0022-2488/89/122895-05$02.50

the velocity on the Riemann curvature tensor in the weak field approximation, but we have a gravitational effect associated with the velocity of matter on a particle moving outside the tube of matter. The space-time surrounding the spinning cosmic string is locally flat everywhere; it is conical with a twist in the time direction. In this case there is a dependence of the energy and of the wave function on the angular momentum, as well as on the mass density of the string.

For moving mass currents the gravitational effect is taken in the weak field approximation and for the spinning cosmic string when $J^2 \sim 0$. It is interesting to observe that in all the cases a local influence of the space-time curvatures is absent, and all the effects come out from global (topological) properties of the space-times associated with the sources under consideration. So, the attributes of the source—mass, angular momentum, etc.—are coded in the global properties of the locally flat variables.

Let us consider a scalar quantum particle imbedded in a classical background gravitational field. Its behavior is described by the covariant Klein–Gordon equation

$$[(1/\sqrt{-g})\partial_\mu(\sqrt{-g}g^{\mu\nu}\partial_\nu) + M^2]\psi = 0, \tag{1.1}$$

where M is the mass of the particle and $\hbar = c = 1$ units are chosen. All space-times that will be considered here are time independent, so the time dependence of the wave function that solves Eq. (1.1) may be separated as e^{-iEt} and one is led to a stationary problem at fixed energy E. Moreover, rotational invariance and invariance along the z axis of the metrics allow us to separate the φ and z dependences. In view of these we choose the solutions of Eq. (1.1), $\psi(t,\rho,\varphi,z)$, in the form

$$\psi(t,\rho,\varphi,z) = \exp(-iEt + il\varphi + ikz)R(\rho), \tag{1.2}$$

where E, l, and k are constants.

II. SCALAR PARTICLE IN THE GRAVITATIONAL FIELD OF A TUBULAR MATTER SOURCE WITH INTERIOR AXIAL MAGNETIC FIELD

We will consider a solution of the combined Einstein–Maxwell field with cylindrical symmetry, corresponding to a tubular matter source with axial interior magnetic field and vanishing exterior magnetic field.[6] The exterior space-time corresponding to this configuration of fields is locally flat with nontrivial topology. It is conical with deficit angle $2\pi e^{-\beta}$, where β is a quantity related to the interior magnetic field and the mass of the tube. The line element corresponding to this case is given by[6]

$$ds^2 = e^{2\beta}(dt^2 - d\rho^2) - \rho^2 d\varphi^2 - dz^2, \tag{2.1}$$

where

$$\beta = \ln\left[\frac{(1+\mathscr{K}_i\rho_1^2)^2}{\rho_2}\right] + \frac{4\mathscr{K}_i\rho_1}{1+\mathscr{K}_i\rho_1^2}\frac{(\rho_2-\rho_1)}{(\gamma+1)},$$

and depends on the intensity of the interior magnetic field through $\mathscr{K}_i$ and on the mass. The quantities ρ_1 and ρ_2 are the interior and exterior radius of the tube of matter and γ is an arbitrary constant.

It is easy to see, by using an appropriate changing of variables, that the space-time given by metric (2.1) is everywhere locally flat. The cross section $z = \text{const}$ is topologically equivalent to a cone.

In the space-time corresponding to a tubular matter source with an axial interior magnetic field, the Klein–Gordon equation [Eq. (1.1)] takes the form

$$\begin{aligned}&\Big\{\partial_t^2 - \frac{1}{\rho}\partial_\rho(\rho\,\partial_\rho) - \frac{1}{\rho^2}e^{2\beta}\\ &\quad\times[\partial_\varphi^2 + (\partial_z^2 - M^2)\rho^2]\Big\}\psi = 0.\end{aligned} \tag{2.2}$$

Using the ansatz given by Eq. (1.2), Eq. (2.2) reduces to

$$\begin{aligned}&\{\rho\,\partial_\rho(\rho\,\partial_\rho)\\ &\quad + [E^2 - e^{2\beta}(k^2+M^2)]\rho^2 - l^2e^{2\beta}\}R(\rho) = 0.\end{aligned} \tag{2.3}$$

Equation (2.3) is a Bessel differential equation with the general solution given by

$$R_{\nu k}(\rho) = C_{\nu k}^{(1)}J_{|\nu|}(\lambda\rho) + C_{\nu k}^{(2)}N_{|\nu|}(\lambda\rho), \tag{2.4}$$

where $\lambda^2 = E^2 - e^{2\beta}(k^2+M^2)$, $\nu = le^{\beta}$, $C_{\nu k}^{(1)}$ and $C_{\nu k}^{(2)}$ are normalization constants, and $J_{|\nu|}(\lambda\rho)$ and $N_{|\nu|}(\lambda\rho)$ are Bessel functions of the first and second kind, respectively.

We assume that the scalar quantum particle is restricted to move in a region outside the tubular matter source bounded by the cylindrical surfaces $\rho = a$ and $\rho = b$, where $b > a > \rho_2$ (external radius of the tube). The boundary conditions

$$R(a) = R(b) = 0 \tag{2.5}$$

determine the energy levels of the particle in the stationary state between the cylindrical surfaces $\rho = a$ and $\rho = b$. This condition yields the following equation for the energy spectrum of the particle:

$$J_{|\nu|}(\lambda a)N_{|\nu|}(\lambda b) - J_{|\nu|}(\lambda b)N_{|\nu|}(\lambda a) = 0. \tag{2.6}$$

Using Hankel's asymptotic expansion[10] when ν is fixed, $\lambda a \gg 1$, and $\lambda b \gg 1$, we get

$$\begin{aligned}J_{|\nu|}(\lambda a) \sim \sqrt{\frac{2}{\pi\lambda a}}\Big[&\cos\Big(\pi\lambda a - \frac{\nu}{2}\pi - \frac{\pi}{4}\Big)\\ &- \frac{4\nu^2-1}{8\lambda a}\sin\Big(\lambda a - \frac{\nu}{2}\pi - \frac{\pi}{4}\Big)\Big],\end{aligned} \tag{2.7}$$

$$\begin{aligned}N_{|\nu|}(\lambda a) \sim \sqrt{\frac{2}{\pi\lambda a}}\Big[&\sin\Big(\lambda a - \frac{\nu}{2}\pi - \frac{\pi}{4}\Big)\\ &+ \frac{4\nu^2-1}{8\lambda a}\cos\Big(\lambda a - \frac{\nu}{2}\pi - \frac{\pi}{4}\Big)\Big],\end{aligned} \tag{2.8}$$

and similar expressions for $J_{|\nu|}(\lambda b)$ and $N_{|\nu|}(\lambda b)$ with b interchanged for a. Putting Eqs. (2.7) and (2.8) in the condition given by Eq. (2.6), we get

$$\lambda^2 \sim (n\pi/(b-a))^2 + (4l^2e^{2\beta} - 1)/4ab, \tag{2.9}$$

where we used the fact that $\nu = le^{\beta}$. Remembering that $\lambda^2 = E^2 - e^{2\beta}(k^2+M^2)$, we get from Eq. (2.9) that

$$E = e^{\beta}\sqrt{M^2 + k^2 + \frac{l^2}{ab} + \frac{[4ab(n\pi)^2 - (b-a)^2]e^{-2\beta}}{4ab(b-a)^2}}, \tag{2.10}$$

where we have taken only the positive sign.

From Eq. (2.10) we see that when $b \to a$ (particle moving on the cylindrical surface), $E \to \infty$, so that to get the limit

$E \to$ const, we have to introduce an attractive potential in the cylindrical hollow ($a < \rho < b$) to compensate for the increasing of the energy of the radial modes in this limit. Doing this we get

$$E = e^{\beta} E^{(0)}, \tag{2.11}$$

where

$$E^{(0)} = \sqrt{M^2 + k^2 + l^2/a^2}.$$

From Eq. (2.11) we see that the energy spectrum depends on the factor e^{β} (as well as the wave function) relative to the Minkowski case. But outside the tubular matter source the space-time is locally flat; the Riemann curvature tensor vanishes everywhere. So, the fact that this space-time is locally flat but not globally (it is conical with deficit angle $2\pi e^{-\beta}$) deforms the energy spectrum proportionally to e^{β}, which is the same factor that affects the height of the centrifugal barrier. This is a gravitational analog of the electromagnetic bound state Aharonov–Bohm effect in which the energy eigenvalues depend upon the magnetic flux inside the solenoid, which is inaccessible to the charged particle.

The previous result can be extended to the cosmic string solution[11] as well as to the point[9] particle solution in $(2+1)$-dimensional gravity. It is sufficient to identify formally $e^{-\beta}$ with $\alpha = 1 - 4\mu$ (where μ is the mass density of the string) and make an appropriate changing of variables to put the metric corresponding to the previous case in the cosmic string form. So, doing this we get for the energy of the scalar particle moving on a cylindrical surface $\rho = a$, in the space-time of a cosmic string, the expression

$$E = \sqrt{M^2 + k^2 + l^2/\alpha^2 a^2}, \tag{2.12}$$

which gives us the dependence of the energy spectrum on $\alpha = 1 - 4\mu$.

If one restricts to the cross section $z =$ const of the cosmic string solution we get the point particle solution. For this case the energy spectrum is given by

$$E = \sqrt{M^2 + \frac{l^2}{\alpha^2 a^2}},$$

where now $\alpha = 1 - 4m$, where m is the mass of the particle. The particle is restricted to moves on the circle $\rho = a$.

III. SCALAR PARTICLE IN THE GRAVITATIONAL FIELD OF MOVING MASS CURRENTS

As a source of gravitational field, we consider a cylindrical distribution of matter with uniform density along the z axis, which moves slowly with velocity $v(v^2 \sim 0)$ in the z direction. The z extension of matter is taken to be much larger than the other lengths of the distribution. The line element corresponding to this distribution of matter is given, in weak field approximation, by[7]

$$ds^2 = (1 - \Phi)dt^2 - (1 + \Phi)(d\rho^2 + \rho^2 d\varphi^2 + dz^2) + 4v\Phi\, dt\, dz, \tag{3.1}$$

where Φ represents the Newtonian potential produced by the distribution of matter. Since the gravitational field is sufficiently weak, we can write the metric tensor corresponding to the line element given by Eq. (3.1) as

$$g_{\mu\nu} = \eta_{\mu\nu} + h_{\mu\nu}, \tag{3.2}$$

where $\eta_{\mu\nu} = \eta^{\mu\nu} = \mathrm{diag}(1, -1, -1, -1)$ and $h_{\mu\nu}$ is given by the obvious identifications. In this case the Riemann tensor is

$$R_{\alpha\mu\beta\nu} = \tfrac{1}{2}(h_{\alpha\nu,\mu\beta} + h_{\mu\beta,\alpha\nu} - h_{\mu\nu,\alpha\beta} - h_{\alpha\beta,\mu\nu}). \tag{3.3}$$

We can see from Eq. (3.3) that the curvature outside the distribution of matter does not depend on its velocity in the considered linear approximation. This means that for the weak gravitational field associated with slowly moving mass current, the local effects of curvature associated with the velocity of the distribution of matter are absent outside it.

Assume that the massive scalar quantum particle is restricted to move in a region outside the distribution of matter and bounded by the cylindrical surfaces $\rho = \rho_1$ and $\rho = \rho_2$ ($\rho_2 > \rho_1 >$ radius of the distribution of matter).

In the space-time corresponding to moving mass currents, the Klein–Gordon equation [Eq. (1.1)] takes the form

$$[(1 + 2\Phi)\partial_t^2 - (1/\rho)\partial_\rho(\rho\, \partial_\rho) - (1/\rho^2)\partial_\rho^2 - 4v\Phi\, \partial_t \partial_z - \partial_z^2 + M^2(1 + \Phi)]\psi = 0. \tag{3.4}$$

Using $\psi(t,\rho,\varphi,z)$ in the form given by Eq. (1.2), Eq. (3.4) reads

$$\{\rho\, \partial_\rho(\rho\, \partial_\rho) + [(1 + 2\Phi)E^2 + 4v\Phi kE - k^2 - M^2(1 + \Phi)]\rho^2 - l^2\}R(\rho) = 0. \tag{3.5}$$

The boundary conditions $R(\rho_1) = R(\rho_2) = 0$ determine the energy levels of the particles in the stationary states. As we can see from Eq. (3.5) these energy levels depend on the mass distribution and its velocity. To obtain the energy spectrum explicitly we will consider a narrow hollow for which ρ_1 and ρ_2 are close together so that the Newtonian potential can be considered as constant and equal to Φ_1 (Newtonian potential over the cylindrical surface $\rho = \rho_1$). So, in this case Eq. (3.5) turns to a Bessel differential equation with the general solution given by

$$R_{lk}(\rho) = C_{lk}^{(1)} J_{|l|}(\lambda\rho) + C_{lk}^{(2)} N_{|l|}(\lambda\rho), \tag{3.6}$$

where $C_{lk}^{(1)}$ and $C_{lk}^{(2)}$ are normalization constants, $J_{|l|}(\lambda\rho)$ and $N_{|l|}(\lambda\rho)$ are Bessel functions of the first and second kind, respectively, and

$$\lambda^2 = (1 + 2\Phi_1)E^2 + 4v\Phi_1 Ek - M^2(1 + \Phi_1) - k^2.$$

The boundary conditions $R(\rho_1) = R(\rho_2) = 0$ yield the following equation for the energy spectrum of the particle:

$$J_{|l|}(\lambda\rho_1)N_{|l|}(\lambda\rho_2) - J_{|l|}(\lambda\rho_2)N_{|l|}(\lambda\rho_1) = 0. \tag{3.7}$$

To get the energy spectrum we consider $\lambda\rho_1 \gg 1$ and $\lambda\rho_2 \gg 1$, and use the asymptotic expressions for the Bessel functions. From Hankel's asymptotic expansion[10] we obtain in the linear approximation

$$E_{lk} = E_{lk}^{(0)} - E_{lk}^{(0)}\Phi_1 + M^2\Phi_1/2E_{lk}^{(0)} - 2v\Phi_1 k, \tag{3.8}$$

where

$$E_{lk}^{(0)} = \sqrt{M^2 + k^2 + \frac{l^2}{\rho_1\rho_2} + \frac{4\rho_1\rho_2(n\pi)^2 - (\rho_2 - \rho_1)^2}{4\rho_1\rho_2(\rho_2 - \rho_1)^2}},$$

with n an integer that appears in the expression for λ and comes out from Hankel's asymptotic expansion and Eq. (3.7).

As we can see from the expression for $E_{lk}^{(0)}$ in the limiting case $\rho_1 = \rho_2$ the energy of the radial modes becomes infinitely large and the considered particle turns to a relativistic quantum one restricted to move on the cylindrical surface $\rho = \rho_1$. So, we can do our calculations in this case but it is necessary to introduce into the cylindrical hollow a potential that compensates for the increase of the energy of the radial modes as ρ_2 tends to ρ_1. By doing this we obtain the same expression for E_{lk} [Eq. (3.8)] where now

$$E_{lk}^{(0)} = \sqrt{M^2 + k^2 + l^2/\rho_1^2}.$$

In Eq. (3.8) the last term gives the influence of the velocity of the matter distribution on the energy spectrum of the particle. Observe the fact that the velocity of the moving mass deforms the energy spectrum in relation to the static case. Through the region of motion of the scalar particle, in the linear approximation, the Riemann curvature tensor does not depend on the velocity of the matter (in fact, it vanishes), but the energy of the particle as well as its wave function are influenced by this velocity. This alteration of the energy spectrum by a parameter (velocity) that does not affect the local curvature (in linear approximation) of the region of motion of the particle, represents a gravitational analog of the Aharonov–Bohm effect for bounds states in electrodynamics.

IV. SCALAR PARTICLE IN THE GRAVITATIONAL FIELD OF A SPINNING COSMIC STRING

The metric for a spinning cosmic string is a simple generalization[8] of the spinning particle solution found by Deser *et al.*[9] In ordinary cylindrical coordinates ($0 \leqslant \varphi \leqslant 2\pi$) it reads

$$ds^2 = (dt + 4J\,d\varphi)^2 - d\rho^2 - \alpha^2\rho^2\,d\varphi^2 - dz^2, \quad (4.1)$$

where $\alpha = 1 - 4\mu$, μ is the mass per unit length, J is the angular momentum of the string, and G is chosen equal to 1.

The Klein–Gordon equation in the field of a spinning cosmic string reads

$$\Big[\partial_t^2 - \frac{1}{\rho}\partial_\rho(\rho\,\partial_\rho) - \frac{1}{\alpha^2\rho^2}(4J\,\partial_t - \partial_\varphi)^2 - \partial_z^2 + M^2\Big]\psi = 0. \quad (4.2)$$

With $\psi(t,\rho,\varphi,z)$ given in the form of Eq. (1.2), we get, from Eq. (4.2),

$$\Big[\rho\,\partial_\rho(\rho\,\partial_\rho) + (E^2 - k^2 - M^2)\rho^2 - \Big(\frac{4JE + l}{\alpha}\Big)^2\Big]R(\rho) = 0. \quad (4.3)$$

This is a Bessel equation whose solution is given by

$$R_{lk}(\rho) = C_{lk}^{(1)} J_{|l+4JE|/\alpha}(\lambda\rho) + C_{lk}^{(2)} N_{|l+4JE|/\alpha}(\lambda\rho), \quad (4.4)$$

where $\lambda^2 = E^2 - k^2 - M^2$.

Assuming that the particle is restricted to move between the cylindrical surfaces $\rho = \rho_1$ and $\rho = \rho_2$, the boundary conditions $R(\rho_1) = R(\rho_2) = 0$ yield

$$J_{|l+4JE|/\alpha}(\lambda\rho_1) N_{|l+4JE|/\alpha}(\lambda\rho_2) - J_{|l+4JE|/\alpha}(\lambda\rho_2) N_{|l+4JE|/\alpha}(\lambda\rho_1) = 0. \quad (4.5)$$

As in the previous cases we consider $\lambda\rho_1 \gg 1$, $\lambda\rho_2 \gg 1$, and use Hankel's asymptotic expansion[11] for the Bessel functions, obtaining in the case $J^2 \sim 0$ and for a narrow hollow

$$E_{lk} = 4Jl/\rho_1\rho_2\alpha^2 + E_{lk}^{(0)} \quad (4.6)$$

where

$$E_{lk}^{(0)} = \sqrt{M^2 + k^2 + \frac{l^2}{\rho_1\rho_2\alpha^2} + \frac{4\rho_1\rho_2(n\pi)^2 - (\rho_2 - \rho_1)^2}{4\rho_1\rho_2(\rho_2 - \rho_1)^2}}.$$

From the expression for $E_{lk}^{(0)}$ we can see that it tends to infinity when ρ_2 tends to ρ_1. So in the case $\rho_1 = \rho_2$ we have a relativistic quantum rotator. To obtain a finite result for the energy we have to add a compensating term (potential) when we consider a particle restricted to move on the cylindrical surface $\rho = \rho_1$. With this procedure we obtain the expression given by Eq. (4.6), where

$$E_{lk}^{(0)} = \sqrt{M^2 + k^2 + l^2/\rho_1^2\alpha^2}$$

which corresponds to the energy of the relativistic motion of a particle of mass M on the cylindrical surface $\rho = \rho_1$.

From Eq. (4.6) we see that the energy levels depend on the angular momentum of the string, which is a quantity localized along the string. The space-time of the spinning string is flat outside it. This is like the Aharonov–Bohm effect in electrodynamics.

If we take $J = 0$ in the above results we obtain the energy spectrum corresponding to the cosmic string[11] case of Sec. II, as well as the results for the point particle solution,[9] if, in addition to this, we take the cross section $z = \text{const}$.

A similar result can be obtained with the infinite solenoid potential if we translate the electromagnetic situation into a gravitational one. Thus we may make the analogy in weak field approximation $e \leftrightarrow m$ and $A_\mu \leftrightarrow h_\mu = (h_{0i}, \frac{1}{2}h_{00})$. So, using this translation, the space-time corresponding to the potential $\mathbf{A} = (0,0,\Phi/\rho)$ is given by

$$ds^2 = dt^2 - d\rho^2 - \rho^2\,d\varphi^2 - dz^2 + 2\Phi\,d\varphi\,dt, \quad (4.7)$$

where $\Phi \equiv$ magnetic flux$/(2\pi/e)$.

Equation (4.7) is formally the same as Eq. (4.1) in the limit $J^2 \sim 0$ and for $\alpha = 1$. Therefore, proceeding as in the spinning cosmic string case we get the energy levels of a scalar quantum particle restricted to move on the cylindrical surface $\rho = \rho_1$, in the space-time given by Eq. (4.7). The result is

$$E_{lk} = 4\Phi l/\rho_1^2 + E_{lk}^{(0)}, \quad (4.8)$$

where

$$E_{lk}^{(0)} = \sqrt{M^2 + k^2 + l^2/\rho_1^2}.$$

From Eq. (4.8) we see that the energy levels depend on the magnetic flux, but the particle moves in a region inaccessible to this flux. In fact, it moves in a region where the Riemann curvature tensor is zero everywhere, in the weak approximation, but its energy levels are not the same as the ones in Minkowski space-time. They are affected by the inaccessible magnetic flux. This is the gravitational version of the Aharonov–Bohm effect in electrodynamics for the Aharonov–Bohm potential.

V. FINAL REMARKS

The effect of local influence of the gravitational field that arises from a tubular matter source with an interior magnetic field on the energy levels of a particle as well as on the wave function is absent. The same occurs in the case of moving mass currents in relation to the velocity. The velocity deforms the energy spectrum only globally. Analogously we can say the same in relation to the influence of the angular momentum in the case of a spinning cosmic string. In fact, these dependences come out because the allowed values of the height of the centrifugal barrier depend on $e^{-\beta}$, $v\Phi$, and J in the cases of a tubular matter source with interior magnetic field, moving mass currents, and spinning cosmic string, respectively.

The fact that the attributes of the sources—like mass, angular momentum, etc.—are coded in the global properties of the locally flat variables evidenciate the importance of the topological structure of the space-time in the description of the physics of a given system. A tubular matter source with interior axial magnetic field, a long cylinder of moving mass currents, in the weak field approximation, and a spinning cosmic string produce an effect analogous to the one produced by an infinite solenoid with magnetic flux inaccessible to the particle, on a charged particle.

Therefore all the effects here discussed represent gravitational analogs of the Aharonov–Bohm effect for bound states in electrodynamics.

We can understand all these effects in terms of the phase factors, as emphasized before, associated with a given connection corresponding to these space-times. In the Safko–Witten model[6] we can see the effect as a Yang–Mill-type one due to the conical defect $2\pi e^{-\beta}$, which induces a discontinuity in the connection. The situation is similar for scalar particles in the gravitational field due to moving mass currents. In this case the topological effect can be understood incorporating the velocity effects within the connection. The same analysis can be applied to the case of a scalar particle in the space-time of a spinning cosmic string and through the analogy $e \leftrightarrow m$ and $A_\mu \leftrightarrow h_\mu = (h_{0i}, \frac{1}{2}h_{00})$ to the case of a scalar particle in the space-time corresponding to an infinite solenoid potential.

ACKNOWLEDGMENTS

I would like to thank Professor R. Jackiw for discussions and a critical reading of the manuscript. I would also like to thank Professor J. L. Safko for his comments on the original version of this paper.

This work was supported in part by funds provided by the U.S. Department of Energy (D.O.E.) under Contract No. DE-AC02-76ER03069 and by Conselho Nacional de Desenvolvimento Científico e Tecnológico (CNPq), Brazil.

[1] Y. Aharonov and D. Bohm, Phys. Rev. **115**, 485 (1959).
[2] T. T. Wu and C. N. Yang, Phys. Rev. D **12**, 3845 (1975).
[3] E. J. Post, Rev. Mod. Phys. **39**, 475 (1967); A. Ashtekar and A. Magnon, J. Math. Phys. **16**, 341 (1875). A phase difference between plane waves interfering in a gravitational field was obtained by J. Anandan, Phys. Rev. D **15**, 1448 (1977).
[4] J. S. Dowker, Nuovo Cimento B **52**, 129 (1967); J. M. Cohen and D. R. Brill, *ibid.* **56**, 209 (1968); K. Kraus, Ann. Phys. (NY) **50**, 102 (1969); L. H. Ford and A. Vilenkin, J. Phys. A **14**, 2353 (1981); J. Stachel, Phys. Rev. D **26**, 1281 (1982); J. Audretsch and C. Lämmerzahl, J. Phys. A **16**, 2457 (1983); C. J. C. Burges, Phys. Rev. D **32**, 504 (1985); V. B. Bezerra, *ibid.* **35**, 2031 (1987); **38**, 506 (1988).
[5] J. K. Lawrence, D. Leiter, and G. Szamosi, Nuovo Cimento B **17**, 113 (1973); V. P. Frolov, V. D. Skarzhinsky, and R. W. John, *ibid.* **99**, 67 (1987); P. de Sousa Gerbert and R. Jackiw, Commun. Math. Phys. (in press).
[6] J. L. Safko and L. Witten, Phys. Rev. D **5**, 293 (1972); J. Math. Phys. **12**, 257 (1971).
[7] C. G. Oliveira and J. Tiomno, Nuovo Cimento **24**, 672 (1962).
[8] P. O. Mazur, Phys. Rev. Lett. **57**, 929 (1986).
[9] S. Deser, R. Jackiw, and G. 't Hooft, Ann. Phys. (NY) **152**, 220 (1984).
[10] *Handbook of Mathematical Functions*, edited by M. Abramowitz and I. A. Stegun (Dover, New York, 1970), p. 364.
[11] L. Marder, Proc. R. Soc. London Ser. A **252**, 45 (1959); A. Vilenkin, Phys. Rev. D **23**, 852 (1981); J. R. Gott, Astrophys. J **288**, 422 (1985); W. A. Hiscock, Phys. Rev. D **31**, 3288 (1985).

Scattering of a wave packet by an interval of random medium

William G. Faris and Woody J. Tsay
Program in Applied Mathematics, University of Arizona, Tucson, Arizona 85721

(Received 28 March 1989; accepted for publication 19 July 1989)

This paper deals with the scattering theory for the one-dimensional discrete Schrödinger equation with a random potential having large support. The main result is that a fluctuation deep within the scattering region has a very small effect on the scattering of wave packets; the region of random potential is effectively opaque.

I. INTRODUCTION

This paper deals with the scattering theory for the one-dimensional discrete Schrödinger equation with a random potential having support in an interval of length L. For each frequency k there are associated reflection and transmission coefficients $r_L(k)$ and $t_L(k)$. It is known[1–4] that when L is large, then $|t_L(k)|^2$ is close to zero with high probability. A wave packet is made from a superposition of waves of different frequencies. It is shown that the transmission of such a wave packet is also close to zero. Furthermore, it is shown that the phase of the reflected wave, which is determined by $r_L(k)$, is not affected by fluctuations deep within the medium. Thus for such problems the medium is effectively opaque.

It is important to note that the extension from fixed frequency to wave packets involves some subtlety. The reason is that for each fixed frequency only a very special solution occurring with low probability will penetrate the random medium. However, as the frequency varies these solutions will be encountered at certain special frequencies. Thus if the wave packet were to be concentrated near these frequencies, the wave could propagate a long way in the medium. However, in the case of a random medium, a wave packet that is chosen independent of the medium will have a low probability of being concentrated at these frequencies.

It is known from the work of Furstenberg[5] that the Lyapunov exponent measuring the growth of solutions of random initial value problems is strictly positive. By using a result from the theory of moment Lyapunov exponents[6] it is possible to obtain large deviation results on the growth of the solutions. These are used to obtain the bounds on the transmission of wave packets.

A more detailed analysis of the relation between two different solutions gives the results on insensitivity of the reflected wave to fluctuations in the interior. This depends on the remarkable fact[4] that two solutions with the same random function, but with different initial conditions becomes asymptotically proportional (in some random direction). The reason for this is that the area of the parallelogram spanned by the vectors is constant, so as the vectors grow in length they must also become parallel.

II. SCATTERING

This section contains a brief review of scattering theory for the one-dimensional discrete Schrödinger equation. Let $h>0$ be a lattice spacing constant. (This parameter will be fixed in all that follows.) Write $h\mathbf{Z}$ for the set of integer multiples of this constant and $h\mathbf{Z}_+$ for the set of integer multiples $\geqslant 0$. We shall take the Hilbert space to be the space of square-summable complex sequences indexed by one of these sets. Thus it is either $l^2(\mathbf{Z})$ (two-sided case) or $l^2(\mathbf{Z}_+)$ (one-sided case).

We consider a real function V defined on $\mathbf{Z}$. (We are interested in the situation when V is a sample function of a shift-invariant random process.) Let L be an integer multiple of h. A function V_L with bounded support is then defined, setting $V_L(x) = V(x)$ when $0<x\leqslant L$ and $V_L(x)=0$ elsewhere. Define the forward and backward difference operators $D_\pm$ by

$$D_\pm \phi(x) = [\phi(x\pm h) - \phi(x)]/\pm h. \tag{1}$$

Then the central second-difference operator is

$$D_-D_+\phi(x) = D_+D_-\phi(x) = [\phi(x+h) - 2\phi(x) + \phi(x-h)]/h^2. \tag{2}$$

The Schrödinger operator is defined by

$$H\phi(x) = -D_-D_+\phi(x) + V_L(x)\phi(x), \tag{3}$$

where the second term is multiplication by the function V_L. In the one-sided case the functions in the domain of H are restricted to vanish at the origin.

The Schrödinger wave equation is

$$i\frac{du(t)}{dt} = Hu(t). \tag{4}$$

In scattering theory we look for solutions that are far from the support of V_L at large times. The way to find such solutions is to fix k with $-\pi<kh<\pi$ and look for a solution for

$$H\psi_k = E(k)\psi_k \tag{5}$$

satisfying *scattering boundary conditions.* When $0<kh<\pi$ these solutions are

$$\psi_k(x) = e^{-ikx} + r_L(k)e^{ikx} \tag{6}$$

for $x\geqslant L$ and

$$\psi_k(x) = t_L(k)e^{-ikx} \tag{7}$$

for $x\leqslant 0$. [In the half-line case we replace this by $\psi_k(0)=0$, which in effect says that $t_L(k)=0$.] Such a solution can always be found by starting with the solution at 0, integrating forward, and then normalizing. It is easy to see that we must have $E(k) = (2-2\cos(kh))/h^2 = 4\sin^2(kh/2)/h^2$.

The $r_L(k)$ is the *reflection coefficient* and $t_L(k)$ is the *transmission coefficient*. These coefficients satisfy

$$|r_L(k)|^2 + |t_L(k)|^2 = 1, \tag{8}$$

which means, as we shall see, that the incoming wave is either reflected or transmitted.

Now choose a complex function $\hat{f}$ and a closed subinterval I of $(0,\pi/h)$.

Then

$$u(t,x) = \int_I e^{-iE(k)t}\psi_k(x)\hat{f}(k)\,\frac{dk}{2\pi}. \tag{9}$$

is the desired solution.

Standard scattering theory [for example, the method of stationary phase] shows that solution (9) is asymptotically of the form

$$u(t,x) \sim f(t,x)$$

$$= \int_I e^{-iE(k)t}e^{-ikx}\hat{f}(k)\,\frac{dk}{2\pi} \tag{10}$$

as $t\to -\infty$. Thus solution (9) represents a wave incoming from the right. Furthermore, solution (9) also satisfies

$$u(t,x) \sim f_R(t,x) + f_T(t,x) \tag{11}$$

as $t\to\infty$. Here

$$f_R(t,x) = \int_I e^{-iE(k)t}e^{ikx}r_L(k)\hat{f}(k)\frac{dk}{2\pi} \tag{12}$$

and

$$f_T(t,x) = \int_I e^{-iE(k)t}e^{-ikx}t_L(k)\hat{f}(k)\frac{dk}{2\pi}. \tag{13}$$

Equations (12) and (13) represent, respectively, a reflected wave going out on the right and a transmitted wave going out on the left. Furthermore,

$$\|f(t)\|^2 = \|f_R(t)\|^2 + \|f_T(t)\|^2, \tag{14}$$

where the norm is the l^2 norm defined in the usual way; for example,

$$\|f(t)\|^2 = \sum |f(t,x)|^2 h = \int_I |\hat{f}(k)|^2\,\frac{dk}{2\pi}. \tag{15}$$

Consider an arbitrary solution ϕ_k of the eigenvalue equation

$$H\phi_k = E(k)\phi_k. \tag{16}$$

We define the *current* as

$$J(x) = (1/2i)(\bar{\phi}_k(x)D_+\phi_k(x) - D_+\bar{\phi}_k(x)\phi_k(x)). \tag{17}$$

Since ϕ_k is a solution of the eigenvalue equation this current is constant. If we evaluate the current for the eigenfunction ψ_k satisfying the scattering boundary conditions (6) and (7) we see that

$$[\sin(kh)/h](1 - |r_L(k)|^2) = [\sin(kh)/h]|t_L(k)|^2, \tag{18}$$

that is, the reflection and transmission parts add up to 1 at each fixed frequency.

The eigenvalue equation with the function V may be written in matrix form as

$$\begin{pmatrix}\phi_k(x)\\ \dot{\phi}_k(x)\end{pmatrix} = \begin{pmatrix}1 & h\\ h(V(x)-E(k)) & 1+h^2(V(x)-E(k))\end{pmatrix} \times \begin{pmatrix}\phi_k(x-h)\\ \dot{\phi}_k(x-h)\end{pmatrix}, \tag{19}$$

where the second component $\dot{\phi}$ is the forward difference of ϕ. Note that all the matrices have determinant 1. We may rewrite (19) as

$$\mathbf{y}_k(x) = M(x)\mathbf{y}_k(x-h), \tag{20}$$

where the $M(x)$ have determinant 1. We use Eq. (20) for $x = h,2h,3h,\ldots,L$ to obtain the final $\mathbf{y}(L)$ from the initial $\mathbf{y}(0)$.

III. LYAPUNOV EXPONENTS

We assume that the real function V is a sample path of a stationary ergodic random process. The random variable $V(x)$ should have some finite moment. As before, the function V_L in the Schrödinger equation is equal to V in the interval $(0,L]$ and to zero elsewhere.

It is known that if $\mathbf{y}_k(0)$ is a vector that is independent of the random function V, then

$$\lim_{x\to\infty} (1/x)\,\mathbf{E}\log(\|\mathbf{y}_k(x)\|) = \lambda_k. \tag{21}$$

This limit is called the *Lyapunov exponent*. Furthermore, it is known[7–9] that if the V processs is nondeterministic, then the exponent $\lambda_k > 0$. On the other hand, it is not particularly easy to compute the exponent.[10]

From now on we make the assumption that the $V(x)$ are independent. It is known[4] that in this situation λ_k is continuous in k, so we may assume that it is bounded away from zero for k in the interval I.

There is a closely related quantity involving moments. For p real, define the *moment Lyapunov exponent* $g_k(p)$ by

$$\lim_{x\to\infty} (1/x)\log\mathbf{E}(\|\mathbf{y}_k(x)\|^p) = g_k(p). \tag{22}$$

This is a convex function of p and $g_k'(0) = \lambda_k$ (Refs. 11 and 12).

One consequence of $\lambda_k > 0$ and $g_k'(0) = \lambda_k$ is that there is an interval of $p<0$ such that $g_k(p)<0$. Take a p in this interval. We change notation to $q = -p$ and $0<\alpha< -g_k(p)$. Since our main interest is scattering and not Lyapunov exponents, we merely summarize the facts we need in the form of a hypothesis on the moment Lyapunov exponent.

Hypothesis 1: There exist $q>0$ and $\alpha>0$ such that for all k in I and for all initial conditions $\mathbf{y}_k(0)$ independent of the potential the inequality

$$\mathbf{E}((\|\mathbf{y}_k(0)\|/\|\mathbf{y}_k(x)\|)^q) \leqslant e^{-\alpha x} \tag{23}$$

is satisfied for sufficiently large x.

Hypothesis 1 has been verified in the case at hand of independent random variables.[6] It should be emphasized that Hypothesis 1 is about expectations for the solutions of the random differential equation for a fixed value of the parameter k. Our main purpose is to show that this has implications for the scattering problem, which involves the simultaneous consideration of solutions of all the equations with k ranging over an interval.

It follows from Hypothesis 1 and Chebyshev's inequality that

$$\Pr(\|\mathbf{y}_k(0)\|/\|\mathbf{y}_k(x)\|\geqslant e^{-\beta x})\leqslant e^{-(\alpha-q\beta)x}, \tag{24}$$

which is exponentially decreasing for sufficiently small $\beta>0$.

IV. TRANSMISSION

In this section we are interested in the whole line case in which a wave arrives from the right and is transmitted to the left. We shall see that if there is a large interval of random medium, then the transmission is small.

Take $\mathbf{y}_k(x)$ to be the solution corresponding to $\phi_k(x)$ with $\phi_k(x)=e^{-ikx}$ for $x\leqslant 0$. Then $\|\mathbf{y}_k(0)\|=\sqrt{1+E(k)}$. Also, the initial condition at the origin is independent of the potential and $\psi_k(x)=t_L(k)\phi_k(x)$. It is easy to bound $\|\mathbf{y}_k(L)\|$ in terms of a multiple of $1/|t_L(k)|$ and obtain

$$|t_L(k)|^2\leqslant 4[\|\mathbf{y}_k(0)\|^2/\|\mathbf{y}_k(L)\|^2]. \tag{25}$$

It is not necessarily true that the expectation of the rhs of inequality (25) is small for large L. However, if we note that $|t_L(k)|^2\leqslant 1$, this bound and the probability estimate above give

$$\mathbf{E}(|t_L(k)|^2)\leqslant e^{(-\alpha-q\beta)L}+4e^{-2\beta L}\leqslant 5e^{-\beta L} \tag{26}$$

if we take $\beta=\alpha/(2+q)$.

We see from Fubini's theorem that for normalized $\hat{f}$ with support in the interval I we have the following theorem.

Theorem 1: Consider scattering through an interval of random medium of length L. Under the moment hypothesis 1 the expected portion of the wave that is transmitted is exponentially small:

$$\mathbf{E}(\|f_T(t)\|^2)=\mathbf{E}\left(\int_I|t_L(k)|^2|\hat{f}(k)|^2\frac{dk}{2\pi}\right)\leqslant 5e^{-\beta L}. \tag{27}$$

From Chebyshev's inequality we conclude that the transmission of a wave packet is improbable. One formulation is the following corollary.

Corollary 1: Under the moment hypothesis 1 the probability that a significant portion of the wave is transmitted is small:

$$\Pr(\|f_T(t)\|^2\geqslant e^{-\gamma L}) =\Pr\left(\int_I|t_L(k)|^2|\hat{f}(k)|^2\frac{dk}{2\pi}\geqslant e^{-\gamma L}\right)\leqslant 5e^{-(\beta-\gamma)L}. \tag{28}$$

V. REFLECTION

In this section we consider either the two-sided case or the one-sided case and are interested in the reflection of a wave coming in from the right. The goal is to show that the reflection is not significantly affected by the potential deep within the interval of random medium.

We want to study the reflection from random media defined by two random potentials V_L and $\widetilde{V}_L$. We consider an additional parameter L' with $0<L'\leqslant L$ and require that the two potentials coincide for $L-L'<x\leqslant L$.

We may construct such potentials as follows. Let $\widetilde{V}$ be another process with the same distribution as V. Define $\widetilde{V}_L(x)=\widetilde{V}(x)$ for $0<x\leqslant L-L'$ and $\widetilde{V}_L(x)=V(x)$ for $L-L'<x\leqslant L$. Set $\widetilde{V}_L(x)=0$ elsewhere. Then $\widetilde{V}_L$ has the support $(0,L]$ and is independent of V_L on $(0,L-L']$ and equal to V_L on $(L-L',L]$.

We want to compare solutions of the two equations involving V_L and $\widetilde{V}_L$. We start with initial conditions at $x=0$ that are independent of the potential and obtain the solutions

$$\phi_k(x)=c_L(k)(e^{-ikx}+r_L(k)e^{ikx}) \tag{29}$$

and

$$\tilde{\phi}_k(x)=\tilde{c}_L(k)(e^{-ikx}+\tilde{r}_L(k)e^{ikx}) \tag{30}$$

for $x\geqslant L$.

Now we want to compare the two solutions. Define the relative current

$$\widetilde{J}(x)=(1/2i)(\bar{\phi}_k(x)D_+\tilde{\phi}_k(x)-D_+\bar{\phi}_k(x)\tilde{\phi}_k(x)). \tag{31}$$

Then

$$D_-J(x)=(1/2i)(\widetilde{V}_L(x)-V_L(x))\bar{\phi}_k(x)\tilde{\phi}_k(x). \tag{32}$$

It follows that $\widetilde{J}(L)=\widetilde{J}(L-L')$. We may also compute directly that

$$\widetilde{J}(L)=-[\sin(kh)/h]\bar{c}_L(k)\tilde{c}_L(k)(1-\bar{r}_L(k)\tilde{r}_L(k)). \tag{33}$$

Consider again the solutions $\mathbf{y}_k(x)$ and $\tilde{\mathbf{y}}_k(x)$ whose first components are $\phi_k(x)$ and $\tilde{\phi}_k(x)$. Then

$$2|\widetilde{J}(x)|\leqslant\|\mathbf{y}_k(x)\|\|\tilde{\mathbf{y}}_k(x)\|. \tag{34}$$

Therefore,

$$|1-\bar{r}_L(k)\tilde{r}_L(k)|^2=\frac{h^2|\widetilde{J}(L-L')|^2}{\sin^2(kh)|c_L(k)|^2|\tilde{c}_L(k)|^2} \leqslant\frac{h^2\|\mathbf{y}_k(L-L')\|^2\|\tilde{\mathbf{y}}_k(L-L')\|^2}{4\sin^2(kh)|c_L(k)|^2|\tilde{c}_L(k)|^2}. \tag{35}$$

Since we may bound the norm of the solution $\mathbf{y}_k$ at L in terms of the coefficient $c_L(k)$, we have

$$1/|c_L(k)|^2\leqslant 4(1+E(k))/\|\mathbf{y}_k(L)\|^2 \tag{36}$$

and we obtain

$$|1-\bar{r}_L(k)\tilde{r}_L(k)|^2\leqslant\frac{4h^2(1+E(k))^2}{\sin^2(kh)} \times\frac{\|\mathbf{y}_k(L-L')\|^2\|\tilde{\mathbf{y}}_k(L-L')\|^2}{\|\mathbf{y}_k(L)\|^2\|\tilde{\mathbf{y}}_k(L)\|^2}. \tag{37}$$

We may regard the values at $x=L-L'$ as initial conditions and apply the Lyapunov exponent argument on the interval $[L-L',L]$. As before, we obtain the bounds

$$\Pr(\|\mathbf{y}_k(L-L')\|/\|\mathbf{y}_k(L)\|\geqslant e^{-\beta L'})\leqslant e^{-(\alpha-q\beta)L'} \tag{38}$$

and similarly for $\tilde{\mathbf{y}}_k$. The probability of one or the other event is bounded by the sum of the probabilities. Thus we have

$$\Pr\left(|1-\bar{r}_L(k)\tilde{r}_L(k)|^2\geqslant\frac{4h^2(1+E(k))^2}{\sin^2(kh)}e^{-4\beta L'}\right) \leqslant 2e^{-(\alpha-q\beta)L'}. \tag{39}$$

Therefore,

$$\mathbf{E}(|1-\bar{r}_L(k)\tilde{r}_L(k)|^2)\leqslant 4e^{-(\alpha-q\beta)L'}+Ce^{-4\beta L'}, \tag{40}$$

where $C(k) = 4h^2(1 + E(k))/\sin^2(kh)$. Since $|r_L(k)| \leqslant 1$, we can take $\beta = \alpha/(4+q)$ and obtain the final bound

$$\mathbf{E}(|r_L(k) - |r_L(k)|^2 \tilde{r}_L(k)|) \leqslant (4 + C(k))e^{-\beta L'}. \quad (41)$$

In the one-sided case, $|r_L(k)| = 1$; this is the desired result. [In the two-sided case we can use the previous bound for $|t_L(k)|^2$ to bound the expectation of $|r_L(k) - \tilde{r}_L(k)|$.]

All these bounds are for probabilities conditioned on the behavior of the potential in the interal $(0, L - L']$. However, the bounds do not depend on the behavior, so we obtain the same bounds for the unconditioned probabilities.

For simplicity we state the theorem in the one-sided case. As before, we consider normalized $\hat{f}$ with support in the interval I (bounded away from zero).

Theorem 2: Consider a distribution of the $V(x)$ for which the moment hypothesis 1 is satisfied. Consider $0 < L' \leqslant L$ and two realizations V_L and $\tilde{V}_L$ of the random potential with support $(0,L]$ which coincide on $(L\text{-}L', L]$. Consider the two random reflection coefficients $r_L(k)$ and $\tilde{r}_L(k)$ determined by the two scattering problems. Then for large L' the scattering from the two potentials is almost the same in the sense that

$$\mathbf{E}(\|f_R(t) - \tilde{f}_R(t)\|^2) = \mathbf{E}\left(\int_I |r_L(k) - \tilde{r}_L(k)|^2 |\hat{f}(k)|^2 \frac{dk}{2\pi}\right) \leqslant Ce^{-\beta L'}. \quad (42)$$

The conclusion of this theorem may be summarized by saying that when $\beta L'$ is large the actual reflected scattering does not significantly depend on a layer deeper than L'.

One obtains a bound on probabilities from Chebyshev's inequality in the usual way.

ACKNOWLEDGMENTS

We thank Volker Wihstutz and Peter Baxendale for useful discussions at the Midwest Probability Colloquium.

The research of WGF was supported by National Science Foundation Grant No. DMS 810215.

[1]G. Papanicolaou and J. B. Keller, SIAM J. Appl. Math. **21**, 287 (1971).
[2]L. Pastur and E. P. Fel'dman, Sov. Phys. JETP **40**, 241 (1975).
[3]R. Johnston and H. Kunz, J. Phys. C: Solid State Phys. **16**, 3895 (1983).
[4]P. Bougerol and J. Lacroix, *Products of Random Matrices with Applications to Schrödinger Operators* (Birkhäuser, Boston, 1985).
[5]H. Furstenberg, Trans. Am. Math. Soc. **108**, 377 (1963).
[6]R. Carmona, A. Klein, and F. Martinelli, Commun. Math. Phys. **108**, 41 (1987).
[7]S. Kotani, in *Proceedings of the Taniguchi Symposium on Probabilistic Methods in Mathematical Physics*, Katata, edited by K. Ito and N. Ikeda (Kodansha, Tokyo, 1985), p. 219.
[8]B. Simon, Commun. Math. Phys. **89**, 227 (1983).
[9]F. Ledrappier, in *Lyapunov Exponents*, edited by L. Arnold and V. Wihstutz (Springer, Berlin, 1986), pp. 55–73.
[10]L. Arnold, G. Papanicolaou, and V. Wihstutz, SIAM J. Appl. Math. **46**, 427 (1986).
[11]L. Arnold, SIAM J. Appl. Math. **44**, 793 (1984).
[12]P. H. Baxendale, in *Proceedings of the Taniguchi Symposium on Probabilistic Methods in Mathematical Physics*, Katata, edited by K. Ito and N. Ikeda (Kodansha, Tokyo, 1985), p. 31.

Weyl-ordered fermions and path integrals

Giuliano M. Gavazzi
Istituto Nazionale di Fisica Nucleare, Sezione di Torino, Via P. Giuria 1, 10125 Torino, Italy

(Received 25 April 1989; accepted for publication 19 July 1989)

It is shown that a correspondence exists between the Weyl-ordered Hamiltonian and the midpoint prescription in the discrete path integral for fermions. It is then proven that the Feynman rules obtained from the discrete and continuous path integral are equivalent.

I. INTRODUCTION

The correspondence between classical and quantum functions of the dynamical variables is an old problem in quantum theory. On one hand (the canonical one) the question arises on how to order noncommuting operators. On the other hand (the functional one) it is the action in the path integral that presents ambiguities.[1,2]

A well-known result in the bosonic case[3] is that the Weyl-ordered form of operators is equivalent to the midpoint prescription in the path integral provided that one uses $\langle b|b\rangle = \int dx\, b^*b$ for the scalar product. In other words,

$$\langle x_2|\exp[-i\epsilon \mathbf{H}_{WO}]|x_1\rangle = \int dp \exp\left\{i\epsilon\left[p\,\frac{x_2-x_1}{\epsilon} - H\left(p, \frac{x_1+x_2}{2}\right)\right]\right\}, \quad (1)$$

where $\mathbf{H}_{WO}$ is derived from the classical Hamiltonian $H(p,x)$ considered as a symmetrized function of the canonical operators $\mathbf{x}$ and $\mathbf{p}$. [Equation (1), as in the others that follow, is valid up to the order ϵ included.] In our language the properly stated Weyl ordering acts on the products of the operators $\mathbf{Q}_i$ by symmetrizing them, i.e.,

$$(\mathbf{Q}_1\cdots\mathbf{Q}_K)_{WO} = \frac{1}{K!}\sum_{\text{perm}} \mathbf{Q}_{i_1}\cdots\mathbf{Q}_{i_K}, \quad (2)$$

so that we should specify the canonical coordinate operators to be symmetrized. (An operator Weyl ordered in one coordinate system may not be such after a canonical transformation.)

The rhs of Eq. (1) is the infinitesimal form of the phase space path integral written on the time lattice, where the Hamiltonian H is evaluated at the midpoint. The corresponding finite path integral is a product of infinitesimal ones integrated over the intermediate coordinates; as shown by Sato,[4] in the limit of zero lattice spacing this reproduces the same Feynman rules as the continuous path integral. Since in the last decade the path integral has been used extensively as a tool for quantization, it would be useful to include the other half of the world, the fermions, in this picture; we do this in what follows.

II. A REPRESENTATION FOR FERMIONS

Let us now consider the quantum mechanics of fermions. We have operators obeying anticommutation rules:

$$\{\Psi^\mu,\Psi^\nu\} = \{\overline{\Psi}_\mu,\overline{\Psi}_\nu\} = 0;$$
$$\{\Psi^\mu,\overline{\Psi}_\nu\} = \delta^\mu_\nu, \quad \mu,\nu = 1,\dots,N. \quad (3)$$

In order to establish a correspondence between the canonical and functional forms of the propagator, let us introduce the eigenvectors of Ψ[5]:

$$\langle\psi|\Psi^\mu = \psi^\mu\langle\psi|, \quad \Psi^\mu|\psi\rangle = (-1)^N|\psi\rangle\psi^\mu, \quad (4)$$

with the completeness

$$\int|\psi\rangle \overleftarrow{d^N\psi}\langle\psi| = 1, \text{ where } \overleftarrow{d^N\psi} = \overleftarrow{d\psi^1}\cdots\overleftarrow{d\psi^N}. \quad (5)$$

We use left arrows as in $\int f(z)\overleftarrow{dz} = f(z)\overleftarrow{\partial}/\partial z$, corresponding to the right derivative, and we use the transposed relation (right arrows) for the left derivative; when unnecessary the arrows will be suppressed. The relationship (5) holds only if each component Ψ^μ represents one fermionic degree of freedom: The linear space generated by 1 and ψ^μ has dimension 2 indeed. The ψ's are generators of a Grassmann algebra and

$$\int\psi\,\overleftarrow{d\psi} = -\int\overleftarrow{d\psi}\,\psi = 1, \quad \int\overleftarrow{d\psi} = 0 \quad (6)$$

as usual. The following formulas hold:

$$\langle\psi|\psi'\rangle \equiv \delta(\psi-\psi') = (\psi'-\psi)^N\cdots(\psi'-\psi)^1, \quad (7)$$

$$\langle\psi|\overline{\Psi}_\mu = \frac{\overrightarrow{\partial}}{\partial\psi^\mu}\langle\psi|, \quad \overline{\Psi}_\mu|\psi\rangle = (-1)^N|\psi\rangle\frac{\overleftarrow{\partial}}{\partial\psi^\mu}. \quad (8)$$

For the eigenvectors of $\overline{\Psi}$ we have

$$\overline{\Psi}_\mu|\overline{\psi}\rangle = |\overline{\psi}\rangle\overline{\psi}_\mu, \quad \langle\overline{\psi}|\overline{\Psi}_\mu = (-1)^N\overline{\psi}_\mu\langle\overline{\psi}|. \quad (4')$$

As is readily seen Eqs. (4′)[6] are the conjugate of the corresponding ones for Ψ. The same is true for the remaining equations; for instance,

$$\int|\overline{\psi}\rangle\overrightarrow{d^N\overline{\psi}}\langle\overline{\psi}| = 1, \quad \overrightarrow{d^N\overline{\psi}} = \overrightarrow{d\overline{\psi}_N}\cdots\overrightarrow{d\overline{\psi}_1}, \quad (5')$$

and

$$\int\overrightarrow{d\overline{\psi}}\,\overline{\psi} = -\int\overline{\psi}\,\overrightarrow{d\overline{\psi}} = 1. \quad (6')$$

The connection between these bases is given by

$$\langle\overline{\psi}|\psi\rangle = Ce^{\overline{\psi}_\mu\psi^\mu}, \quad \langle\psi|\overline{\psi}\rangle = C^{-1}e^{\psi^\mu\overline{\psi}_\mu} \quad (9)$$

(without loss of generality, $C = 1$).

 0022-2488/89/122904-03$02.50

III. WEYL ORDERING AND THE MIDPOINT RULE

The Weyl ordering is now defined ($\mathbf{Q}$ denotes Ψ or $\overline{\Psi}$) as

$$(\mathbf{Q}_1 \cdots \mathbf{Q}_K)_{WO} = \frac{1}{K!} \sum_{\text{perm}} (-1)^{\sigma(\text{perm})} \mathbf{Q}_{i_1} \cdots \mathbf{Q}_{i_K}. \quad (10)$$

The sign of the permutation is due to the statistics.

A useful property of the Weyl ordering is

$$(\mathbf{Q}_1\mathbf{Q}_2 \cdots \mathbf{Q}_K)_{WO} = \tfrac{1}{2}(\mathbf{Q}_1(\mathbf{Q}_2 \cdots \mathbf{Q}_K)_{WO} \pm (\mathbf{Q}_2 \cdots \mathbf{Q}_K)_{WO}\mathbf{Q}_1), \quad (11)$$

valid for fermions and bosons, with the minus sign when $\mathbf{Q}_1$ and $(\mathbf{Q}_2 \cdots \mathbf{Q}_K)_{WO}$ are fermions. [The proof relies on the fact that thanks to the symmetry of Weyl ordering, the nonvanishing (anti)commutators cancel.]

Now we can prove in general that to a Weyl-ordered fermionic Hamiltonian corresponds the midpoint prescription in the path integral, as sketched by several authors.[7] We proceed by induction.

Any Hamiltonian is a polynomial in the fermionic variables, so it is sufficient to prove the theorem for a product of Ψ's and $\overline{\Psi}$'s. For $\mathbf{H} = \overline{\Psi}_1 \cdots \overline{\Psi}_K$, which is Weyl ordered, we have

$$\begin{aligned}\langle \psi_2 | \exp(-i\epsilon\mathbf{H}) | \psi_1 \rangle &= \int \langle \psi_2 | (1 - i\epsilon\mathbf{H}) | \bar{\psi} \rangle d^N \bar{\psi} \langle \bar{\psi} | \psi_1 \rangle \\ &= \int \exp\left\{ i\epsilon \left[i\bar{\psi}_\mu \frac{\psi_2^\mu - \psi_1^\mu}{\epsilon} - H(\bar{\psi}) \right] \right\} d^N \bar{\psi}. \end{aligned} \quad (12)$$

It is easy to recognize the exponential of the action under the integral. Here the theorem holds: Since the Hamiltonian depends only on $\overline{\Psi}$ we have no ordering ambiguities and no midpoint at all.

Now let us suppose the correspondence is valid for a Weyl-ordered operator $(\mathbf{A})_{WO}$ corresponding to the classical function $A(\psi,\bar{\psi})$ and consider $B = \psi A$; we have

$$\mathbf{B}_{WO} \equiv (\Psi\mathbf{A})_{WO} = \tfrac{1}{2}(\Psi\mathbf{A}_{WO} \pm \mathbf{A}_{WO}\Psi), \quad (13)$$

having used Eq. (11). Finally,

$$\begin{aligned}\langle \psi_2 | \exp(-i\epsilon\mathbf{B}_{WO}) | \psi_1 \rangle &\equiv \left\langle \psi_2 \left| \exp\left[\frac{-i\epsilon(\Psi\mathbf{A}_{WO} \pm \mathbf{A}_{WO}\Psi)}{2} \right] \right| \psi_1 \right\rangle \\ &= \int \exp\left\{ i\epsilon \left[i\bar{\psi}_\mu \frac{\psi_2^\mu - \psi_1^\mu}{\epsilon} - \frac{(\psi_2 A(\bar{\psi},\psi_{MP}) \pm A(\bar{\psi},\psi_{MP})\psi_1)}{2} \right] \right\} d^N \bar{\psi} \\ &= \int \exp\left\{ i\epsilon \left[i\bar{\psi}_\mu \frac{\psi_2^\mu - \psi_1^\mu}{\epsilon} - B(\bar{\psi},\psi_{MP}) \right] \right\} d^N \bar{\psi}, \quad \text{with } \psi_{MP} = \frac{\psi_1 + \psi_2}{2}. \end{aligned} \quad (14)$$

Now any product of Ψ's and $\overline{\Psi}$'s can be generated by iteration of (13), starting from $\mathbf{A}$ equal to a product of $\overline{\Psi}$'s; thus for a general Hamiltonian H we have

$$\langle \psi_2 | \exp(-i\epsilon\mathbf{H}_{WO} | \psi_1 \rangle = \int d^N \bar{\psi} \exp\left\{ i\epsilon \left[i\bar{\psi}_\mu \frac{\psi_2^\mu - \psi_1^\mu}{\epsilon} - H(\bar{\psi},\psi_{MP}) \right] \right\}. \quad (15)$$

Equation (15) proves our first statement. In Sec. IV we show how the corresponding finite form of the discrete path integral gives rise to the naive Feynman rules which we can obtain from the more usual continuous path integral: Since the former is well defined we may gain more insight of the latter.

IV. EQUIVALENCE BETWEEN THE MIDPOINT AND CONTINUOUS PATH INTEGRALS

Let us recall a few points about the quantization of fermions via path integrals.

Given a Hamiltonian $H = H_0 + H_I$, where $H_0 = \omega\bar{\psi}\psi$, we have, for the action integral,

$$S = \int dt \{ i\bar{\psi}\dot{\psi} - \omega\bar{\psi}\psi - H_I(\bar{\psi},\psi) \}. \quad (16)$$

The fundamental object is the generator of the Green's functions (in the following we take $N = 1$)

$$\begin{aligned} Z(\bar{\chi},\chi) &= \int D\psi \, D\bar{\psi} \exp\left\{ i\left[S + \int dt(\bar{\chi}\psi + \bar{\psi}\chi) \right] \right\} \\ &= \exp\left[-i \int dt \, H_I\left(\frac{\overleftarrow{\delta}}{i\delta\chi}, \frac{\overrightarrow{\delta}}{i\delta\bar{\chi}} \right) \right] Z_0(\bar{\chi},\chi), \end{aligned} \quad (17)$$

where

$$Z_0(\bar{\chi},\chi) = \int D\psi \, D\bar{\psi} \exp\left[i \int dt (i\bar{\psi}\dot{\psi} - \omega\bar{\psi}\psi + \bar{\chi}\psi + \bar{\psi}\chi) \right]. \quad (18)$$

The interaction Hamiltonian defines the vertices, while the derivatives of Z_0 give the free propagators:

$$\langle \psi(t)\bar{\psi}(t') \rangle_0 = \frac{\overrightarrow{\delta}}{i\delta\bar{\chi}(t)} \log Z_0 \frac{\overleftarrow{\delta}}{i\delta\chi(t')} \bigg|_{\bar{\chi},\chi = 0}. \quad (19)$$

Eventually this machinery generates the usual Feynman rules of the perturbation expansion.

To compute the propagators we complete the square in Z_0 to obtain

$$Z_0(\bar{\chi},\chi) = \det\left[\frac{d}{dt} + i\omega \right] \exp\left[-\int dt \, \bar{\chi}\left(\frac{d}{dt} + i\omega \right)^{-1} \chi \right]. \quad (18')$$

Finally,

$$\begin{aligned}\langle \psi(t)\bar{\psi}(t') \rangle_0 &= \left(\frac{d}{dt} - i\omega \right) \Delta_F(t - t') \\ &= \theta(t - t')\exp[-i\omega(t - t')], \end{aligned} \quad (20)$$

where

$$\Delta_F(t) = \int_{-\infty}^{+\infty} \frac{d\omega'}{2\pi} \frac{e^{i\omega' t}}{\omega^2 - \omega'^2 - i\epsilon}$$

is the Feynman propagator.

Now we turn our attention to the discrete formalism. On the time lattice, where $t = t_0 + k\epsilon$, the generator of the exact Green's functions takes the form

$$Z^{\text{lat}}(\bar{\chi},\chi) = \lim_{\substack{T\to+\infty \\ t_0\to-\infty}} \int \langle 0|\psi(T)\rangle \prod_{k=1}^{T} d\psi(k) d\bar{\psi}(k)$$

$$\times \exp\left\{ i\epsilon \sum_{k=1}^{T} \left[i\bar{\psi}(k) \frac{\psi(k) - \psi(k-1)}{\epsilon} - H(\bar{\psi}(k),\psi(k)) + \bar{\chi}(k)\psi(k) + \bar{\psi}(k)\chi(k) \right] \right\} d\psi(0) \langle \psi(0)|0\rangle. \tag{21}$$

Now $\langle 0|\psi\rangle \propto 1$ and $\langle \psi|0\rangle \propto \psi$ (see Ref. 6), so that in the exponential we can consider $\psi(0) = 0$ because of the presence of the last factor; also, the integral over $\psi(0)$ is nothing but the normalization of the vacuum state: $\langle 0|0\rangle = 1$. Thus we have

$$Z^{\text{lat}}(\bar{\chi},\chi) = \lim_{\substack{T\to+\infty \\ t_0\to-\infty}} \exp\left[-i \sum_{k=1}^{T} H_I\left(\frac{\overleftarrow{\partial}}{i\epsilon\partial\chi}, \frac{\overrightarrow{\partial}}{i\epsilon\partial\bar{\chi}} \right)\epsilon \right] \times Z_0^{\text{lat}}(\bar{\chi},\chi) \tag{22}$$

and

$$Z_0^{\text{lat}}(\bar{\chi},\chi) = \lim_{\substack{T\to+\infty \\ t_0\to-\infty}} \det A \times \exp\left[-\epsilon^2 \sum \bar{\chi}(i) A^{-1}(i,j) \chi(j) \right], \tag{23}$$

with

$$A(i,j) = [(1 + i\epsilon\omega)\delta_{ij} - \delta_{i,j+1}]. \tag{24}$$

The inverse of this matrix is the free propagator

$$\langle \psi(j+m)\bar{\psi}(j)\rangle_0^{\text{lat}} = \frac{\overrightarrow{\partial}}{i\epsilon\partial\bar{\chi}(j+m)} \log Z_0^{\text{lat}} \frac{\overleftarrow{\partial}}{i\epsilon\partial\chi(j)}\bigg|_{\bar{\chi},\chi=0} = A^{-1}(j+m,j). \tag{25}$$

After some algebra we obtain

$$A^{-1}(j+m,j) = \begin{cases} 0, & m<0, \\ (1+i\epsilon\omega)^{-(m+1)}, & m\geqslant 0, \end{cases} \tag{25'}$$

and in the limit $\epsilon\to 0$ we have

$$\langle \psi(j+m)\bar{\psi}(j)\rangle_0^{\text{lat}} = \begin{cases} 0, & m<0, \\ \exp(-i\epsilon\omega m), & m\geqslant 0. \end{cases} \tag{26}$$

We see that the "discrete" propagator (26) is equivalent to the "continuous" one (20) when $m\epsilon = t - t' \neq 0$. For $t - t' = 0$ we have $\langle \psi(t)\bar{\psi}(t')\rangle_0 = (d/dt - i\omega)\Delta_F(0) = \frac{1}{2}$, although it is not clear to what it corresponds on the time lattice formalism. Let us clarify this point.

Until now we have ignored the ambiguity in the form of the action integral on the time lattice. This means that we can evaluate the Hamiltonian in Eq. (21) at any point between $\psi(k-1)$ and $\psi(k)$. Actually, no choice modifies the discrete free propagator (26); however, in order to have full equivalence between the discrete and continuous formalisms we must cure the $t - t' = 0$ disease. This is what the midpoint choice does; as a matter of fact, when we let $\psi(k) \to \psi_{MP}(k) = [\psi(k-1) + \psi(k)]/2$ in the interaction Hamiltonian H_I we have, for the equal-time propagator actually appearing in the perturbation expansion,

$$\langle \psi_{MP}(j)\bar{\psi}(j)\rangle_0^{\text{lat}} = \tfrac{1}{2}(0+1) = \tfrac{1}{2} = \langle \psi(t)\bar{\psi}(t)\rangle_0$$

and so we have succeeded in achieving the aim.

V. CONCLUSIONS

In this paper we have shown how the Weyl-ordered form of a fermionic Hamiltonian corresponds to the midpoint discrete path integral and in turn the latter corresponds, in the perturbation scheme, to the naive Feynman rules of the continuous path integral. This result may serve, for instance, as a useful tool in the study of the behavior of the quantum theory under coordinate transformation. In fact, to this purpose the operator or discrete path integral formalisms are the most reliable ways to work,[2] while the continuous path integral is the most direct approach. For instance, if an extra fermionic potential contribution is needed to make the Weyl Hamiltonian covariant, it must be added to the continuous fermionic Lagrangian as well in order to obtain the covariant Feynman rules.

In addition, we want to point out that the proof follows the same lines as in the bosonic case; this stresses the symmetry between bosons and fermions.

ACKNOWLEDGMENT

The author is grateful to Professor V. de Alfaro for valuable discussions.

[1] R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965); L. S. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).

[2] G. Gavazzi, Nuovo Cimento A **101**, 241 (1989).

[3] F. A. Berezin, Theor. Math. Phys. **6**, 194 (1971); M. Mizrahi, J. Math. Phys. **16**, 2201 (1975).

[4] M. Sato, Prog. Theor. Phys. **58**, 1262 (1977).

[5] T. D. Lee, *Particle Physics and Introduction to Field Theory* (Harwood, London, 1981).

[6] In this notation the bra $\langle\psi|$ is *not* the conjugate of $|\psi\rangle$: $|\psi\rangle^\dagger \neq \langle\psi| = |\bar{\psi}\rangle^\dagger$ up to a phase factor. For instance, we have $|\psi\rangle = -|0\rangle + |1\rangle\psi$ and $\langle\psi| = \psi\langle 0| + \langle 1|$, where $|0\rangle$ and $|1\rangle$ are the usual occupation number eigenvectors.

[7] V. de Alfaro, S. Fubini, G. Furlan, and M. Roncadelli, Nucl. Phys. B **269**, 402 (1988), see Appendix C; P. Salomonson and J. W. van Holten, Nucl. Phys. B **196**, 509 (1982).

Reduction of the Einstein equations in 2+1 dimensions to a Hamiltonian system over Teichmüller space

Vincent Moncrief
Department of Mathematics and Department of Physics, Yale University, P. O. Box 6666, New Haven, Connecticut 06511

(Received 14 June 1989; accepted for publication 2 August 1989)

In this paper the ADM (Arnowitt, Deser, and Misner) reduction of Einstein's equations for three-dimensional "space-times" defined on manifolds of the form $\Sigma\times\mathbf{R}$, where Σ is a compact orientable surface, is discussed. When the genus g of Σ is greater than unity it is shown how the Einstein constraint equations can be solved and certain coordinate conditions imposed so as to reduce the dynamics to that of a (time-dependent) Hamiltonian system defined on the $12g-12$-dimensional cotangent bundle, $T^*\mathscr{T}(\Sigma)$, of the Teichmüller space, $\mathscr{T}(\Sigma)$, of Σ. The Hamiltonian is only implicitly defined (in terms of the solution of an associated Lichnerowicz equation), but its existence, uniqueness, and smoothness are established by standard analytical methods. Similar results are obtained for the case of genus $g=1$, where, in fact, the Hamiltonian can be computed explicitly and Hamilton's equations integrated exactly (as was found previously by Martinec). The results are relevant to the problem of the reduction of the $3+1$-dimensional Einstein equations (formulated on circle bundles over $\Sigma\times\mathbf{R}$ and with a spacelike Killing field tangent to the fibers of the chosen bundle) and to the recent discussion by Witten of the possible exact solvability of the "topological dynamics" associated with Einstein's equations in $2+1$ dimensions.

I. INTRODUCTION

In a recent paper Witten has argued that the vacuum Einstein equations in $2+1$ dimensions are an exactly solvable system both classically and quantum mechanically.[1] Locally of course this seems quite reasonable, since Einstein's equations in three dimensions imply that space-time is flat. Globally, however, there are "topological degrees of freedom" to consider and Witten's conclusion is far from obvious.

One of the main steps in Witten's argument was to show that Einstein's equations in three dimensions can be viewed as the Euler–Lagrange equations determined by a Chern–Simons action for the "gauge group" ISO(2,1), i.e., for the Poincaré group in three dimensions. Since the dynamics of Yang–Mills connections (for compact, semisimple gauge groups), based on a Chern–Simons action, had previously been shown to be exactly solvable,[2] the corresponding result appeared to follow at once for Einstein's equations by analogy with the Yang–Mills case.

There are, however, several significant differences between the Yang–Mills theory and the Einstein one. For the Chern–Simons, Yang–Mills theory (on a manifold of the form $\Sigma\times\mathbf{R}$, where Σ is a compact two-manifold) the "space of classical solutions" coincides with the space of flat connections modulo gauge transformations defined on Σ. Work by Atiyah and Bott had shown the latter space to be a finite-dimensional compact, symplectic manifold.[3] Witten pointed out that, in the temporal gauge, the evolution equations are trivial and thus that the solution of the full set of field equations reduces to the solution of the initial value (or "constraint") equations—a problem solved previously by the work of Atiyah and Bott.

For the Einstein equations, however, the trivializing choice of "temporal gauge" is not permitted (at least classically) since, among other things, it corresponds to setting the time components of the orthonormal frame fields to zero. This in turn implies the vanishing of the lapse and shift fields of the space-time metric and thus the failure of this "metric" to have Lorentz signature. A vanishing choice for the lapse and shift fields does indeed trivialize the dynamics, but only in the relatively trivial way of preventing one from evolving the initial data at all.

The Einstein problem also differs from the Yang–Mills one in having a noncompact "gauge group" and a noncompact space of classical solutions. In particular, the analysis of Atiyah and Bott does not apply to this problem though, as Witten shows, the space of classical solutions can nevertheless be identified. It is essentially the cotangent bundle, $T^*\mathscr{T}(\Sigma)$, of the Teichmüller space, $\mathscr{T}(\Sigma)$, associated to the compact two-manifold Σ.

Insofar as one is only interested in "labeling" the classical solutions, rather than actually computing them, Witten's analysis would seem to be sufficient. But one could do essentially as much in $3+1$ dimensions! The space of classical solutions in that case (for, say, vacuum space-times with compact Cauchy surfaces) is of course infinite dimensional, and not everywhere a manifold,[4–7] but is nevertheless reasonably well understood (including the structure of its singularities and the relationship of the singular points to the isometry groups of the corresponding classical solutions). Lacking a means to characterize the solutions of the evolution equations, however, one would certainly not regard the four-dimensional Einstein equations as exactly solvable.

Locally, on the other hand, the Einstein evolution equations in three dimensions can indeed by solved explicitly by imposing, for example, Gaussian normal coordinate condi-

 0022-2488/89/122907-08$02.50

tions (lapse equal to unity, shift equal to zero). This choice eliminates all the spatial derivatives in the evolution equations and thus effectively reduces them to decoupled systems of ordinary differential equations along each normal geodesic. Certain particular solutions (including those obtained by taking quotients of Minkowski space by suitably chosen discrete subgroups of the Lorentz group) can in fact be globally foliated by Gaussian normal slicings. In general, however, one expects Gaussian normal coordinate systems to develop singularities unrelated to the natural boundaries of the space-times under study. In any case the solution of Einstein's equations through the use of Gaussian coordinates does not really realize Witten's objective to reduce these equations to a finite-dimensional Hamiltonian system defined on the cotangent bundle of Teichmüller space.

Our aim in this paper is to show that one can in fact reduce the Einstein equations to a Hamiltonian system on $T^*\mathscr{T}(\Sigma)$ but one that is, somewhat contrary to the spirit of Witten's discussion, both time-dependent and only implicity defined (except when the genus of Σ is less than 2). Our interest in this problem arose through its connection with the study of four-dimensional Einstein space-times defined on circle bundles over manifolds of the form $\Sigma\times\mathbf{R}$. If the four-dimensional metrics considered are required to have one spacelike Killing field (tangent to the fibers of the chosen circle bundle), then Einstein's equations can be projected (in the manner of Kaluza, Klein, and Jordan) to a set of field equations on the base manifold $\Sigma\times\mathbf{R}$. In the simplest cases the projected field equations take the form of Einstein's equations coupled to a set of harmonic map equations.[8–10] If the harmonic map field is taken to be a constant map then its stress–energy tensor vanishes and the projected field equations reduce to the pure vacuum Einstein equations on $\Sigma\times\mathbf{R}$. In any case one expects the Lorentzian metric on $\Sigma\times\mathbf{R}$ to have only "topological degrees of freedom" and thus that the original system should be reducible to an unconstrained system involving only harmonic map fields and "Teichmüller parameters." This has indeed been demonstrated for the special case $\Sigma\approx S^2$ in Ref. 8 (and, for the Einstein–Maxwell equations over the same base, in Ref. 9). Of course there are no Teichmüller parameters at all in this case and the associated "Witten problem" is vacuous. As a first step toward the study of the corresponding reduction problem for higher genus base manifolds we are thus led to consider the pure vacuum Einstein equations on $\Sigma\times\mathbf{R}$ when the genus of Σ is greater than zero—in other words, to consider the same problem posed by Witten.

In Sec. II A we recall the Arnowitt, Deser, and Misner (ADM) variational formulation of Einstein's equations in $2+1$ dimensions and apply standard techniques to solve the constraint equations for the cases in which the genus of Σ is greater than unity. In Sec. II B we discuss the reduced phase space and the construction (in principle) of the reduced Hamiltonian defined over it. Section II C describes the reconstruction of the space-time metric from a solution of Hamilton's equations and Sec. II D treats the special case for which the genus of Σ is one. In this case, the reduction can be carried out explicitly, as was discussed previously by Martinec.[11] Section III contains some concluding remarks.

II. REDUCTION OF EINSTEIN'S EQUATIONS

A. Variational formulation and solution of the constraint equations

Let $M=\mathbf{R}\times\Sigma$, where Σ is a compact orientable two-manifold of genus $g>1$. We may parametrize Lorentzian metrics on M, which have $t=$ const, spacelike hypersurfaces diffeomorphic to Σ, in the Arnowitt, Deser, and Misner (ADM) form[12,8,9]

$$\begin{aligned} ds^2 &= {}^3g_{\mu\nu}\,dx^\mu\,dx^\nu \\ &= -N^2\,dt^2 + g_{ab}(dx^a+N^a\,dt)(dx^b+N^b\,dt). \end{aligned} \tag{2.1}$$

Here μ,ν range over $\{0,1,2\}$ where $x^0=t$ is the "time" and a, b range over $\{1,2\}$. The space-time metric induces a Riemannian metric,

$$d\sigma^2 = g_{ab}\,dx^a\,dx^b \tag{2.2}$$

on each $t=$ const hypersurface.

The ADM action for Einstein's equations takes the form[13,8,9]

$$I=\int_{\mathscr{I}\times\Sigma} d^3x\{\pi^{ab}g_{ab,t}-N\mathscr{H}-N^a\mathscr{H}_a\}, \tag{2.3}$$

where $\mathscr{I}=[t_0,t_1]$ is some interval in the time coordinate t and in which the Hamiltonian and momentum constraints $\{\mathscr{H},\mathscr{H}_a\}$ are given by

$$\begin{aligned} \mathscr{H} &= (1/\sqrt{^2g})[\pi^{ab}\pi_{ab}-(\pi^a_a)^2]-\sqrt{^{(2)}g}\,{}^{(2)}R, \\ \mathscr{H}_a &= -2\,{}^{(2)}\nabla_b\pi^b_a. \end{aligned} \tag{2.4}$$

Here indices are raised and lowered with the two-metric g_{ab}, ${}^{(2)}g$ is the determinant of this metric, ${}^{(2)}R$ is its scalar curvature, and ${}^{(2)}\nabla_a$ represents covariant differentiation with respect to it. In the following it will be convenient to make a choice of units and then to treat the metric components and the coordinates as dimensionless quantities. Thus the physical line element is really our line element (2.1) multiplied by a fixed positive constant of dimension $(\text{length})^2$.

Since the genus of Σ is greater than unity by assumption, any smooth Riemannian metric g_{ab} on Σ is globally conformal, with a unique smooth conformal factor, to a metric h_{ab} with scalar curvature equal to -1. In other words, there exists a unique, smooth function λ: $\Sigma\to\mathbf{R}$ such that

$$g_{ab}=e^{2\lambda}h_{ab},\ {}^{(2)}R(h)=-1. \tag{2.5}$$

This is proven directly by solving the associated nonlinear elliptic equation for λ in Ref. 14 and by appeal to results from Riemann surface theory in Ref. 15. Since the dimension of Σ is 2, h_{ab} has constant sectional curvature as well.

The gravitational momentum π^{ab} is a symmetric tensor density closely related to the second fundamental form of the embedded hypersurface with Cauchy data (g_{ab},π^{ab}). For any such hypersurface, let λ be the unique function and h_{ab} be the unique constant curvature metric defined by Eq. (2.5). Also define

$$\tau = g_{ab}\pi^{ab}/\sqrt{^{(2)}g}, \tag{2.6}$$

the mean curvature of the hypersurface. Then, as in Refs. 8 and 9, one can uniquely decompose π^{ab} into three L^2-orthogonal summands,

$$\pi^{ab} = \pi^{ab^{\rm TT}} + \tfrac{1}{2}\tau\sqrt{^{(2)}g}\,g^{ab} + e^{-2\lambda}\sqrt{^{(2)}g}\,({}^{(2)}\nabla^b Y^a + {}^{(2)}\nabla^a Y^b - g^{ab\,(2)}\nabla_c Y^c), \tag{2.7}$$

where $\pi^{ab^{\rm TT}}$ is "transverse and traceless," i.e., satisfies

$$\begin{aligned} {}^{(2)}\nabla_b \pi^{ab^{\rm TT}} &= 0, \\ g_{ab}\pi^{ab^{\rm TT}} &= 0, \end{aligned} \tag{2.8}$$

and where Y^a is uniquely determined by solving the linear elliptic equation

$$ {}^{(2)}\nabla_a(\pi^{ab} - \tfrac{1}{2}\tau\sqrt{^{(2)}g}\,g^{ab}) = {}^{(2)}\nabla_a[e^{-2\lambda}\sqrt{^{(2)}g}\,({}^{(2)}\nabla^b Y^a + {}^{(2)}\nabla^a Y^b - g^{ab(2)}\nabla_c Y^c)]. \tag{2.9}$$

We now wish to impose the momentum constraint $\mathscr{H}_a = 0$ as a restriction upon the choice of π^{ab} and later to impose the Hamiltonian constraint to determine the allowed values of λ (hence g_{ab}). To decouple these two procedures, we shall need to impose the slicing condition that (Σ, g_{ab}, π^{ab}) be a hypersurface of constant mean curvature, i.e., that ${}^{(2)}\nabla_a\tau = 0$. Computing ${}^{(2)}\nabla_b\pi^{ab}$ under this assumption one obtains [cf. Eq. (3.9) of Ref. 8]

$$ {}^{(2)}\nabla_b(h)\{\sqrt{^{(2)}h}\,[{}^{(2)}\nabla^b(h)(h_{ac}Y^c) + {}^{(2)}\nabla_a(h)Y^b - \delta_a^{b(2)}\nabla_c(h)Y^c]\} = -\tfrac{1}{2}\tau_{,a}e^{2\lambda}\sqrt{^{(2)}h} = 0, \tag{2.10}$$

where ${}^{(2)}h$ is the determinant of h_{ab} and ${}^{(2)}\nabla_b(h)$ and ${}^{(2)}\nabla^b(h) = h^{ab(2)}\nabla_a(h)$ represent covariant differentiation with respect to this metric. Equation (2.10) has the unique solution $Y^a = 0$ (since a compact surface of constant negative curvature has no conformal Killing fields) and thus (2.7) reduces to

$$\pi^{ab} = \pi^{ab^{\rm TT}} + \tfrac{1}{2}\tau\sqrt{^{(2)}g}\,g^{ab}. \tag{2.11}$$

It is worth noting here that the mixed form of $\pi^{ab^{\rm TT}}$, defined as usual by $\pi_a^{b^{\rm TT}} = g_{ac}\pi^{cb^{\rm TT}}$, is divergence-free relative to any metric conformal to g_{ab}, in particular relative to h_{ab}. Thus

$$ {}^{(2)}\nabla_b(h)\pi_a^{b^{\rm TT}} = 0, \quad \pi_a^{a^{\rm TT}} = 0, \tag{2.12}$$

and therefore $p^{ab^{\rm TT}}$, defined by

$$p^{ab^{\rm TT}} = h^{ac}\pi_c^{b^{\rm TT}} = e^{2\lambda}g^{ac}\pi_c^{b^{\rm TT}} = e^{2\lambda}\pi^{ab^{\rm TT}}, \tag{2.13}$$

is symmetric (since it is proportional to $\pi^{ab^{\rm TT}}$) and transverse and traceless with respect to h_{ab}.

The Hamiltonian constraint, $\mathscr{H} = 0$, may now be written[8,9]

$$ {}^{(2)}\Delta_h\lambda = p_1 e^{2\lambda} - p_2 e^{-2\lambda} + p_3, \tag{2.14}$$

where

$$\begin{aligned} p_1 &= \tfrac{1}{4}\tau^2, \\ p_2 &= \tfrac{1}{2}[h_{ac}h_{bd}p^{ab^{\rm TT}}p^{cd^{\rm TT}}/({}^{(2)}h)], \\ p_3 &= \tfrac{1}{2}{}^{(2)}R(h) = -\tfrac{1}{2}, \end{aligned} \tag{2.15}$$

and where

$$ {}^{(2)}\Delta_h\lambda = (1/\sqrt{^{(2)}h})\,\partial_a(\sqrt{^{(2)}h}\,h^{ab}\lambda_{,b}), \tag{2.16}$$

the Laplacian of λ relative to the metric h_{ab}. This follows from combining Eqs. (2.4), (2.5), (2.11), and (2.13) with the conformal identity

$$ {}^{(2)}R(g) = {}^{(2)}R(e^{2\lambda}h) = e^{-2\lambda}[{}^{(2)}R(h) - 2\,{}^{(2)}\Delta_h\lambda]. \tag{2.17}$$

A complete existence and uniqueness theorem for Eq. (2.14) was proved in Ref. 8 for the case in which $\Sigma \approx S^2$ and ${}^{(3)}g_{\mu\nu}$ was coupled to certain "harmonic map" fields. The same analysis works for the higher genus two-manifolds and, for genus $g > 1$, leads to the following results: (i) no solutions of Eq. (2.14) exist unless $\tau^2 > 0$ [this follows upon integration of (2.14) over Σ and using the fact that $\tau = $ const by assumption]: (ii) if $\tau^2 > 0$ and $(h_{ab}, p^{ab^{\rm TT}})$ are smooth (i.e., C^∞), then a unique solution λ always exists and λ is smooth; and (iii) if $\tau^2 > 0$ and $(h_{ab}, p^{ab^{\rm TT}}) \in H_s \times H_{s-1}$ for $s > 2$ (where H_s is the Sobolev space of square integrable fields with square integrable derivatives up through order s) then a unique solution λ always exists and $\lambda \in H_{s+1}$.

To summarize the above results, one obtains the general solution of the constraint equations $\mathscr{H} = \mathscr{H}_a = 0$ on a spacelike hypersurface of constant mean curvature $\tau \neq 0$ by carrying out the following steps: (1) choosing a Riemannian metric h_{ab} on Σ having ${}^{(2)}R(h) = -1$; (2) choosing an arbitrary transverse-traceless symmetric tensor density $p^{ab^{\rm TT}}$ [i.e., one that satisfies $h_{ab}p^{ab^{\rm TT}} = 0$, ${}^{(2)}\nabla_a(h)p^{ab^{\rm TT}} = 0$, $p^{ab^{\rm TT}} = p^{ba^{\rm TT}}$]; (3) solving Eq. (2.14) for its unique solution λ; and (4) setting

$$\begin{aligned} g_{ab} &= e^{2\lambda}h_{ab}, \\ \pi^{ab} &= e^{-2\lambda}p^{ab^{\rm TT}} + \tfrac{1}{2}\tau\sqrt{^{(2)}h}\,h^{ab} = \pi^{ab^{\rm TT}} + \tfrac{1}{2}\tau\sqrt{^{(2)}g}\,g^{ab}. \end{aligned}$$

It is worth noting here that the conformal factor λ, determined by Eq. (2.14), depends smoothly upon the data occurring in that equation [i.e., upon $(h_{ab}, p^{ab^{\rm TT}}, \tau)$]. This follows as in Refs. 15 and 16 from the inverse function theorem and the observation that the linearization of Eq. (2.14) in λ yields an isomorphism between H_{s+1} and H_{s-1}.

B. The reduced phase space and ADM Hamiltonian

Let $\mathscr{M}^s$, for $s > 2$, denote the (Hilbert) manifold of H^s Riemannian metrics on Σ and let $\mathscr{M}^s_{-1}$ denote the set of metrics in $\mathscr{M}^s$ that have scalar curvature equal to -1. In Ref. 15, Fischer and Tromba show that $\mathscr{M}^s_{-1}$ is an (infinite-dimensional) closed C^∞ submanifold of $\mathscr{M}^s$. Let $\mathscr{M}$ denote the space of C^∞ Riemannian metrics on Σ and $\mathscr{M}_{-1}$ the C^∞ Riemannian metrics with scalar curvature equal to -1. One can view $\mathscr{M}$ and $\mathscr{M}_{-1}$ as (dense) subsets of $\mathscr{M}^s$ and $\mathscr{M}^s_{-1}$ (respectively) or, upon introducing a suitable differentiable structure such as the ILH (inverse limit of Hilbert) structure discussed by Fischer and Tromba, regard $\mathscr{M}$ and $\mathscr{M}_{-1}$ as differentiable manifolds in their own right. For simplicity, let us restrict our attention to the C^∞ case.

Now let $\mathscr{D}$ denote the group of C^∞ diffeomorphisms of Σ and let $\mathscr{D}_0 \subset \mathscr{D}$ denote the component of the identity. Since the equation ${}^{(2)}R(h) = -1$ is preserved by diffeomorphisms (under pullback of metrics), the groups $\mathscr{D}$ and $\mathscr{D}_0$ both act on $\mathscr{M}_{-1}$. Tromba and Fischer show that the quotient space, $\mathscr{M}_{-1}/\mathscr{D}_0$ is a smooth $6g-6$-dimensional manifold diffeomorphic to the Teichmüller space $\mathscr{T}(\Sigma)$ of Σ, which in turn is diffeomorphic to $\mathbf{R}^{6g-6}$.

Earlier work by Earle, Eells, and Sampson[17,18] had in fact already shown that $\pi: \mathscr{M}_{-1} \to \mathscr{M}_{-1}/\mathscr{D}_0$ (where π is the natural projection onto the quotient) is a trivial principal fiber bundle and that one can construct a global smooth cross section of this bundle through the use of harmonic maps. In physics terminology, the existence of a global cross section of the bundle $\pi: \mathscr{M}_{-1} \to \mathscr{M}_{-1}/\mathscr{D}_0$ corresponds to the existence of a globally valid choice of "gauge" for the diffeomorphism group $\mathscr{D}_0$.

The main result we shall need from this analysis is the existence of a smooth $6g-6$-dimensional global cross section of $\pi: \mathscr{M}_{-1} \to \mathscr{M}_{-1}/\mathscr{D}_0$, diffeomorphic to $\mathbf{R}^{6g-6}$ [and hence to the Teichmüller space $\mathscr{T}(\Sigma)$]. Letting $\{q^\alpha | \alpha = 1,2,...,6g-6\}$ be a global coordinate system on $\mathscr{T}(\Sigma) \approx \mathbf{R}^{6g-6}$, we can express any such cross section as a smooth $6g-6$ parameter family of C^∞ metrics on Σ with scalar curvature equal to -1:

$$\{q^\alpha\} \to h_{ab}(x^c, q^\alpha), \quad {}^{(2)}R(h(q^\alpha)) = -1. \tag{2.18}$$

We can now supplement the temporal coordinate condition of constant mean curvature slicing with the spatial coordinate condition that the conformal metric h_{ab} remain (during the course of time evolution) within the chosen global cross section of $\mathscr{M}_{-1}$. Since the initial value equations and the equations defining the conformal projection are covariant with respect to diffeomorphisms of Σ, there is clearly no loss of generality involved in assuming that h_{ab} lies initially on the chosen cross section. To ensure that as g_{ab} evolves the conformal metric h_{ab} remains within this cross section requires a particular choice of the shift vector field N^a, which is thereby implicitly determined. The lapse function N is also implicitly determined by the requirement that the evolution preserve the constant mean curvature slicing condition. We shall derive below the elliptic equations determining N and N^a, but for now shall simply assume that they have already been imposed.

Let $h_{ab}(x^c, \mathring{q}^\alpha)$ be an arbitrary point of the chosen global cross section of $\mathscr{M}_{-1}$ and let $p^{ab\,\mathrm{TT}}(x^c)$ be an arbitrary symmetric tensor density that is transverse and traceless with respect to $h_{ab}(x^c, \mathring{q}^\alpha)$. We define the "components" $\{p_\alpha | \alpha = 1,2...,6g-6\}$ of $p^{ab\,\mathrm{TT}}$ by

$$p_\alpha = \int_\Sigma \left(p^{ab\,\mathrm{TT}}(x^c) \frac{\partial h_{ab}}{\partial q^\alpha}(x^c, q^\beta) \right) \bigg|_{q^\alpha = \mathring{q}^\alpha} d^2x. \tag{2.19}$$

These components uniquely determine $p^{ab\,\mathrm{TT}}$ since the space of transverse traceless tensor densities at $h_{ab}(x^c, \mathring{q}^\alpha)$ is in fact $6g-6$-dimensional[15,17] and since the tangent vectors $(\partial h_{ab}/\partial q^\alpha)|_{q^\alpha = \mathring{q}^\alpha}$ span a $6g-6$-dimensional space transversal to the orbits of $\mathscr{D}_0$ through $h_{ab}(x^c, \mathring{q}^\alpha)$, whereas an arbitrary vector tangent to the $\mathscr{D}_0$ orbit [i.e., a tensor field of the form $(\mathscr{L}_{{}^{(2)}X} h)_{ab}$ for some vector field ${}^{(2)}X^a$ on Σ] annihilates $p^{ab\,\mathrm{TT}}$:

$$\int_\Sigma \{ p^{ab\,\mathrm{TT}} (\mathscr{L}_{{}^{(2)}X} h)_{ab} \} d^2x = -2 \int_\Sigma \{ {}^{(2)}\nabla_a(h) p^{ab\,\mathrm{TT}} h_{bc} {}^{(2)}X^c \} d^2x = 0. \tag{2.20}$$

This same calculation shows that the p_α's are independent of the particular cross section chosen to represent the Teichmüller space $\mathscr{T}(\Sigma)$ since, on the one hand, the integral expression in Eq. (2.19) is invariant with respect to diffeomorphisms and, on the other, any two representatives of the same tangent vector to $\mathscr{T}(\Sigma)$ [corresponding to different choices of cross section through $h_{ab}(x^c)$] would differ only by a field of the form $(\mathscr{L}_{{}^{(2)}X} h)_{ab}$, which annihilates $p^{ab\,\mathrm{TT}}$.

We may regard the p_α's as the components of an arbitrary covector at the point $\{\mathring{q}^\alpha\}$ of $\mathscr{T}(\Sigma)$. Thus the coordinates $\{(q^\alpha, p_\alpha) | \alpha = 1,...,6g-6\}$ give a global chart on $T^*\mathscr{T}(\Sigma) \approx \mathbf{R}^{12g-12}$, the cotangent bundle of the Teichmüller space of Σ. From the results of the preceding section it is clear that points of $T^*\mathscr{T}(\Sigma)$ uniquely label the $\mathscr{D}_0$ equivalence classes of solutions of the constraint equations on a hypersurface of constant mean curvature $\tau \neq 0$. Any choice of $\{q^\alpha, p_\alpha\} \in T^*\mathscr{T}(\Sigma)$ determines $h_{ab}(x^c, q^\alpha)$, then $p^{ab\,\mathrm{TT}}(x^c)$; Eq. (2.14) then yields λ, which, in turn, fixes g_{ab} and π^{ab}. It is worth noting that τ occurs both in the equation for λ and in the formula for π^{ab}, and thus must be specified together with the point of $T^*\mathscr{T}(\Sigma)$.

To determine the dynamics induced on $T^*\mathscr{T}(\Sigma)$ by Einstein's equations, we apply the ADM technique of reduction of the action. Letting $\{g_{ab}, \pi^{ab}\}(x^c, q^\alpha, p_\alpha, \tau)$ represent an arbitrary solution of the constraint equations parametrized by q^α, p_α, and τ, we substitute this solution into the expression (2.3). The constraint terms now vanish leaving

$$\begin{aligned}
I^* &= \int_{\mathscr{I}\times\Sigma} d^3x \{\pi^{ab} g_{ab,t}\} = \int_{\mathscr{I}\times\Sigma} d^3x \left\{ \left(e^{-2\lambda} p^{ab\,\mathrm{TT}} + \frac{1}{2}\tau\sqrt{{}^{(2)}h}\, h^{ab} \right) \cdot (e^{2\lambda} h_{ab})_{,t} \right\} \\
&= \int_{\mathscr{I}\times\Sigma} d^3x \left\{ p^{ab\,\mathrm{TT}} \frac{\partial h_{ab}}{\partial t} + \frac{\partial}{\partial t}[\tau e^{2\lambda}\sqrt{{}^{(2)}h}] - e^{2\lambda} h_{ab} \frac{\partial}{\partial t}\left(\frac{1}{2}\tau\sqrt{{}^{(2)}h}\, h^{ab} \right) \right\} \\
&= \int_{\mathscr{I}\times\Sigma} d^3x \left\{ p^{ab\,\mathrm{TT}} \frac{\partial h_{ab}}{\partial q^\alpha} \frac{dq^\alpha}{dt} - \frac{d\tau}{dt} e^{2\lambda}\sqrt{{}^{(2)}h} + \frac{\partial}{\partial t}[\tau e^{2\lambda}\sqrt{{}^{(2)}h}] - e^{2\lambda} h_{ab} \tau \frac{\partial}{\partial t}\left(\frac{1}{2}\sqrt{{}^{(2)}h}\, h^{ab} \right) \right\} \\
&= \int_{\mathscr{I}} dt \left\{ p_\alpha \frac{dq^\alpha}{dt} - \frac{d\tau}{dt} \int_\Sigma d^2x \sqrt{{}^{(2)}g} \right\} + \int_\Sigma d^2x [\tau\sqrt{{}^{(2)}g}]\,\Big|_{t_0}^{t_1},
\end{aligned} \tag{2.21}$$

where the term involving $(\partial/\partial t)(\frac{1}{2}\sqrt{{}^{(2)}h}\, h^{ab})$ has dropped out in the last step by virtue of the fact that $(\partial/\partial t)\, h_{ab}$ is tangent to $\mathscr{M}_{-1}$, and thus has an L^2-orthogonal decomposition of the form[15,19]

$$\frac{\partial h_{ab}}{\partial t} = k_{ab}^{\mathrm{TT}} + (\mathscr{L}_{{}^{(2)}X} h)_{ab}, \tag{2.22}$$

where $k_{ab}^{\rm TT}$ is transverse and traceless with respect to h_{ab}. This leads to an expression for $(\partial/\partial t)(\frac{1}{2}\sqrt{^{(2)}h}\, h^{ab})$, which is traceless with respect to h_{ab} and thus vanishes when contracted with this metric, as in Eq. (2.21).

To complete the determination of the reduced ADM action, we now fix the time coordinate uniquely by demanding that $\tau = t$ and drop the boundary terms from the right-hand side of Eq. (2.21), since they make no contribution to the equations of motion. The resulting ADM action may thus be written

$$I_{\rm ADM} = \int_{\mathscr{I}} dt \left\{ p_\alpha \frac{dq^\alpha}{dt} - H(q^\alpha, p_\alpha, t) \right\}, \tag{2.23}$$

where

$$\begin{aligned} H(q^\alpha, p_\alpha, t) &= \int_\Sigma d^2x \sqrt{^{(2)}g} \\ &= \int_\Sigma d^2x \{ e^{2\lambda} \sqrt{^{(2)}h} \} (q^\alpha, p_\alpha, t) \end{aligned} \tag{2.24}$$

is the area functional of the hypersurface $\tau = t$ expressed in terms of the canonical variables and the time through the solution of the Hamiltonian constraint equation for λ. Thus when the mean curvature is chosen to play the role of time the area functional becomes the corresponding Hamiltonian. Since this Hamiltonian depends explicitly upon the time it fails to be conserved by the evolution. The dynamics on $T^*\mathscr{T}(\Sigma)$ is that of a time-dependent Hamiltonian system—the time dependence corresponding to the nonconstancy of the area of the constant mean curvature hypersurfaces.

The expression for H is, in fact, independent of the particular choice of global cross section of $\mathscr{M}_{-1}$, which is used to define it and thus is intrinsically defined on $T^*\mathscr{T}(\Sigma)\times\mathbf{R}$. Unfortunately, however, the explicit form of H seems difficult to determine since it depends upon the solution of the nonlinear elliptic equation (2.14). The solution can be guessed whenever

$$p_2 = \tfrac{1}{2}[h_{ac} h_{bd} p^{ab\,\rm TT} p^{cd\,\rm TT} / (^{(2)}h)]$$

is constant on Σ, since λ is also constant in that case. For an arbitrary choice of data, however, this condition is not satisfied (except for the case of genus $g = 1$, which we have not considered yet) and therefore λ is not constant on Σ. To see that p_2 is not in general constant, one can appeal to the representation of transverse-traceless symmetric two-tensors in terms of holomorphic quadratic differentials, as discussed, for example, by Tromba and Fischer.[15,20] Thus the form of H and hence that of Hamilton's equations is only implicitly determined.

C. Determination of the lapse and shift fields

Using the definition (2.6) one can compute $\partial\tau/\partial t$ by means of Hamilton's equations for g_{ab} and π^{ab}. The result is

$$\begin{aligned} \frac{\partial\tau}{\partial t} &= -{}^{(2)}\nabla_a {}^{(2)}\nabla^a N + \frac{N}{(^{(2)}g)} \pi^b_a \pi^a_b \\ &= -{}^{(2)}\nabla_a {}^{(2)}\nabla^a N + \frac{N}{(^{(2)}h)} [e^{-4\lambda} h_{ab} h_{cd} p^{ac\,\rm TT} p^{bd\,\rm TT} \\ &\quad + \tfrac{1}{2}\tau^2 (^{(2)}h)], \end{aligned} \tag{2.25}$$

which can also be written as

$$\begin{aligned} e^{2\lambda}\frac{\partial\tau}{\partial t} &= -{}^{(2)}\Delta_h N + N\left\{ \frac{e^{-2\lambda}}{(^{(2)}h)} [h_{ab} h_{cd} p^{ac\,\rm TT} p^{bd\,\rm TT} \right. \\ &\quad \left. + \tfrac{1}{2} e^{4\lambda} (^{(2)}h)\tau^2] \right\} \equiv -{}^{(2)}\Delta_h N + Nq. \end{aligned} \tag{2.26}$$

Since, according to the results of the previous section, the function q is strictly positive, this linear elliptic equation always has a unique smooth solution for N. A straightforward application of the maximum principle shows that N will be strictly positive on Σ if and only if $\tau(t)$ satisfies $\partial\tau/\partial t > 0$. Thus our choice $\tau = t$ always yields a unique, smooth, strictly positive solution for N.

The determination of the shift field $N^a(\partial/\partial x^a)$ is complicated by the fact that it depends upon the particular choice of cross section of the bundle $\mathscr{M}_{-1} \to \mathscr{M}_{-1}/\mathscr{D}_0$, which one has made. For any such choice, however, the tangent vector to a solution curve in $\mathscr{M}_{-1}$ has a decomposition of the form (2.22), where the summand $(\mathscr{L}_{^{(2)}X} h)_{ab}$, which is tangent to the orbit of the diffeomorphism group $\mathscr{D}_0$ passing through h_{ab}, is uniquely determined, in terms of $k_{ab}^{\rm TT}$ by the choice of cross section. Thus $\partial h_{ab}/\partial t$ lies in the $6g-6$-dimensional tangent space (transversal to the $\mathscr{D}_0$ orbit), which is defined at h_{ab} by the chosen cross section.

To avoid having to compute the time derivative of the conformal factor λ it is convenient to evaluate the time derivative of $\sqrt{^{(2)}h}\, h^{ab} = \sqrt{^{(2)}g}\, g^{ab}$ instead of h_{ab} itself. Using the decomposition (2.22) on the one hand and the ADM evolution equation for $g_{ab,t}$ on the other, one finds that

$$\begin{aligned} &\sqrt{^{(2)}h}\,\{ h^{ac} h^{bd} k_{cd}^{\rm TT} + {}^{(2)}\nabla^a(h) X^b + {}^{(2)}\nabla^b X^a - h^{ab\,(2)}\nabla_c(h) X^c \} \\ &\quad = 2N(\pi^{ab} - \tfrac{1}{2} g^{ab}\pi^c_c) + \sqrt{^{(2)}h}\,\{ {}^{(2)}\nabla^a(h) N^b \\ &\qquad + {}^{(2)}\nabla^b(h) N^a - h^{ab\,(2)}\nabla_c(h) N^c \}, \end{aligned} \tag{2.27}$$

where the reexpression of the terms involving N^a in terms of h_{ab}, ${}^{(2)}\nabla_a(h)$ (instead of g_{ab}, ${}^{(2)}\nabla_a$) follows from the identity (3.7) given in Ref. 8.

As with $h_{ab,t}$ itself the tangent vector $(\sqrt{^{(2)}h}\, h^{ab})_{,t}$ to the curve induced in the space of contravariant densities, $\{\sqrt{^{(2)}h}\, h^{ab}\}$, lies in a $6g-6$-dimensional subspace transversal to the $\mathscr{D}_0$ orbit directions, which are, in turn, spanned by tensor densities of the form $\sqrt{^{(2)}h}\,({}^{(2)}\nabla^a(h) Y^b + {}^{(2)}\nabla^b(h) Y^a - h^{ab\,(2)}\nabla_c(h) Y^c)$. Let $\mathbf{P}_h$ designate the L^2-orthogonal projection onto the complement of this $6g-6$-dimensional tangent space (i.e., such that $I - \mathbf{P}_h$ is the projection onto the tangent space to the cross section at h_{ab}). Then the equation determining the shift field can be written as

$$\begin{aligned} &\mathbf{P}_h\{ \sqrt{^{(2)}h}\,[{}^{(2)}\nabla^a(h) N^b + {}^{(2)}\nabla^b(h) N^a - h^{ab\,(2)}\nabla_c(h) N^c] \} \\ &\quad = -\mathbf{P}_h\{ 2N(\pi^{ab} - \tfrac{1}{2} g^{ab}\pi^c_c) \}, \end{aligned} \tag{2.28}$$

which follows from applying $\mathbf{P}_h$ to Eq. (2.27). That a unique solution for N^a always exists follows, on the one hand, from the fact that the conformal Killing operator on vector fields has trivial kernel for $g > 1$ and, on the other, from the fact that $2N(\pi^{ab} - \frac{1}{2} g^{ab}\pi^c_c)$, being traceless, has a decomposition, relative to h_{ab}, of the same type as that for $(\sqrt{^{(2)}h}\, h^{ab})_{,t}$ occurring on the left-hand side of Eq. (2.27). The transverse-traceless part of $2N(\pi^{ab} - \frac{1}{2} g^{ab}\pi^c_c)$ determines $k_{ab}^{\rm TT}$, whereas the remainder has precisely the form of the conformal Killing operator applied to a vector field $Y^a(\partial/\partial x^a)$. Equation (2.28) thus reduces to

$$\sqrt{^{(2)}h}\{^{(2)}\nabla^a(h)(X^b - Y^b - N^b) + {}^{(2)}\nabla^b(h)$$
$$\times(X^a - Y^a - N^a) - h^{ab\,(2)}\nabla_c(h)$$
$$\times(X^c - Y^c - N^c)\} = 0, \tag{2.29}$$

which has the unique solution $N^a = X^a - Y^a$. Equation (2.28) for $N^a(\partial/\partial x^a)$ has the advantage of depending only upon the chosen cross section at h_{ab} (through $\mathbf{P}_h$) and not upon the particular solution curve in question.

We can summarize our results as follows. A solution curve of Hamilton's equations, $\{q^\alpha(t),p_\alpha(t)\}$, with $t=\tau$ determines, through the chosen cross section (2.18), a curve $h_{ab}(x^c,t) = h_{ab}(x^c,q^\alpha(t))$ in the space of metrics with scalar curvature -1 and, through the inverse of (2.19), a curve in the space of transverse-traceless tensor densities, $p^{ab^{\rm TT}}(x^c,t) = p^{ab^{\rm TT}}(x^c,q^\alpha(t),p_\alpha(t))$. The unique, smoothly varying solution of Eq. (2.14) for λ then yields

$$g_{ab}(x^c,t) = (e^{2\lambda}h_{ab})(x^c,t)$$
$$\pi^{ab}(x^c,t) = (e^{-2\lambda}p^{ab^{\rm TT}} + \tfrac{1}{2}\tau\sqrt{^{(2)}h}\,h^{ab})(x^c,t) \tag{2.30}$$
$$= (\pi^{ab^{\rm TT}} + \tfrac{1}{2}\tau\sqrt{^{(2)}g}g^{ab})(x^c,t).$$

Finally, the constructions of the present section yield $N(x^a,t),N^a(x^b,t)$ and hence the space-time metric (2.1) in the chosen coordinate system. Though the resulting form of the metric depends, of course, upon the chosen cross section of $\mathscr{M}_{-1}\to\mathscr{M}_{-1}/\mathscr{D}_0$, any global smooth cross section would suffice as well as any other. The temporal coordinate condition of setting $t=\tau$ has, however, played a more crucial role in allowing one to decouple the Hamiltonian and momentum constraints.

D. Reduction for the case of genus $g=1$

If the genus of Σ is zero (i.e., if $\Sigma\approx S^2$) there are no solutions to the constraint equations for a hypersurface of constant mean curvature. This follows from the fact that transverse-traceless symmetric two-tensors vanish identically on S^2 (since the Teichmüller space for S^2 has zero dimension) and thus that the function p_2, defined in Eq. (2.15) above, vanishes identically, whereas p_3 is now positive (in fact $p_3 = 1$ when h_{ab} is taken to be the canonical metric on the "unit" two-sphere). It follows at once that Eq. (2.14) has no solutions.

Therefore the only case remaining is that for which $g=1$ (i.e., $\Sigma\approx T^2$). Any Riemannian metric g_{ab} on T^2 is globally conformal to a flat metric $h_{ab} = e^{-2\lambda}g_{ab}$. One proves this by solving the linear elliptic equation,

$$(2/\sqrt{^{(2)}g})\partial_a(\sqrt{^{(2)}g}g^{ab}\lambda_{,b}) = -{}^{(2)}R(g), \tag{2.31}$$

for which the Gauss–Bonnet theorem provides the needed integrability condition.[21] Since λ is only determined up to an additive constant one can assume, without loss of generality, that a normalization condition such as

$$\int_{T^2}\sqrt{^{(2)}h}\,d^2x = (2\pi)^2 \tag{2.32}$$

is satisfied by h_{ab}.

Decomposing π^{ab} as in Eq. (2.7), one finds that Y^a is now only determined up to a conformal Killing field of g_{ab} (and hence of h_{ab}). Since h_{ab} is flat, it follows by a straightforward argument (upon taking the divergence of the conformal Killing equation) that Y^a is in fact covariantly constant (hence Killing) with respect to h_{ab}. Such fields form a two-dimensional space that, however (since they are conformally Killing with respect to g_{ab}), does not disturb the uniqueness of the decomposition (2.7).

Defining τ as in Eq. (2.6) and setting $\tau=$ const to decouple the constraint equations one finds, as before, that the momentum constraint implies that Y^a must be purely a conformal Killing field of g_{ab} and hence that π^{ab} reduces to the form given in Eq. (2.11). Defining $p^{ab^{\rm TT}}$, as in Eq. (2.13), one now finds that $p^{ab^{\rm TT}}$ is an arbitrary traceless, symmetric tensor density on T^2 that is covariantly constant relative to h_{ab}. This follows, for example, from analyzing the divergence condition on $p^{ab^{\rm TT}}$ in charts for which h_{ab} is itself constant [e.g., charts based upon the closed geodesics of (T^2,h_{ab})]. For fixed h_{ab} the space of $p^{ab^{\rm TT}}$'s is therefore two dimensional.

Since $p_3 = \frac{1}{2}\,{}^{(2)}R(h) = 0$, the Hamiltonian constraint now becomes

$${}^{(2)}\Delta_h\lambda = p_1e^{2\lambda} - p_2e^{-2\lambda}, \tag{2.33}$$

where (since $p^{ab^{\rm TT}}$ is covariantly constant) both p_1 and p_2 are constant on Σ.

If either p_1 or p_2 is zero, then a solution of Eq. (2.33) exists only if both p_1 and p_2 are zero and, in this case, λ is an arbitrary constant. Conversely, if both p_1 and p_2 are nonzero, then one can guess a particular solution given by

$$e^{4\lambda} = p_2/p_1 = \text{const} > 0. \tag{2.34}$$

By the argument given in the appendix of Ref. 8, this solution is unique. In either case, it follows that λ is constant and hence g_{ab} is flat and therefore that π^{ab} is covariantly constant with respect to g_{ab}.

The group $\mathscr{D}_0$ acts on the space of flat metrics on T^2 and, since the constraint equations are covariant with respect to $\mathscr{D}_0$, one can pass to a slice for this group action in which the flat metrics are represented by (spatially) constant metrics on T^2. In this "gauge" π^{ab} (since covariantly constant) is also constant on T^2. More precisely, if $\{x^1,x^2\}$ are periodic coordinates on T^2 then one can represent an arbitrary point (g_{ab},π^{ab}) within the slice by constant fields relative to the chosen coordinates (with, of course, g_{ab} and π^{ab} symmetric and with g_{ab} positive definite). The constraint equations reduce, within the chosen slice, to the single, purely algebraic equation,

$$\pi^{ab}\pi_{ab} - (\pi^a_a)^2 = 0, \tag{2.35}$$

which, for each fixed g_{ab}, defines a cone in the three-dimensional momentum space of π^{ab}'s. The solution set of the constraint equations within the slice thus reduces to a manifold except at the points of conical singularity defined by

$$\mathscr{S} = \{(g_{ab},\pi^{ab})\,|\,\pi^{ab}=0\}. \tag{2.36}$$

This is an example, in $2+1$ dimensions, of the well-known conical singularities that arise in the solution set of the constraint equations whenever the isometry group of the corresponding space-times changes discontinuously.[4,5] Here any vacuum space-time with Cauchy data lying in $\mathscr{S}$ has a globally defined timelike Killing field whereas solutions determined by data in the complement of $\mathscr{S}$ are nonstationary. Every solution has $\partial/\partial x^1$ and $\partial/\partial x^2$ as spacelike Killing fields but, since there is no possibility of breaking these symmetries in $2+1$ dimensions, there are no additional conical singularities associated with their occurrence.

To sidestep the complications arising from the singular points of the constraint set, let us simply cut out the points belonging to the singular set $\mathscr{S}$ and work on the complementary manifold. Thus we demand that the mean curvature τ be nonzero and set

$$g_{ab} = e^{2\lambda} h_{ab}, \quad \pi^{ab} = e^{-2\lambda} p^{ab^{\mathrm{TT}}} + \tfrac{1}{2}\tau\sqrt{{}^{(2)}h}\, h^{ab}, \tag{2.37}$$

where

$$e^{2\lambda} = \sqrt{(2/\tau^2)\left[h_{ac}h_{bd}p^{ab^{\mathrm{TT}}}p^{cd^{\mathrm{TT}}}/({}^{(2)}h)\right]} \tag{2.38}$$

and where h_{ab} and $p^{ab^{\mathrm{TT}}}$ are arbitrary constant symmetric tensor fields (with h_{ab} positive definite) satisfying the algebraic conditions

$$h_{ab}p^{ab^{\mathrm{TT}}} = 0, \quad \det({}^{(2)}h) = 1. \tag{2.39}$$

The last of these is equivalent to the normalization condition (2.32) introduced above, provided x^1 and x^2 are periodic coordinates (defined modulo 2π) on T^2.

For definiteness, let us take the case for which $\tau > 0$ and impose the temporal coordinate condition

$$\tau = \exp[t/(2\pi)^2]. \tag{2.40}$$

Upon choosing an appropriate set of coordinates $\{q^\alpha\} = \{q^1, q^2\}$ for the two-dimensional "Teichmüller space" of constant metrics with fixed volume elements and defining their conjugate momenta $\{p_\alpha\} = \{p_1, p_2\}$ via Eq. (2.19), one obtains, as in Eq. (2.21), the reduced action

$$I^* = \int_{\mathscr{S}} dt \left\{ p_\alpha \frac{dq^\alpha}{dt} - \left(\frac{2 p_a^{b^{\mathrm{TT}}} p_b^{a^{\mathrm{TT}}}}{({}^{(2)}h)} \right)^{1/2} \right\} + \text{boundary term}, \tag{2.41}$$

in which h_{ab} and $p_a^{b^{\mathrm{TT}}} = h_{ac}p^{cb^{\mathrm{TT}}}$ are understood to be expressed in terms of the cononical variables q^α and p_α. The choice (2.40) for the time coordinate condition has, in this case, ensured a time-independent Hamiltonian function. A convenient choice of canonical variables has been given by Martinec, who proceeded to solve the evolution equations explicitly.[11]

With our choice of temporal coordinate condition the unique solution of Eq. (2.25) turns out to be

$$N = [1/(2\pi)^2\tau]. \tag{2.42}$$

The shift vector field $N^a(\partial/\partial x^a)$ is only determined up to an arbitrary time-dependent linear combination of the two Killing fields $\partial/\partial x^1$ and $\partial/\partial x^2$. The simplest choice is simply to take $N^a = 0$.

III. CONCLUDING REMARKS

The methods of this paper could be used to study several related problems such as the reduction of Einstein's equations in the presence of a cosmological constant or suitable sources. One should be prepared for the possibility that the existence or uniqueness of the solution of the Hamiltonian constraint equation may fail, depending upon the sign of the chosen cosmological constant or the nature of the chosen sources.

For the (finite-dimensional) Hamiltonian systems studied here one could also consider the problem of quantization. However, as the explicit example of the genus $g = 1$ case shows, one should not expect the Hamiltonian to be a polynomial in the momentum variables $\{p_\alpha\}$ and therefore one should not expect the associated "Schrödinger equation" (formulated with a suitable choice of operator ordering in the coordinate representation) to be a conventional partial differential equation. The $g = 1$ case is tractable but already involves the square root of a certain Laplacian defined on the associated "Teichmüller space." While one can reasonably identify the appropriate Hilbert space as square integrable **C**-valued functions over Teichmüller space (as Witten has done), the possibility of actually solving the associated Schrödinger equation in the higher genus cases seems currently rather remote.

ACKNOWLEDGMENT

This research was supported in part by National Science Foundation Grant No. PHY-8503072 to Yale University.

[1] E. Witten, Nucl. Phys. B **311**, 46 (1988).
[2] E. Witten, Commun. Math. Phys. **121**, 351 (1989).
[3] M. F. Atiyah and R. Bott, Philos. Trans. R. Soc. London Ser. A **308**, 523 (1982).
[4] J. Arms, J. Marsden, and V. Moncrief, Ann. Physics (NY) **144**, 81 (1982); see also Commun. Math. Phys. **78**, 455 (1981) by the same authors.
[5] A. Fischer, J. Marsden, and V. Moncrief, Ann. Inst. H. Poincaré **33**, 147 (1980).
[6] J. Isenberg and J. Marsden, J. Geom. Phys. **1**, 85 (1984); see also Phys. Rep. **89**, 179 (1982) by the same authors.
[7] J. Isenberg, Phys. Rev. Lett. **59**, 2389 (1987).
[8] V. Moncrief, Ann. Phys. (NY) **167**, 118 (1986); see also the proceedings of the Colloquium "Geometrie et Physique" in honor of Andre Lichnerowicz, Paris, 1986 (Hermann, 1988) for a review by the same author.
[9] V. Moncrief, "Reduction of the Einstein–Maxwell and Einstein–Maxwell–Higgs Equations for Cosmological Spacetimes with Spacelike U(1) Isometry Groups," Yale preprint, 1989.
[10] The reduction problem for Einstein's equations on circle bundles on $\Sigma \times \mathbf{R}$, for the higher genus cases, is currently under study by J. Cameron and V. Moncrief.
[11] E. Martinec, Phys. Rev. D **30**, 1198 (1984).
[12] See, for example, Chap. 21 of *Gravitation* by C. Misner, K. Thorne, and J. Wheeler (Freeman, San Francisco, 1973).
[13] The ADM variational principle for the vacuum Einstein equations in $2+1$ dimensions results from setting the harmonic map fields in Refs. 8 and 9 to constant maps. The corresponding expression (for an arbitrary number of space-time dimensions) was given by E. Martinec in Ref. 11.
[14] T. Aubin, *Nonlinear Analysis on Manifolds. Monge–Ampère Equations* (Springer, New York, 1982); see, in particular, the discussion of "Berger's problem" on pp. 119–122.
[15] A. Fischer and A. Tromba, Math. Ann. **267**, 311 (1984).
[16] The argument needed here is essentially the same as that given in the proof of Theorem 7.2 of Ref. 15.
[17] C. Earle and J. Eells, J. Diff. Geom. **3**, 19 (1969).
[18] The continuity and later the smoothness of the global cross section defined by Earle and Eells in Ref. 17 were established (respectively) by Sampson and Eells. For a brief discussion of this point see the proof of Theorem 9.4 in Ref. 15.

[19]The decomposition defined on the right-hand side of Eq. (2.22) is equivalent to that derived in the proof of Theorem 8.2 of Ref. 15.

[20]Using the (purely local) formulas satisfied by "quadratic differentials" (see, for example, the discussion in the proof of Theorem 8.9 of Ref. 15) it is not hard to show that there is, at most, a two-parameter family of transverse-traceless symmetric two-tensors for which the function p_2 is constant. Since the dimension of the space of such two-tensors is $6g-6$, for genus $g>1$, p_2 cannot in general be constant. For the special case of $g=1$, as discussed here in Sec. II D, p_2 is always constant and the solution of Lichnerowicz's equation can be given explicitly.

[21]This problem is discussed and solved by M. Berger in *Nonlinearity and Functional Analysis* (Academic, New York, 1977); see also Ref. 14.

Static "semi-plane-symmetric" metrics yielded by plane-symmetric electromagnetic fields

Jian-zeng Li
Center of Theoretical Physics, CCAST (World Lap), Beijing, People's Republic of China, and Department of Physics, Beijing Broadcasting Institute, Beijing, People's Republic of China

Can-bin Liang
Center of Theoretical Physics, CCAST (World Lap), Beijing, People's Republic of China, and Department of Physics, Beijing Normal University, Beijing, People's Republic of China

(Received 4 January 1989; accepted for publication 26 July 1989)

The task of seeking a general static solution to the Einstein–Maxwell equations representing "semi-plane-symmetric" metrics yielded by plane-symmetric electromagnetic fields is reduced to solving a single ordinary differential equation. A special solution is given, showing that there does exist some electrovac metric that does not share some of the symmetries of the electromagnetic fields.

I. INTRODUCTION

Einstein–Maxwell fields in which the metrics g_{ab} and the corresponding electromagnetic fields F_{ab} admit the same symmetries are well known. There also exist Einstein–Maxwell fields in which F_{ab} does not share some of the symmetries of g_{ab},[1] for instance, the general solution to Einstein–Maxwell equations representing plane-symmetric metrics yielded by the "semi-plane-symmetric" electromagnetic fields given by Li and Liang[2] and Kuang *et al.*[3] It is interesting to ask whether the opposite case is possible, i.e., whether there exists an electrovac metric that does not share some of the symmetries of the electromagnetic field. The main purpose of this paper is to give an affirmative answer to this question by giving a static solution to Einstein–Maxwell equations in which the electromagnetic field F_{ab} is plane-symmetric while the metric g_{ab} is only "semi-plane-symmetric," in the sense that there exists a coordinate system (t,x,y,z) such that $\mathscr{L}_{\xi_{(i)}} F_{ab} = 0$ ($i = 1,2,3$ and $\mathscr{L}$ denotes the Lie derivative) with

$$\xi_{(1)} = \frac{\partial}{\partial x}, \quad \xi_{(2)} = \frac{\partial}{\partial y}, \quad \xi_{(3)} = x\left(\frac{\partial}{\partial y}\right) - y\left(\frac{\partial}{\partial x}\right),$$

and where g_{ab} can be written as

$$dS^2 = E(t,z)(-dt^2 + dz^2) + G(t,z)dx^2 + H(t,z)dy^2, \quad (1)$$

where $E(t,z)$, $G(t,z)$, and $H(t,z)$ are arbitrary positive functions. Note that $\xi_{(1)}$ and $\xi_{(2)}$ are Killing vectors of metric (1) while $\xi_{(3)}$ is not if $G(t,z) \neq H(t,z)$, although it is a symmetry vector field of F_{ab}. This is why we refer to such a metric as "semi-plane-symmetrical." Incidentally, metric (1) is locally a special case of the cylindrically symmetric metrics [formula (20.1) of Ref. 1 with $A = 0$].

II. THE EQUATIONS

Denote $x^0 = t$, $x^1 = x$, $x^2 = y$, and $x^3 = z$. It follows from Eq. (1) that $R_{01} = R_{02} = R_{12} = R_{13} = R_{23} = 0$; thus the Einstein equations require

$$T_{01} = T_{02} = T_{12} = T_{13} = T_{23} = 0. \quad (2)$$

Define the electric and magnetic fields **E** and **B** with respect to this coordinate system to be $E_1 = F_{01}$, $E_2 = F_{02}$, $E_3 = F_{03}$, $B_1 = F_{32}$, $B_2 = F_{13}$, $B_3 = F_{21}$. It then follows from Eq. (2) that

$$\begin{aligned} EE_2B_3 &= HE_3B_2, \\ EE_1B_3 &= GE_3B_1, \\ E_1E_2 &= -B_1B_2, \\ HE_1E_3 &= -EB_1B_3, \\ GE_2E_3 &= -EB_2B_3. \end{aligned} \quad (3)$$

Equation (3) is a necessary condition imposed on F_{ab} for yielding a "semi-plane-symmetric" metric. It is not difficult to show that there can be only two cases of F_{ab} that satisfy requirement (3).

(i) $E_1 = E_2 = B_1 = B_2 = 0$. Calculation of $^*F_{ab}\ ^*F^{ab}$ (where $^*F_{ab} \equiv F_{ab} + i\widetilde{F}_{ab}$, $\widetilde{F}_{ab}$ being the dual form of F_{ab}) shows that in this case F_{ab} must be a non-null electromagnetic field. Restricting F_{ab} to be source-free, one obtains from the source-free Maxwell equations that

$$B_3^2 = a_1^2, \quad E_3^2 = a_2^2 E^2/GH, \quad a_1, a_2 \text{ constants.} \quad (4)$$

This implies that E_3 and B_3 are independent of x and y, and hence F_{ab} is plane-symmetric.

(ii) $E_3 = B_3 = 0$. It follows from the source-free Maxwell equations that E_1, E_2, B_1, and B_2 are also independent of x and y, but, since $\mathscr{L}_{\xi_{(3)}} F_{ab} \neq 0$ [$\xi_{(3)} \equiv x(\partial/\partial y) - y(\partial/\partial x)$], F_{ab} is only "semi-plane-symmetric." Calculation of $^*F_{ab}\ ^*F^{ab}$ shows that F_{ab} can be either null or non-null, unlike the case of plane-symmetric metrics yielded by "semi-plane-symmetric" electromagnetic fields F_{ab}, where F_{ab} must be null.[2,3]

The main focus of this paper involves only case (i). Expressing the nonvanishing components of the Ricci tensor in terms of E, G, and H, one obtains from the Einstein equations the following equations:

$$\frac{\dot{E}\dot{G} + E'G'}{2EG} + \frac{\dot{E}\dot{H} + E'H'}{2EH} + \frac{\dot{G}^2 + G'^2}{4G^2} - \frac{\ddot{G} + G''}{2G} + \frac{\dot{H}^2 + H'^2}{4H^2} - \frac{\ddot{H} + H''}{2H} = 0, \quad (5)$$

 0022-2488/89/122915-03$02.50

$$\frac{\dot G^2-G'^2}{4G^2}-\frac{\ddot G-G''}{2G}+\frac{\dot H^2-H'^2}{4H^2}-\frac{\ddot H-H''}{2H}+\frac{EE''-E'^2-E\ddot E+\dot E^2}{E^2}=8\pi a^2\frac{E}{GH}, \tag{6}$$

$$\frac{\ddot G-G''}{2G}-\frac{\ddot H-H''}{2H}-\frac{\dot G^2-G'^2}{4G^2}+\frac{\dot H^2-H'^2}{4H^2}=0, \tag{7}$$

$$\frac{\ddot G-G''}{2G}+\frac{\ddot H-H''}{2H}-\frac{\dot G^2-G'^2}{4G^2}-\frac{\dot H^2-H'^2}{4H^2}+\frac{\dot G\dot H-G'H'}{2GH}=8\pi a^2\frac{E}{GH}, \tag{8}$$

$$\frac{\dot EG'+E'\dot G}{4EG}+\frac{E'\dot H+\dot EH'}{4EH}-\frac{\dot G'}{2G}-\frac{\dot H'}{2H}+\frac{\dot GG'}{4G^2}+\frac{\dot HH'}{4H^2}=0, \tag{9}$$

where $a^2=a_1^2+a_2^2$ is a scalar characterizing the electromagnetic field, and an overdot and a prime represent "$\partial/\partial t$" and "$\partial/\partial z$," respectively. We have not yet found any solutions to Eqs. (5)–(9) when E, G, and H are functions of both z and t. However, if E, G, and H are restricted to be functions of z only (and hence the metric is static), these equations can be reduced to a single ordinary differential equation. This is shown in the next section and a special solution is given in Sec. IV.

III. THE REDUCED ORDINARY DIFFERENTIAL EQUATION

Since in the static case Eq. (9) is satisfied automatically, we have only four equations [(5)–(8)] left. Introduce two functions $M(z)$ and $N(z)$ such that

$$H(z)=M(z)N(z),\quad G(z)=M(z)N^{-1}(z), \tag{10}$$

Eqs. (7) and (8) then yield, respectively,

$$(N'/N)'+M'N'/MN=0, \tag{11}$$

and

$$8\pi a^2E=-MM''. \tag{12}$$

The general solution to Eq. (11) is

$$N=C_2\exp\left[C_1\int dz[\sqrt{3}M(z)]^{-1}\right],\quad C_1,C_2 \text{ constants.} \tag{13}$$

Since $a\neq 0$, otherwise there would be no electromagnetic field, it follows from Eq. (12) that

$$E=-MM''/8\pi a^2,$$

and hence Eqs. (5) and (6) become, respectively,

$$\frac{M'M'''}{MM''}+\frac{3M'^2}{2M^2}-\frac{M''}{M}-\frac{N'^2}{2N^2}=0, \tag{14}$$

$$\frac{3M''}{M}+\frac{N'^2}{2N^2}-\frac{3M'^2}{2M^2}+\left(\frac{M'''}{M''}\right)'=0. \tag{15}$$

Substitution of (13) into Eq. (14) gives

$$\left(\frac{M'''}{M''}\right)'=-\frac{C_1^2}{6M^2}-\frac{C_1^2M''}{6MM'^2}+\left(\frac{M''}{M'}\right)'-\frac{3M''}{2M}+\frac{3M'^2}{2M^2},$$

which, together with Eq. (15), yields

$$\left(\frac{M''}{M'}\right)'-\frac{C_1^2M''}{6MM'^2}+\frac{3M''}{2M}=0. \tag{16}$$

On the other hand, Eq. (14) can be rewritten as

$$\frac{M'^2}{MM''}\left[\left(\frac{M''}{M'}\right)'-\frac{C_1^2M''}{6MM'^2}+\frac{3M''}{2M}\right]=0. \tag{17}$$

Since $M'\neq 0$ or E would vanish, Eqs. (16) and (17) are equivalent. Therefore we have only one equation [(16)] restricting the function $M(z)$, and hence

$$E=-MM''/8\pi a^2,$$

$$G=M\exp\left[-C_1\int dz[M(z)]^{-1}\right],$$

$$H=M\exp\left[C_1\int dz[M(z)]^{-1}\right],$$

with $M(z)$ any solution to Eq. (16), is a static solution to Eqs. (5)–(9). Therefore the task of solving a system of partial differential equations has been reduced to solving an ordinary differential equation.

IV. A SPECIAL STATIC SOLUTION

A special solution (more precisely, a family of solutions) to Eq. (16) is

$$M(z)=C_1z+C_3z^{2/3},\quad C_3 \text{ constant,}$$

and the corresponding static solution to Eqs. (5)–(9) is

$$E=C_3z^{-2/3}(C_1z^{1/3}+C_3)/36\pi a^2,$$

$$G=z^{2/3}(C_1z^{1/3}+C_3)^{1-\sqrt{3}},$$

$$H=z^{2/3}(C_1z^{1/3}+C_3)^{1+\sqrt{3}}.$$

The metric, after a trivial coordinate transformation, reads

$$dS^2=z^{-2/3}(C_1z^{1/3}+C_3)(-dt^2+dz^2)+z^{2/3}(C_1z^{1/3}+C_3)^{1-\sqrt{3}}dx^2+z^{2/3}(C_1z^{1/3}+C_3)^{1+\sqrt{3}}dy^2. \tag{18}$$

This is an example of a "semi-plane-symmetric" metric yielded by a plane-symmetric electromagnetic field, illustrating that there does exist some electrovac metric that does not share some of the symmetries of the corresponding electromagnetic field.

The following observations are also worth noting.

(A) If $a=0$, then there is no electromagnetic field (vacuum) and it follows from Eq. (12) that

$$M''=0,$$

$$M=\alpha z+\beta,\quad \alpha,\beta \text{ constants.} \tag{19}$$

In this case, Eq. (13) gives

$$N=C_2(\alpha z+\beta)^{C_1/\sqrt{3}\alpha}. \tag{20}$$

Substitution into Eq. (5) yields

$$E'/E=(C_1^2-3\alpha^2)/6\alpha(\alpha z+\beta);$$

thus

$$E=\gamma(\alpha z+\beta)^{(C_1^2-3\alpha^2)/6\alpha^2},\quad \gamma \text{ constant.}$$

This, together with the expressions for G and H obtained from (19), (20), and (10), gives the following metric:

$$dS^2 = \gamma(\alpha z+\beta)^{(C_1^2-3\alpha^2)/6\alpha^2}(-dt^2+dz^2) + C_2^{-1}(\alpha z+\beta)^{1-(C_1/\sqrt{3}\alpha)}dx^2 + C_2(\alpha z+\beta)^{1+(C_1/\sqrt{3}\alpha)}dy^2. \quad (21)$$

which, by a coordinate transformation, can be transformed into the static cylindrically symmetric vacuum metric given by Levi-Civita [formula (20.8) of Ref. 1], as expected.

(B) If $C_3 = 0$, then a trivial coordinate transformation leads the metric (18) to

$$dS^2 = z^{-1/3}(-dt^2+dz^2) + z^{1-\sqrt{3}/3}dx^2 + z^{1+\sqrt{3}/3}dy^2,$$

which is a special case ($C_1 = C_2 = \alpha = \gamma = 1, \beta = 0$) of the vacuum solution (21), in accordance with the fact that $C_3 = 0$ implies $M'' = 0$, which, in turn, implies $a = 0$, by Eq. (12).

(C) If $C_1 = 0$, then $G = H = M$. The solution (18) is reduced to a plane-symmetric metric yielded by a plane-symmetric electromagnetic field. It is well known that such a solution must be the one given by Patnaik[4] and Letelier and Tabensky[5] [formulas (13.27) and (13.28) of Ref. 1], and a straightforward calculation from our Eqs. (12) and (16) indeed leads to the same result. Thus solution (13.27) and (13.28) of Ref. 1 is a special case of ours.

ACKNOWLEDGMENTS

We would like to thank the referee for pointing out some mistakes caused mainly by our carelessness, and also a few inexact statements in the original manuscript.

This project was supported by the Natural Science Foundation of China.

[1] D. Kramer, H. Stephani, E. Herlt, and M. MacCallum *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980).
[2] J.z. Li and C.b. Liang, Gen. Relat. Gravit. **17**, 1001 (1985).
[3] Z-q. Kuang, J.z. Li, and C.b. Liang, Gen. Relat. Gravit. **19**, 345 (1987).
[4] S. Patnaik, Proc. Camb. Philos. Soc. **67**, 127 (1970).
[5] P. S. Letelier and R. R. Tabensky, J. Math. Phys. **15**, 594 (1974).

Thermodynamic perfect fluid. Its Rainich theory

Bartolomé Coll
Département de Mécanique Relativiste, UA 766 CNRS – Université de Paris VI, Paris, France

Joan Josep Ferrando
Departament de Fisica Teòrica, Universitat de València, Burjassot (València), Spain

(Received 14 April 1989; accepted for publication 12 July 1989)

The conditions for a relativistic perfect fluid to admit a thermodynamic scheme are considered, and the necessary and sufficient requirements for a metric to define a thermodynamic perfect fluid space-time are given.

I. INTRODUCTION

Let g be the metric tensor of (a region of) a space-time, S its Einstein tensor, and let (M,T) be the pair of the definition equations M of a medium and of its energy tensor T. We call here *Rainich theory of the medium* the set of necessary and sufficient conditions on g insuring the existence of the pair (M,T) such that the Einstein equations $S = T$ (Ref. 1) hold.

It is clear that this definition is nothing but a direct extension to other media of results developed by Rainich[2] for the regular electromagnetic field; in it, T is the Maxwell–Minkowski energy tensor and M is the set of the vacuum Maxwell equations.

Rainich worked out his theory about seven years after the Einstein paper on the foundation of the general theory of relativity,[3] where both media, the perfect fluid and the electromagnetic field, were explicitly considered. It seems rather paradoxical that the perfect fluid had not, up to now, been the object of a work analogous to Rainich's one on the electromagnetic field.[4] We would like to comment here on four of the factors that have contributed to this situation.

(i) The apparent simplicity of the barotropic case. A Rainich theory involves two sets of equations: a first, algebraic, set ensuring that S has the same algebraic structure as T, and a second, generally differential set translating in terms of g (and its differential concomitants) the definition equations M. In the barotropic case, the second set reduces to the expression of the functional dependence of the two algebraically independent invariant scalars of S, so that to complete the Rainich theory of the barotropic perfect fluid one only needs to know the algebraic characterization of the perfect fluid energy tensor. It is true that to obtain it is an easy task. Nevertheless, because of the Lorentzian character of the metric, it is not so easy a task as it has been evoked in the literature;[5] in addition to imposing T to have a triple eigenvalue and be of algebraic type I, one must give the condition insuring that to the simple eigenvalue corresponds a *timelike* eigenvector. For symmetric tensors, the general problems of finding the causal character of the eigenspace associated to a given eigenvalue, and its application to the perfect fluid, have been solved only very recently;[6] we will need here these results.

(ii) The apparent multiplicity of fluid thermodynamics. Both the equations of electromagnetism and relativistic continuous media have been largely analyzed, discussed, and criticized from the beginning of relativity. But, meanwhile, the matter for the electromagnetic field has been, in general, to find for it a nonlinear system.[7] For thermodynamic continuous media, the matter has been to *establish the basic system* of equations, playing the role analogous to the Maxwell ones. And, as it is well known, there are many proposed versions for this basic system. This situation would indicate that thermodymamics is not yet ripe to be incorporated in relativistic continuous media. Nevertheless, Marle's work[8] pointed out in the opposed sense: many of these versions[9] may be obtained from a unique relativistic kinetic theory, their differences corresponding essentially to the different methods used to approximate the Boltzmann equation.[10] Furthermore, any two arbitrary versions differ in at least one of the following three aspects: the form of the conserved quantities (stress energy, momentum), the thermodynamic closure (generalized Fourier law, entropy balance), and the physical definition of the variables appearing in the equations. What is important here for us is that, generically,[11] the proposed versions, when reduced to the thermodynamic perfect fluid, differ at most in the third aspect,[12] that is to say: the thermodynamic perfect fluid is generically unique, up to an eventual redefinition of some of its variables.

(iii) The apparent independence of the thermodynamics from the energy tensor. In the usual presentation of the thermodynamic perfect fluid, the thermodynamic scheme is obtained by adding to the standard energy tensor a conserved matter current, an entropy relation, and an equation of state. It would seem that the existence of these three elements could not be deduced from the metric and the energy tensor itself, so that a Rainich theory would not be possible. Nevertheless, we shall see that a unique condition from the energy tensor guarantees the existence of such a thermodynamic scheme.

(iv) The wideness of Rainich's work. The work developed by Rainich[2] to geometricize the electromagnetic field was, fortunately, superabundant. In particular, he revealed the (weighted) $(2 + 2)$ almost-product structure associated to the electromagnetic field[13] and obtained the necessary and sufficient equations that the volume element U of the structure must verify in order to have a solution of the Maxwell equations. As similarly, a perfect fluid has an associated (weighted) $(1 + 3)$ almost-product structure, the extension of the Rainich work to the perfect fluid would implicate correspondingly the obtainment of the necessary and sufficient

 0022-2488/89/122918-05$02.50

equations that the volume element u (Ref. 14) of the structure must verify in order to have a solution of the hydrodynamic equations.[15] Rainich considered also the uniqueness of the Maxwell field, which he solved, but globally, the corresponding uniqueness of the thermodynamic scheme would need to introduce some rather artificial *ad hoc* conditions. It is to palliate these features that we have chosen our above definition of a *Rainich theory*, which includes only *a part* of Rainich's work.

The above analysis shows that a Rainich theory of the thermodynamic perfect fluid may be boarded. But, is it worthwhile? We think there are, at least, four reasons to construct it: (i) A general medium may not admit a Rainich theory. What are the media admitting it? According to Misner and Wheeler's geometrical point of view,[16] the existence of a Rainich theory would be a necessary condition for such a medium to be realistic. In any case, these media admit such a particular physical characterization [see (iv) below] that the question about the *existence* of a Rainich theory is already an interesting question. (ii) A Rainich theory offers an alternative method[17] of integration of the Einstein equations: the set of all unknowns being reduced to the *metric coefficients*, the completed system of equations (the Einstein ones plus those corresponding to the set M) is now an *overdetermined system* (unless $M=\varnothing$), and the corresponding methods of compatibility conditions may be applied. (iii) This last consideration may be of interest in the study of those conjectures about perfect fluids which do not restrict the space of solutions of the hydrodynamic (test) equations, but restrict the space-times with which they are coupled;[18] due to this fact, it seems plausible that the Rainich theory may help their analysis. (iv) In the penultimate phase of a Rainich theory, the set M is reduced to a system of equations on the energy tensor: a medium which admits a Rainich theory is a medium which may be completely described in terms of the sole energy tensor variables. This fact may be of interest for practical purposes;[19] it is certainly of interest for conceptual and epistemological analysis.[20]

In Sec. II we find a simple, necessary, and sufficient condition for a perfect fluid to admit a thermodynamic scheme (Theorem 1), and in Sec. III we give the equations of the Rainich theory for it (Theorem 4). The barotropic and polytropic particular cases are given explicitly (Corollaries 2 and 3).

The results without proof of this paper were communicated to the Spanish relativistic meeting E.R.E. 87.[21]

II. CHARACTERIZATION OF THE THERMODYNAMIC PERFECT FLUID

A. Thermodynamic scheme

The energy conservation equations $\delta T=0$ (Ref. 22) for a perfect fluid $T=(\rho+p)u\otimes u-pg$ (Ref. 23) may be written

$$dp=(\rho+p)a+\dot{p}u, \tag{1}$$

$$(\rho+p)\theta+\dot{\rho}=0, \tag{2}$$

where a and θ are, respectively, the *acceleration* and the *expansion* of u: $a\equiv\dot{u}$, $\theta\equiv-\delta u$.

From the evolution point of view, the system (1), (2) is open. A usual algebraic closure is obtained by imposing a *barotropic* condition $\rho=\rho(p)$; however acceptable in some cases, it is known that this condition is too restrictive in many other interesting physical situations.[24] The standard general closure to the energy conservation system is the differential closure consisting of a *thermodynamic scheme.*

Let r be the (Eckart) *matter density*[25] of the fluid; denoting by $E\equiv\rho-r$ the *internal energy density* and by $\epsilon\equiv E/r$ the *specific internal energy*, one has

$$\rho=r(1+\epsilon). \tag{3}$$

When an equation of state, depending only on the internal structure of the fluid, is known,

$$\epsilon=\epsilon(p,r), \tag{4}$$

the one-form $d\epsilon+p\,dv$ is integrable, $v=1/r$ being the *specific volume*. Then, the *temperature* Θ of the fluid may be identified, by a classical argument, with an integrant divisor, and the *specific entropy* s is given, up to an additive constant, by

$$\Theta\,ds=d\epsilon+p\,dv. \tag{5}$$

As far as creation or annihilation of baryons do not take place,[26] the equation of conservation of matter holds:

$$\delta(ru)=0. \tag{6}$$

The relation (5) allows us to write Eq. (2) in the form

$$\delta(ru)=[r\Theta/f]\dot{s}, \tag{7}$$

where $f\equiv1+\epsilon+pv$ is the *enthalpy index* of the fluid;[27] Eq. (7) shows the intimate relation existing between the local adiabatic motion and matter conservation.

It is interesting to note that, while in classical thermodynamics, because of the nonequivalence between mass and energy, the internal energy E_V of a given volume V is determined *up* to an additive constant; in relativistic thermodynamics this energy is *univocally* determined once the matter density is given. However, this fact does not imply that the zero of the internal energy E_V be fixed in relativity; because of its noninertial character, the matter density is only determined up to a constant factor and, as a consequence, there still exists indeterminacy of E_V by an additive constant. Thus, if M and $M'=kM$ denote two mass balances[28] of the particles contained in V one has $r=M/V$, $r'=M'/V$ and it results in $E'_V=(1-k)M+E_V$. This observation is pertinent, for example, in the study of reaction fronts, where it allows us to localize conveniently the binding specific energy of the reaction,[29] or in the study of those hot perfect gases for which the limit $\Theta\to0$ is meaningless.[30]

B. Characterization theorem

Einstein equations for the thermodynamic perfect fluid being not easy to solve, one often, in a first step, looks for a solution to the general perfect fluid and, once obtained, in a second step, considers the admissibility by this solution of a thermodynamic scheme.

The existence of a thermodynamic scheme for a perfect fluid verifying Eqs. (1) and (2) amounts to the existence of functions ϵ and r such that Eqs. (3), (4), and (6) hold. As a consequence of (6), the equation in the function F,

$$\dot{F} = \theta, \tag{8}$$

must admit at least one solution of the form

$$F(x) = F(\rho(x), p(x)). \tag{9}$$

If this is the case, the one-form $\Gamma \equiv d\rho + (\rho + p)dF$ is integrable, the variables r, ϵ, and f may be defined by $r \equiv e^{-F}$, $\epsilon = \rho e^F - 1$, and $f = (\rho + p)e^F$, and, to every integral factor $D > 0$ for Γ, it may be associated an absolute temperature $\Theta = e^F/D$ and a specific entropy s such that $ds = D\Gamma$. Thus we have the following.

Lemma 1: The necessary and sufficient condition for a perfect fluid to admit a thermodynamic scheme is the existence of solutions of the form $F = F(\rho,p)$ to the equation $\dot{F} = \theta$. Then, every pair $\{F,D\}$ where $D > 0$ is an integral factor of the one-form $d\rho + (\rho + p)dF$, determines a thermodynamic scheme.

For a thermodynamic perfect fluid, Eq. (9) may be written in the equivalent form

$$dF = h(\rho,p)d\rho + g(\rho,p)dp, \tag{10}$$

which implies

$$\dot{F} = h\dot{\rho} + g\dot{p}. \tag{11}$$

On the other hand, from (2) and (8) one obtains $\dot{\rho} + (\rho + p)\dot{F} = 0$, so that (11) becomes

$$h\dot{\rho} + g\dot{p} = -\dot{\rho}/(\rho + p). \tag{12}$$

Suppose $\dot{\rho} = 0$; from (12) it is either $\dot{p} = 0$ or $g = 0$. If $\dot{p} = 0$, every arbitrary function $F = F(\rho,p)$ verifies (8); meanwhile if $g = 0$, they are the functions of the form $F = F(\rho)$ which verify (8). Suppose $\dot{\rho} \neq 0$; then if $g = 0$, from (12) we have $h = -1/(\rho + p)$ and (10) implies that $p = p(\rho)$: the fluid is barotropic. Finally, if $g \neq 0$ we have from (12)

$$\dot{p}/\dot{\rho} = [-1/g(\rho,p)]\{1/(\rho + p) + h(\rho,p)\}, \tag{13}$$

which implies that $\dot{p}/\dot{\rho}$ is a function of state:

$$\dot{p}/\dot{\rho} \equiv \chi(\rho,p). \tag{14}$$

Conversely, if (14) is verified, we can consider the following first-order partial differential equation:

$$F_\rho' + \chi F_p' = -1/(\rho + p). \tag{15}$$

Then, because of (2) and (11), every solution $F(\rho,p)$ to it is a solution to (8). Differentiating (14) and multiplying by $\dot{\rho}^2$, we obtain an equivalent expression which is identically satisfied for $\dot{\rho} = 0$, and thus we have the following.

Theorem 1: The necessary and sufficient condition for a perfect fluid $T = (\rho + p)u \otimes u - pg$ to admit a thermodynamic scheme is

$$(\dot{\rho}d\dot{p} - \dot{p}d\dot{\rho}) \wedge d\rho \wedge dp = 0. \tag{16}$$

Let $\lambda = \lambda(\rho,p)$ and $\mu = \mu(\rho,p)$ be two independent thermodynamic variables, $J = J(\lambda,\mu;\rho,p) \neq 0$. We know that $d\lambda \wedge d\mu = J\, d\rho \wedge dp$ so that, evaluating $\dot{\rho}\, d\dot{p} - \dot{p}\, d\dot{\rho}$ up to terms in $d\lambda$ and $d\mu$, one easily finds the following.

Corollary 1: Let $T = (\rho + p)u \otimes u - pg$ be a perfect fluid and $\lambda = \lambda(\rho,p)$ and $\mu = \mu(\rho,p)$ two independent thermodynamic variables. T admits a thermodynamic scheme iff

$$(\dot{\lambda}\, d\dot{\mu} - \dot{\mu}\, d\dot{\lambda}) \wedge d\lambda \wedge d\mu = 0. \tag{17}$$

III. RAINICH THEORY FOR THE THERMODYNAMIC PERFECT FLUID

Remember that if S is the Einstein tensor of the metric g, and if $\{M,T\}$ is the pair of definition equations of a medium, with T the energy tensor and M the complementary equations, then we call *Rainich theory of the medium* the set of conditions on g and on its differential concomitants, which ensure the existence of the pair $\{M,T\}$ verifying the Einstein equations $S = T$.

As everyone knows, the genuine Rainich theory concerns the regular Einstein–Maxwell equations. The pair $\{M,T\}$ is constituted of the set M of the vacuum Maxwell equations, $\delta F = \delta{*F} = 0$, and of the Minkowski energy tensor T, $2T = F^2 + (*F)^2$. Let us write $\mathbf{R} = \mathrm{Ric}(g)$, $\mathbf{r} = \mathrm{tr}\,\mathbf{R}$, $\mathbf{s} = \mathrm{tr}\,\mathbf{R}^2$, and define the one-form $\psi = \mathbf{s}^{-1} \cdot *(\nabla\mathbf{R} \times \mathbf{R})$ (Ref. 31); the Rainich theory of the regular Einstein–Maxwell space-times consists[2] of the algebraic equations $\mathbf{r} = 0$, $\mathbf{R}^2 = (1/4)\mathbf{s}g \neq 0$, and the differential equations $d\psi = 0$; any metric g verifying these conditions defines an Einstein–Maxwell space-time corresponding to a regular solution to the source-free Maxwell equations.

A. Algebraic conditions

Let us consider the thermodynamic perfect fluid space-times. The pair $\{M,T\}$ consists now of the set M of Eq. (16), ensuring the existence of a thermodynamic scheme, and of the energy tensor $T = (\rho + p)u \oplus u - pg$.

The algebraic set of equations characterizing the perfect fluid energy tensor T were partially given by Taub[5]; we will present here a slightly different form of his result.[32] Let tr and I be, respectively, the trace operator and the identity over the second rank tensors; consider the *trace-removing* operator $\mathbf{Q} \equiv I - (1/4)g$ tr, and, for any second rank tensor T write $\mathbf{t} \equiv \mathrm{tr}\, T$ and $\mathbf{s} \equiv \mathrm{tr}\, T^2$; then we have the following lemma.

Lemma 2 (Taub's lemma): A second rank symmetric tensor T is of algebraic type I and admits a strictly triple eigenvalue if, and only if, it satisfies the following relations:

$$\mathbf{Q}(T^2 - \chi T) = 0,$$

$$4\mathbf{s} > \mathbf{t}^2, \quad 2\chi \neq \mathbf{t}.$$

This result says nothing about the causal character of the associated eigenvectors. Regarding them, the following lemma has been shown elsewhere.[6]

Lemma 3: A necessary and sufficient condition for the eigenvector associated to the single eigenvalue to be timelike is that the expression

$$\epsilon\{2i^2(x)T-\chi\}$$

be positive for any timelike vector x, where ϵ denotes the sign of the quantity $\mathbf{t}^3 - 6\mathbf{ts} + 8 \text{ tr } T^3$.

We are assuming that the perfect fluids considered here correspond to a macroscopic level of description. For this reason it is plausible to submit them to the *Plebański energy conditions*, which state that, for any observer, the energy density is positive definite and the Poynting vector is non-spacelike.[33] In terms of ρ and p, the Plebański conditions for the perfect fluid are equivalent to the inequalities $-\rho<p\leqslant\rho$, which may in turn be expressed as $\epsilon=1$ and $\chi\geqslant0$. Taking into account the above two lemmas, one obtains the following theorem.[6]

Theorem 2: In a space-time of signature -2, a second rank symmetric tensor T defines algebraically a perfect fluid submitted to the Plebański energy conditions if, and only if,

$$\mathbf{Q}(T^2-\chi T)=0,$$
$$4\mathbf{s}>\mathbf{t}^2, \quad \mathbf{t}<2\chi\geqslant0, \tag{18}$$
$$2i^2(x)T>\chi,$$

where $\mathbf{t}=\text{tr } T$, $\mathbf{s}=\text{tr } T^2$, $\mathbf{Q}=I-(1/4)g \text{ tr}$, and x is any timelike unit vector.

The intrinsic decomposition of T may then be obtained according to the following result.[6]

Theorem 3: The total energy ρ, the pression p, and the direction of the unit velocity u of a perfect fluid energy tensor T are given by

$$\rho=1/2(3\chi-\mathbf{t}), \quad p=1/2(\chi-\mathbf{t}), \tag{19}$$
$$u\propto i(x)T+px,$$

where

$$\chi\equiv1/2(\mathbf{t}+z), \quad z=[(4\mathbf{s}-\mathbf{t}^2)/3]^{1/2}, \tag{20}$$

and x is any timelike vector.

B. General case

Let us write $\mathbf{R}\equiv\text{Ric}(g)$, $\mathbf{r}\equiv\text{tr } \mathbf{R}$, and $\mathbf{s}\equiv\text{tr } \mathbf{R}^2$; from Einstein equations, we have (Ref. 1) $\mathbf{R}=T-1/2\mathbf{t}g$ so that $\mathbf{r}=-\mathbf{t}=-\text{tr } T$ and $\mathbf{s}=\text{tr } T^2$. Taking into account these values in definitions (20) and the expressions (19), the Jacobian of $\mathbf{r}$ and $\mathbf{s}$ with respect ρ and p is given by $J(\mathbf{r},\mathbf{s};\rho,p)=-6(2\chi+\mathbf{r})$, which does not vanish under the third of the assumptions (18). Thus according to Corollary 1, the perfect fluid admits a thermodynamic iff (17) holds for $\lambda=\mathbf{r}$ and $\mu=\mathbf{s}$.

If $\dot{\mathbf{r}}=0$, (17) holds trivially; if $\dot{\mathbf{r}}\neq0$, (17) is equivalent to

$$d(\dot{\mathbf{s}}/\dot{\mathbf{r}})\wedge d\mathbf{r}\wedge d\mathbf{s}=0, \tag{21}$$

and we have to evaluate the scalar $\dot{\mathbf{s}}/\dot{\mathbf{r}}$ in terms of the concomitants $\mathbf{R}$, $\mathbf{r}$, and $\mathbf{s}$ of the space-time metric g. To do it, let us observe that the direction of the unit velocity u, as given by the third of the relations (19), is the image of the endomorphism U given by

$$U\equiv T+pg=\mathbf{R}+1/4(z-\mathbf{r})g, \tag{22}$$

so that $u=\lambda i(y)U$, where y is *any* vector field not belonging to the kernel of U: $i(y)U\neq0$. Thus, for any function f we have $f=i(u)df=i(df)u=\lambda i(df)i(y)U$; in particular, taking $f=\mathbf{r}$ and $y=d\mathbf{r}$, we have $\dot{\mathbf{r}}=\lambda i^2(d\mathbf{r})U$, which vanishes only if $d\mathbf{r}$ belongs to the kernel of U. Also, for $f=\mathbf{s}$ we have $\dot{\mathbf{s}}=\lambda i(d\mathbf{r})i(d\mathbf{s})U$ and, consequently,

$$\dot{\mathbf{s}}/\dot{\mathbf{r}}=i(d\mathbf{r})i(d\mathbf{s})U/i^2(d\mathbf{r})U. \tag{23}$$

On the other hand, let us note that the three inequalities expressed by the second and the third of the relations (18) are equivalent to $4\mathbf{s}>\mathbf{r}^2$ and $z\geqslant\mathbf{r}$, which are nothing but $-2\mathbf{s}^{1/2}<\mathbf{r}\leqslant\mathbf{s}^{1/2}$, as it is not difficult to show.

Finally, taking into account this result, Theorem 2, and expressions (21) and (22), we have the following.

Theorem 4 (Rainich theory of the thermodynamic perfect fluid)**:** A metric g defines a thermodynamic perfect fluid space-time with the Plebański energy conditions if, and only if, it verifies

$$-2\mathbf{s}^{1/2}<\mathbf{r}\leqslant\mathbf{s}^{1/2},$$
$$\mathbf{R}^2-2\pi\mathbf{R}+1/4(2\pi\mathbf{r}-\mathbf{s})g=0,$$
$$i^2(x)\mathbf{R}>\pi,$$

and

$$i^2(d\mathbf{r})U=0$$

or

$$d[i(d\mathbf{r})i(d\mathbf{s})U/i^2(d\mathbf{r})U]\wedge d\mathbf{r}\wedge d\mathbf{s}=0,$$

where $\mathbf{R}\equiv\text{Ric}(g)$, $\mathbf{r}\equiv\text{tr}\mathbf{R}$, $\mathbf{s}\equiv\text{tr } \mathbf{R}^2$, $\pi\equiv1/4\{\mathbf{r}+[(4\mathbf{s}-\mathbf{r}^2)/3]^{1/2}\}$, $U\equiv\mathbf{R}+(\pi-\mathbf{r}/2)g$ and x is an arbitrary unit timelike vector field.

As a corollary of Theorem 3, the total energy density ρ, the pression p, and the direction of the unit velocity u of the perfect fluid are then give by

$$\rho=3\pi-\mathbf{r}, \quad p=\pi, \quad u\propto i(x)\mathbf{R}+(\pi-\mathbf{r}/2)x.$$

C. Barotropic case

Let us note that in the barotropic case, since the Jacobian $J(\mathbf{r},\mathbf{s};\rho,p)$ does not vanish, the condition $d\rho\wedge dp=0$ is equivalent to $d\mathbf{r}\wedge d\mathbf{s}=0$. Thus we have the following corollary.

Corollary 2: A metric g is a barotropic perfect fluid space-time with the Plebański energy conditions if, and only if, it verifies the algebraic relations of Theorem 4 and the differential equation $d\mathbf{r}\wedge d\mathbf{s}=0$.

Also, in the case of a polytropic fluid of index γ, $p=(\gamma-1)\rho$, it is easy to show the following result.

Corollary 3: A metric g defines a polytropic perfect fluid space-time with the Plebański energy conditions if, and only if, it verifies the algebraic relations of Theorem 4 and the equation $d(\mathbf{s}/\mathbf{r}^2)=0$. Then, if $\mathbf{s}/\mathbf{r}^2=c$, the polytropic index is given by

$$\gamma=\{4c-1+[(4c-1)/3]^{1/2}\}/(3c-1).$$

ACKNOWLEDGMENTS

One of the authors (J.J.F.) would like to express his thanks to the "Conselleria de Cultura, Educació i Ciència de la Generalitat Valenciana" for the partial support of this work.

[1]In pertinent units.
[2]G. Y. Rainich, Trans. Am. Math. Soc. **27**, 106 (1925).
[3]A. Einstein, Ann. Phys. **49**, 769 (1916).
[4]At least, we have not been able to find it.
[5]The Taub conditions for a (1,1) tensor to be the energy tensor for a perfect fluid [A.H. Taub, "Relativistic Hydrodynamics," in *Lectures in Applied Mathematics* (Am. Math. Soc., Providence, RI), 1967, Vol. 8, p. 170] though not explicitly stated, apply *only in absence of a given metric*; see our Sec. III A.
[6]J. A. Morales, Ph.D. thesis, València, 1988; see also C. Bona, B. Coll, and J. A. Morales, "Caracterización algebraica de un 2-tensor simétrico," in *Actas de los E.R.E.86* (Pub. Univ. València, València, to be published).
[7]Including the Maxwell equations as linear approximation.
[8]C. Marle, Ann. Inst. H. Poincaré **10**, 67 (1969); **10**, 127 (1969).
[9]The Eckart and the Landau–Lifchitz ones, among others.
[10]Marle considers the relativistic versions of the Chapman–Enskog and the Grad classical, methods.
[11]Here, "generically" means "for *almost all* the versions that have been proposed in the literature." Of course, there are always some exceptions; for example, the Arzeliés fluids [H. Arzeliés, *Fluides Relativistes* (Masson, Paris, 1971)].
[12]For example, the Catteneo fluids [C. Catteneo, Rend. Accad. Naz. dei Lincei **46**, Sér. VIII, 699 (1969)].
[13]Rainich called it the *skeleton* of the electromagnetic field.
[14]Here, this volume element is nothing but the unit velocity of the fluid.
[15]This task is not easy. Restricted to the barotropic fluid, it induces an eightfold classification of the unit velocity (see B. Coll and J. J. Ferrando, "On the velocities of the barotropic perfect fluids," to be published).
[16]C. W. Misner and J. A. Wheeler, Ann. Phys. **2**, 525 (1957).
[17]Usually, the corresponding differential equations are presented in the form of Cauchy or underdetermined systems for the metric coefficients *and* some other additional unknowns (pression, electromagnetic field, etc.).
[18]This is the case for Lichnerowicz's conjecture on spherical symmetry under appropriate asymptotic conditions [see H. P. Kunzle, Commun. Math. Phys. **20**, 85 (1971) and references therein], or the Treciokas–Ellis conjecture on vorticity-free or expansion-free consequences under distortion-free conditions [see R. Treciokas and G. F. R. Ellis, Commun. Math. Phys. **23**, 1 (1971), or the more recent analysis by C. B. Collins, J. Math Phys. **26**, 2009 (1985)].
[19]As an application to the Maxwell case, see, for example, B. Coll, F. Fayos, and J. J. Ferrando, J. Math. Phys. **28**, 1075 (1987).
[20]Fortunately, the development of field theory began, historically, with *force* field variables and not with *energy* field variables. Otherwise Maxwell equations should remain undiscovered; to think so, a glance on the nonlinear Rainich complexion equations is largely sufficient.
[21]B. Coll and J. J. Ferrando, "Fluido perfecto termodinamico. Su teoria 'à la Rainich'," in *Actas de los E.R.E.87* (Pub. Inst. Astrof. de Canarias, La Laguna, Spain, 1988).
[22]δ, $i(u)$, $\perp(u)$, ∇, d, $*$, denote, respectively, the divergence, interior product, normal projection, covariant derivative, exterior differentiation, and Hodge dual operators. Newton's notation is used for timelike derivatives: $\dot{x} \equiv i(u)\nabla$, x being any tensorial quantity.
[23]Of course, u is the *proper unit velocity* of the fluid, p the *pression*, and ρ the *total energy density*.
[24]C. B. Collins, J. Math. Phys. **26**, 2009 (1985).
[25]Also called *rest mass density, proper mass density, baryonic* (average) *mass density* or, simply *density* of the fluid.
[26]The definition of r as a mass balance of the baryonic number allows us to include in this scheme the study of the propagation of chemical reactions fronts; see B. Coll, Ann. Inst. H. Poincaré **25**, 363 (1976).
[27]See A. Lichnerowicz, *Relativistic Hydrodynamics and Magnetohydrodynamics* (Benjamin, New York, 1967).
[28]Molecular, atomic, or baryonic mass balances.
[29]See the paper quoted in Ref. 26.
[30]In such a case, one does not have necessarily $\epsilon = C_v \Theta \rightarrow 0$, and every γ-law, $p = (\gamma - 1)\rho$, may be interpreted as a polytropic perfect gas [see B. Coll, C. R. Acad. Sci. Paris A **273**, 1185 (1971)].
[31]The symbol $\times$ denotes the *cross product*; contraction of the adjacent spaces of the tensor product. Of course, the operator $*$ selects the antisymmetric part of $\nabla \mathbf{R} \times \mathbf{R}$.
[32]J. A. Morales, Ref. 6.
[33]See J. Plebański, Acta Phys. Pol. **26**, 963 (1964), especially his prudent analysis (pages 1011 and 1012) on the validity of his two conditions. In *The large scale structure of space-time* (Cambridge U.P., Cambridge, 1973), S. W. Hawking and G. F. R. Ellis call them the *weak* and *dominant* energy conditions, seeming to be unaware of Plebański's work.

On the Cauchy problem for the theory of gravitation with nonlinear Lagrangian

Andrzej Jakubiec
Institute of Mathematical Physics, University of Warsaw, ul. Hoża 74, 00-682, Warsaw, Poland
Jerzy Kijowski
Institute for Theoretical Physics, Polish Academy of Sciences, Al. Lotników 32/46, 02-668, Warsaw, Poland

(Received 21 February 1989; accepted for publication 21 June 1989)

For a theory of gravitation with nonlinear Lagrangian it is shown that the Cauchy problem is well posed.

I. INTRODUCTION

In recent years, a great deal of interest has been devoted to fourth-order theories of gravitation based on nonlinear Lagrangians:

$$L(g,\Gamma,R,\phi,\partial\Phi) = -(1/2\kappa)\sqrt{-\det g}\,F(g_{\mu\nu},R_{\mu\nu}) + L_{\rm mat}(g,\Gamma,\phi,\partial\phi), \tag{1}$$

where $L_{\rm mat}$ is a matter Lagrangian, $R_{\mu\nu} = R_{\mu\nu}(g)$ is the Ricci curvature of a metric tensor g, Γ is the Levi-Civita connection of g, and $\kappa = 8\pi G$ is the gravitational constant. Different generalizations of the Einstein theory can be obtained by a choice of a special form of the function F; e.g., the quadratic Lagrangian

$$F(g_{\mu\nu},R_{\mu\nu}) = aR + bR^2 + cg^{\alpha\beta}g^{\mu\nu}R_{\alpha\mu}R_{\beta\nu}, \tag{2}$$

has been used by many authors.[1] The Cauchy problem for theories derived from (2) has been examined by a number of authors.[2] The purpose of the present paper is to study the Cauchy problem for a general theory based on a Lagrangian (1).

II. THEORY

The Euler–Lagrange equation for the theory

$$\frac{\delta L}{\delta g_{\mu\nu}} = 0 \tag{3}$$

has the form

$$G^{\mu\nu} + \kappa T^{\mu\nu} = 0, \tag{4}$$

where

$$G^{\mu\nu} = -\frac{1}{\sqrt{-\det g}}\frac{\delta(\sqrt{-\det g}\,F)}{\delta g_{\mu\nu}} \tag{5}$$

and

$$T^{\mu\nu} = -\frac{1}{2\sqrt{-\det g}}\frac{\delta L_{\rm mat}}{\delta g_{\mu\nu}}. \tag{6}$$

Let us denote by $\pi^{\mu\nu}$ an auxiliary quantity:

$$\pi^{\mu\nu} = \frac{\partial F}{\partial R_{\mu\nu}}(g,R) \tag{7}$$

(it is interesting that the quantity π can be interpreted[3] as a new metric tensor but this point of view is not relevant for the present paper). The fourth-order differential equation (3) can be replaced by a system of second-order differential equations for g and π treated as independent quantities. More precisely, Eq. (7) is already a second-order differential equation for g. Moreover, due to definition (5) we have

$$G^{\mu\nu} = \frac{1}{2}D_\alpha D_\beta(g^{\mu\nu}\pi^{\alpha\beta} + g^{\alpha\beta}\pi^{\mu\nu} - g^{\beta\mu}\pi^{\alpha\nu} - g^{\beta\nu}\pi^{\alpha\mu}) - \frac{1}{2}g^{\mu\nu}F - \frac{\partial F}{\partial g_{\mu\nu}}. \tag{8}$$

Therefore Eq. (4) can also be treated as a second-order equation for π. It is interesting to notice that the second-order differential operator acting on $\pi^{\mu\nu}$ is universal and does not depend on a special choice of a Lagrangian function (1). It is easy to check the scalar invariance of the function F implies the following identity:

$$\frac{\partial F}{\partial g_{\mu\nu}} = -\pi^{\alpha\nu}R_\alpha{}^\mu. \tag{9}$$

Therefore expression (8) can be rewritten as

$$G^{\mu\nu} = \tfrac{1}{2}D_\alpha D_\beta(g^{\mu\nu}\pi^{\alpha\beta} + g^{\alpha\beta}\pi^{\mu\nu} - g^{\beta\mu}\pi^{\alpha\nu} - g^{\beta\nu}\pi^{\alpha\mu}) - \tfrac{1}{2}g^{\mu\nu}F + \pi^{\alpha\nu}R_\alpha{}^\mu. \tag{10}$$

In a similar way the invariance of the Lagrangian (1) implies the Noether identity

$$D_\nu(G^{\mu\nu} + \kappa T^{\mu\nu}) = 0. \tag{11}$$

Let us assume that our four-dimensional space-time M is a topological product of a three-dimensional manifold Σ and a real line (time axis). Moreover, we assume that the hypersurfaces Σ_t corresponding to $t = \text{const}$ are spacelike. We introduce the following quantities:

$$X_{ij} = R_{ij}, \tag{12}$$

$$Y_0 = R_{00} - g^{ij}R_{ij}/g^{00}, \tag{13}$$

$$Y_j = 2(R_{0j} + g^{0i}R_{ij}/g^{00}). \tag{14}$$

Let us notice that they do not contain any information about the second derivatives $\partial_{00}g_{0\mu}$ with respect to the time variable. The corresponding derivatives of the spatial components of the metric (i.e., $\partial_{00}g_{ij}$) are contained only in X. Let us denote by $\widetilde{F}$ a new function that is given by

$$\widetilde{F}(g,X,Y) = F(g,R(g,X,Y)). \tag{15}$$

Let us define

$$H^{ij} = \frac{\partial \widetilde{F}}{\partial X_{ij}}(g,X,Y). \tag{16}$$

It is easy check that

$$H^{ij} = \pi^{ij} + (1/g^{00})(\pi^{00}g^{ij} - \pi^{0i}g^{0j} - \pi^{0j}g^{0i}) \tag{17}$$

and

$$\pi^{0\mu} = \frac{\partial \widetilde{F}}{\partial Y_{\mu}}(g,X,Y). \tag{18}$$

We assume that the components $g_{0\mu}$ of the metric g (lapse and shift[4]) are known in the entire space-time M. We will prove that the Cauchy problem for field equations (4) and (7) is well posed for the following set of Cauchy data (CD) on Σ: the three-dimensional metric tensor g_{ij}, the three-dimensional tensor H^{ij}, the matter field ϕ, and their first time derivatives; $\partial_0 g_{ij}$, $\partial_0 H^{ij}$, and $\partial_0 \phi$.

Our analysis will be based on the following regularity conditions, which we impose on $\widetilde{F}$ (i.e., on F):

$$\det\left(\frac{\partial \widetilde{F}}{\partial X_{ij}}\right) \neq 0, \tag{19}$$

$$\det\left(\frac{\partial^2 \widetilde{F}}{\partial X_{ij}\,\partial X_{kl}}\right) \neq 0, \tag{20}$$

where the second derivative of $\widetilde{F}$ is treated like a (6×6) matrix. Condition (20) guarantees the existence of a solution of the six algebraic equations (16) with respect to X_{ij} (i.e., R_{ij})

$$R_{ij} = \mathscr{R}_{ij}(g,Y,H), \tag{21}$$

where $\mathscr{R}_{ij}$ are functions uniquely defined by the Lagrangian (1). Field equations (7) are equivalent to (21) and (18), rewritten as follows:

$$\pi^{0\mu} = \mathscr{P}^{\mu}(g,\partial_0 g,H) := \frac{\partial \widetilde{F}}{\partial Y_{\mu}}(g,\mathscr{R}_{ij}(g,Y,H),Y). \tag{22}$$

Equation (21) enables us to calculate the second derivatives $\partial_{00}g_{ij}$ on Σ in terms of g, $\partial_0 g$, and H. It is important that the time derivatives of H do not appear. This allows us to calculate $\partial_0 \pi^{0\mu}$ (on Σ) differentiating Eq. (22):

$$\partial_0 \pi^{0\mu} = Q^{\mu}(g,\partial_0,g,H,\partial_0 H). \tag{23}$$

Using (23) we rewrite the field equations (4) so that the time derivatives of the second order appear explicitly. The spatial part ($G^{ij} + \kappa T^{ij} = 0$) has the form

$$\begin{aligned}\tfrac{1}{2}g^{00}\,\partial_{00}H^{ij} &= \tfrac{1}{2}g^{ij}(F + H^{kl}R_{kl}) - 3\pi^{i\alpha}R_{\alpha}{}^{j}\\ &\quad - \tfrac{1}{2}(1/g_{00})(\pi^{00}g^{ij} - \pi^{0i}g^{0j} - \pi^{0j}g^{0i})R\\ &\quad + G_1(\mathrm{CD}),\end{aligned} \tag{24}$$

where by $G_1(\mathrm{CD})$ we denote the quantity that is determined by the Cauchy data on Σ. The remaining part of (4)

$$G^{0\mu}(g,\partial g,H,\partial H) + \kappa T^{0\mu} = 0 \tag{25}$$

gives us the constraint equation for the Cauchy data:

$$\tfrac{1}{2}g^{0\mu}(F + H^{ij}R_{ij}) - \pi^{0\mu}R + G_2(\mathrm{CD}) = 0. \tag{26}$$

It is known that the Noether identity (11) implies the constraint conservation. Indeed, we have

$$\begin{aligned}\partial_0(G^{0\mu} + \kappa T^{0\mu}) &= -\Gamma_{\lambda}{}^{0}{}_{0}(G^{\lambda\mu} + \kappa T^{\lambda\mu})\\ &\quad - \Gamma_{\lambda}{}^{\mu}{}_{0}(G^{0\lambda} + \kappa T^{0\lambda})\\ &\quad - D_j(G^{j\mu} + \kappa T^{j\mu}).\end{aligned} \tag{27}$$

As a result of the vanishing of $(G^{ij} + \kappa T^{ij})$, the above equations can be considered as four linear homogeneous differential equations for $(G^{0\mu} + \kappa T^{0\mu})$ with the vanishing Cauchy data on Σ. This implies $G^{0\mu} + \kappa T^{0\mu} = 0$ identically.

[1] A. D. Sakharov, Dokl. Akad. Nauk. SSSR **177**, 70 (1967); F. W. Hehl and G. D. Kerlick, Gen. Relativ. Gravit. **9**, 691 (1978); K. I. Macrae and R. J. Riegert, Phys. Rev. D **24**, 2555 (1981); A Frenkel and K. Brecher, *ibid.* **26**, 368 (1982); V. Muller and H.-J. Schmidt, Gen. Relativ. Gravit. **17**, 769, 971 (1985); H. J. Schmidt, Astron. Nachr. **308**, 183 (1987).

[2] K. S. Stelle, Gen. Relativ. Gravit. **9**, 353 (1978); P. Teyssandier and Ph. Tourrenc, J. Math. Phys. **24**, 2793 (1983).

[3] G. Magnano, M. Ferraris, M. Francaviglia, Gen. Relativ. Gravit. **19**, 465 (1987); A. Jakubiec and J. Kijowski, Phys. Rev. D **37**, 1906 (1988).

[4] R. Arnowitt, S. Deser, and C. Misner, in *Gravitation: An Introduction to Current Research*, edited by L. Witten (Wiley, New York, 1962); C. Misner, K. Thorne, and J. A. Wheeler, *Gravitation* (Freeman, San Francisco, 1973).

Teukolsky–Starobinsky identities for arbitrary spin

E. G. Kalnins
Department of Mathematics and Statistics, University of Waikato, Hamilton, New Zealand
W. Miller, Jr.
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455
G. C. Williams
Department of Mathematics and Statistics, University of Waikato, Hamilton, New Zealand

(Received 18 October 1988; accepted for publication 21 June 1989)

The Teukolsky–Starobinsky identities are proven for arbitrary spin s. A pair of covariant equations are given that admit solutions in terms of Teukolsky functions for general s. The method of proof is shown to extend to the general class of space-times considered by Torres del Castillo [J. Math. Phys. **29**, 2078 (1988)].

I. INTRODUCTION

Gravitational and electromagnetic perturbations in Kerr geometry are known to be intimately connected to Teukolsky functions.[1] This came about because of investigations by Teukolsky who showed that in the Newman–Penrose formalism[2] separable solutions were possible for certain Maxwell and Weyl scalars in Kerr geometry. The resulting separable solutions are known as Teukolsky functions. In addition to the problem of gravitational and electromagnetic perturbations these functions reappear when the neutrino[3] and Rarita–Schwinger[4] equations are solved in a background of Kerr geometry. These functions satisfy what are known as the Teukolsky–Starobinsky identities. In this work we prove these identities for any spin s. This result is established relatively easily. One of the difficulties with the Kerr metric is that for $s>2$ these functions do not appear to come from any covariant equation. We rectify this situation by introducing covariant equations that admit Teukolsky functions for general s as their solutions. No claim is made that these equations have physical significance. Finally we note that the method of proof applies to the more general class of space-time studied by Torres del Castillo[5] who proved these results for $s\leqslant 2$.

II. THE TEUKOLSKY–STAROBINSKY IDENTITIES

We consistently use in this article the spinor notation of Penrose and Rindler[6] and the null tetrad formalism of Chandrasekhar.[7] Specifically we restrict ourselves to the Kinnersley null tetrad of vectors with components

$$\begin{aligned} l^a &= (1/\sqrt{2}\Delta)(r^2+a^2,\Delta,0,a),\\ n^a &= (1/\sqrt{2}\tilde\rho\tilde\rho^*)(r^2+a^2,-\Delta,0,a),\\ m^a &= (1/\sqrt{2}\tilde\rho)(ia\sin\theta,0,1,i\csc\theta)\\ \overline{m}^a &= (1/\sqrt{2}\tilde\rho^*)(-ia\sin\theta,0,1,-i\csc\theta), \end{aligned} \tag{1}$$

where

$$\Delta = r^2-2Mr+a^2,\quad \rho^2=r^2+a^2\cos^2\theta$$

and

$$\tilde\rho = r+ia\cos\theta. \tag{2}$$

The Kerr solution has the line element

$$\begin{aligned} ds^2 &= \left(1-\frac{2Mr}{\tilde\rho\tilde\rho^*}\right)dt^2 - \frac{\tilde\rho\tilde\rho^*}{\Delta}dr^2 - \tilde\rho\tilde\rho^*\,d\theta^2\\ &\quad -\left((r^2+a^2)+\frac{2a^2Mr\sin^2\theta}{\tilde\rho\tilde\rho^*}\right)\sin^2\theta\,d\phi^2\\ &\quad +\frac{4aMr\sin^2\theta}{\tilde\rho\tilde\rho^*}\,dt\,d\phi. \end{aligned} \tag{3}$$

The differential operators $\mathscr{D}_n$, $\mathscr{D}_n^\dagger$, $\mathscr{L}_n$, and $\mathscr{L}_n^\dagger$ are defined as

$$\begin{aligned} \mathscr{D}_n &= \partial_r + iK/\Delta + 2n[(r-M)/\Delta],\\ \mathscr{D}_n^\dagger &= \partial_r - iK/\Delta + 2n[(r-M)/\Delta],\\ \mathscr{L}_n &= \partial_\theta + Q + n\cot\theta,\\ \mathscr{L}_n^\dagger &= \partial_\theta - Q + n\cot\theta, \end{aligned} \tag{4}$$

where

$$K=(r^2+a^2)\sigma+am \quad\text{and}\quad Q=a\sigma\sin\theta+m\csc\theta. \tag{5}$$

Teukolsky functions P_{+s} and P_{-s} in the variable r satisfy

$$\begin{aligned} (\Delta\mathscr{D}_{1-s}\mathscr{D}_0^\dagger - 2(2|s|-1)i\sigma r)P_{+s} &= \lambda P_{+s},\\ (\Delta\mathscr{D}_{1-s}^\dagger\mathscr{D}_0 + 2(2|s|-1)i\sigma r)P_{-s} &= \lambda P_{-s}, \end{aligned} \tag{6}$$

where $s=\frac{1}{2},1,\ldots$. The first result proven is the following theorem.

Theorem 1: If $s=\frac{1}{2},1,\ldots$ then

$$\begin{aligned} &\Delta^s\mathscr{D}_0^{2s}[\Delta\mathscr{D}_{1-s}^\dagger\mathscr{D}_0 + 2(2s-1)i\sigma r]\\ &\quad = [\Delta\mathscr{D}_{1-s}\mathscr{D}_0^\dagger - 2(2s-1)i\sigma r]\Delta^s\mathscr{D}_0^{2s}. \end{aligned} \tag{7}$$

Proof: By induction on s. Noting that for $s=\frac{1}{2}$

$$\Delta^{1/2}\mathscr{D}_0(\Delta\,\mathscr{D}_{1/2}^\dagger\mathscr{D}_0) = (\Delta\mathscr{D}_{1/2}\mathscr{D}_0^\dagger)\Delta^{1/2}\mathscr{D}_0. \tag{8}$$

If we now assume the result is true for a given s then

$$\begin{aligned} &\Delta^{s+1/2}\mathscr{D}_0^{2(s+1/2)}[\Delta\mathscr{D}_{1-(s+1/2)}^\dagger\mathscr{D}_0 + 2(2(s+\tfrac{1}{2})-1)i\sigma r]\\ &\quad = \Delta^{s+1/2}\mathscr{D}_0^{2s+1}[\Delta(\mathscr{D}_{1-s}^\dagger - (r-M)/\Delta)\mathscr{D}_0 + 4si\sigma r]\\ &\quad = \Delta^{1/2}\mathscr{D}_{-s}[\Delta^s\mathscr{D}_0^{2s}(\Delta\,\mathscr{D}_{1-s}^\dagger\mathscr{D}_0 + 2(2s-1)i\sigma r)\\ &\qquad + \Delta^s\mathscr{D}_0^{2s}(2i\sigma r - (r-M)\mathscr{D}_0)] \end{aligned} \tag{9}$$

and

$$[\Delta\mathscr{D}_{1-(s+1/2)}\mathscr{D}_0^\dagger - 2(2(s+\tfrac12)-1)i\sigma r]\Delta^{s+1/2}\mathscr{D}_0^{2(s+1/2)}$$
$$=\Delta^{1/2}(\Delta\mathscr{D}_{1-s}\mathscr{D}_{1/2}^\dagger - 4si\sigma r)\mathscr{D}_{-s}\Delta^s\mathscr{D}_0^{2s}$$
$$=\Delta^{1/2}\mathscr{D}_{-s}(\Delta\mathscr{D}_{1/2}^\dagger\mathscr{D}_{-s} - 4si\sigma r)\Delta^s\mathscr{D}_0^{2s} + 4si\sigma\Delta^{s+1/2}\mathscr{D}_0^{2s}$$
$$=\Delta^{1/2}\mathscr{D}_{-s}\left[\Delta\left(\mathscr{D}_{1-s} + (2s-1)\frac{r-M}{\Delta} - \frac{2iK}{\Delta}\right)\left(\mathscr{D}_0^\dagger - 2s\frac{r-M}{\Delta} + \frac{2iK}{\Delta}\right) - 4si\sigma r\right]\Delta^s\mathscr{D}_0^{2s} + 4si\sigma\Delta^{s+1/2}\mathscr{D}_0^{2s}$$
$$=\Delta^{1/2}\mathscr{D}_{-s}[(\Delta\mathscr{D}_{1-s}\mathscr{D}_0^\dagger - 2(2s-1)i\sigma r)\Delta^s\mathscr{D}_0^{2s} + (2i\sigma r - 2s - (r-M)\mathscr{D}_{-s})\Delta^s\mathscr{D}_0^{2s} + 4si\sigma\Delta^s\mathscr{D}_0^{2s-1}]. \quad (10)$$

Note that

$$\Delta^s\mathscr{D}_0^{2s}(2i\sigma r - (r-M)\mathscr{D}_0) = (2i\sigma r - (r-M)\mathscr{D}_{-s})\Delta^s\mathscr{D}_0^{2s} + \Delta^s(4si\sigma\mathscr{D}_0^{2s-1} - 2s\mathscr{D}_0^{2s})$$
$$= (2i\sigma r - 2s - (r-M)\mathscr{D}_{-s})\Delta^s\mathscr{D}_0^{2s} + 4si\sigma\Delta^s\mathscr{D}_0^{2s-1}. \quad (11)$$

Thus subtracting (10) from (9) and making use of (11) we have

$$\Delta^{s+1/2}\mathscr{D}_0^{2(s+1/2)}[\Delta\mathscr{D}_{1-(s+1/2)}^\dagger\mathscr{D}_0 + 2(2(s+\tfrac12)-1)i\sigma r]$$
$$-[\Delta\mathscr{D}_{1-(s+1/2)}\mathscr{D}_0^\dagger - 2(2(s+\tfrac12)-1)i\sigma r]\Delta^{s+1/2}\mathscr{D}_0^{2(s+1/2)}$$
$$=\Delta^{1/2}\mathscr{D}_{-s}[\Delta^s\mathscr{D}_0^{2s}(\Delta\mathscr{D}_{1-s}^\dagger\mathscr{D}_0 + 2(2s-1)i\sigma r) - (\Delta\mathscr{D}_{1-s}\mathscr{D}_0^\dagger - 2(2s-1)i\sigma r)\Delta^s\mathscr{D}_0^{2s}]. \quad (12)$$

A direct consequence of this result is that $\Delta\mathscr{D}_0^{2s}P_{-s}$ is a solution of the Teukolsky equation for P_{+s}. Similarly it may be proven that

$$\Delta^s\mathscr{D}_0^{\dagger 2s}[\Delta\mathscr{D}_{1-s}\mathscr{D}_0^\dagger - 2(2s-1)i\sigma r] = [\Delta\mathscr{D}_{1-s}^\dagger\mathscr{D}_0 + 2(2s-1)i\sigma r]\Delta^s\mathscr{D}_0^{\dagger 2s}, \quad (13)$$

i.e., $\Delta^s\mathscr{D}_0^{\dagger 2s}P_{+s}$ is a solution of the Teukolsky equation for P_{-s}. By suitable choice of the relative normalization of the functions we can write the following results:

$$\begin{aligned}\Delta^s\mathscr{D}_0^{2s}P_{-s} &= D_sP_{+s},\\ \Delta^s\mathscr{D}_0^{\dagger 2s}P_{+s} &= D_s^*P_{-s},\end{aligned} \quad (14)$$

where D_s is some complex constant. These are the Teukolsky–Starobinsky identities known to be true for $s=\frac12, 1, \frac32, 2$. For the variable θ we can prove a similar result.

Theorem 2: If $s=\frac12, 1,\ldots$ then

$$\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s[\mathscr{L}_{1-s}^\dagger\mathscr{L}_s + 2(2s-1)\sigma a\cos\theta]$$
$$=[\mathscr{L}_{1-s}\mathscr{L}_s^\dagger - 2(2s-1)\sigma a\cos\theta]\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s. \quad (15)$$

Proof: Again using induction we note that for $s=\frac12$

$$\mathscr{L}_{1/2}(\mathscr{L}_{1/2}^\dagger\mathscr{L}_{1/2}) = (\mathscr{L}_{1/2}\mathscr{L}_{1/2}^\dagger)\mathscr{L}_{1/2}. \quad (16)$$

Then

$$\mathscr{L}_{1-(s+1/2)}\mathscr{L}_{2-(s+1/2)}\cdots\mathscr{L}_{(s+1/2)-1}\mathscr{L}_{s+1/2}[\mathscr{L}_{1-s}^\dagger\mathscr{L}_s + 2(2(s+\tfrac12)-1)\sigma a\cos\theta]$$
$$=(1/\sqrt{\sin\theta})\mathscr{L}_{-s}\mathscr{L}_{1-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s(\mathscr{L}_s^\dagger\mathscr{L}_s + 4s\sigma a\cos\theta)\sqrt{\sin\theta}$$
$$=(1/\sqrt{\sin\theta})\mathscr{L}_{-s}\mathscr{L}_{1-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s[\mathscr{L}_{1-s}^\dagger\mathscr{L}_s + 2(2s-1)\sigma a\cos\theta - \cot\theta\mathscr{L}_s + 2\sigma a\cos\theta]\sqrt{\sin\theta} \quad (17)$$

and

$$[\mathscr{L}_{1-(s+1/2)}\mathscr{L}_{s+1/2}^\dagger - 2(2(s+\tfrac12)-1)\sigma a\cos\theta]\mathscr{L}_{1-(s+1/2)}\mathscr{L}_{2-(s+1/2)}\cdots\mathscr{L}_{(s+1/2)-1}\mathscr{L}_{(s+1/2)}$$
$$=(1/\sqrt{\sin\theta})(\mathscr{L}_{-s}\mathscr{L}_s^\dagger - 4s\sigma a\cos\theta)\mathscr{L}_{-s}\mathscr{L}_{1-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s\sqrt{\sin\theta}$$
$$=(1/\sqrt{\sin\theta})\mathscr{L}_{-s}(\mathscr{L}_s^\dagger\mathscr{L}_{-s} - 4s\sigma a\cos\theta)\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s\sqrt{\sin\theta}$$
$$-4s\sigma a\sqrt{\sin\theta}\,\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s\sqrt{\sin\theta}. \quad (18)$$

Now note that we can write

$$\mathscr{L}_s^\dagger\mathscr{L}_{-s} - 4s\sigma a\cos\theta = (\mathscr{L}_{1-s} + (2s-1)\cot\theta - 2Q)(\mathscr{L}_s^\dagger - 2s\cot\theta + 2Q) - 4s\sigma a\cos\theta$$
$$=\mathscr{L}_{1-s}\mathscr{L}_s^\dagger - 2(2s-1)\sigma a\cos\theta - \cot\theta\,\mathscr{L}_s^\dagger + 2s\csc^2\theta - 2m\cot\theta\csc\theta \quad (19)$$

observing the identities

$$\mathscr{L}_a\mathscr{L}_{a+1}\cdots\mathscr{L}_{b-1}\mathscr{L}_b\cos\theta = \cos\theta\mathscr{L}_a\mathscr{L}_{a+1}\cdots\mathscr{L}_{b-1}\mathscr{L}_b - (b-a+1)\sin\theta\mathscr{L}_{a+1}\mathscr{L}_{a+2}\cdots\mathscr{L}_{b-1}\mathscr{L}_b \quad (20)$$

and

$$\mathscr{L}_a\mathscr{L}_{a+1}\cdots\mathscr{L}_{b-1}\mathscr{L}_b\cot\theta = \cot\theta\mathscr{L}_{a-1}\mathscr{L}_a\cdots\mathscr{L}_{b-2}\mathscr{L}_{b-1} - (b-a+1)\mathscr{L}_a\mathscr{L}_{a+1}\cdots\mathscr{L}_{b-2}\mathscr{L}_{b-1} \quad (21)$$

and noting that

$$
\begin{aligned}
&(1/\sqrt{\sin\theta})\mathscr{L}_{-s}[\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s(-\cot\theta\mathscr{L}_s+2\sigma a\cos\theta)\\
&\quad-(-\cot\theta\,\mathscr{L}_s^\dagger+2s\csc^2\theta-2m\cot\theta\csc\theta)\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s]\sqrt{\sin\theta}\\
&\quad+4s\sigma a\sqrt{\sin\theta}\,\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s\sqrt{\sin\theta}\\
&\quad=(1/\sqrt{\sin\theta})\mathscr{L}_{-s}[-\cot\theta\mathscr{L}_{-s}\mathscr{L}_{1-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s+2s\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s\\
&\quad+2\sigma a\cos\theta\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s-4s\sigma a\sin\theta\mathscr{L}_{2-s}\mathscr{L}_{3-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s\\
&\quad-(-\cot\theta\mathscr{L}_s+2Q\cot\theta+2s\csc^2\theta-2m\cot\theta\csc\theta)\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s]\sqrt{\sin\theta}\\
&\quad+4s\sigma a\sqrt{\sin\theta}\,\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s\sqrt{\sin\theta}\\
&\quad=(1/\sqrt{\sin\theta})\mathscr{L}_{-s}[-\cot\theta\mathscr{L}_{-s}+2s+2\sigma a\cos\theta+\cot\theta\mathscr{L}_s-2Q\cot\theta-2s\csc^2\theta\\
&\quad+2m\cot\theta\csc\theta]\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s\sqrt{\sin\theta}\\
&\quad=0,
\end{aligned}
\tag{22}
$$

we have established that

$$
\begin{aligned}
&\mathscr{L}_{1-(s+1/2)}\mathscr{L}_{2-(s+1/2)}\cdots\mathscr{L}_{(s+1/2)-1}\mathscr{L}_{s+1/2}[\mathscr{L}_{1-s}^\dagger\,\mathscr{L}_s+2(2(s+\tfrac{1}{2})-1)\sigma a\cos\theta]\\
&\quad-[\mathscr{L}_{1-(s+1/2)}\,\mathscr{L}_{s+1/2}^\dagger-2(2(s+\tfrac{1}{2})-1)\sigma a\cos\theta]\mathscr{L}_{1-(s+1/2)}\mathscr{L}_{2-(s+1/2)}\cdots\mathscr{L}_{(s+1/2)-1}\mathscr{L}_{(s+1/2)}\\
&\quad=(1/\sqrt{\sin\theta})\mathscr{L}_{-s}[\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s(\mathscr{L}_{1-s}^\dagger\mathscr{L}_s+2(2s-1)\sigma a\cos\theta)\\
&\quad-(\mathscr{L}_{1-s}\,\mathscr{L}_s^\dagger-2(2s-1)\sigma a\cos\theta)\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_s]\sqrt{\sin\theta}
\end{aligned}
\tag{23}
$$

and the result is proven. In this case the Teukolsky equations are defined as

$$
\begin{aligned}
&(\mathscr{L}_{1-s}^\dagger\,\mathscr{L}_s+2(2s-1)\sigma a\cos\theta)S_{+s}=-\lambda S_{+s},\\
&(\mathscr{L}_{1-s}\,\mathscr{L}_s^\dagger-2(2s-1)\sigma a\cos\theta)S_{-s}=-\lambda S_{-s}.
\end{aligned}
\tag{24}
$$

Consequently Theorem 2 tells us that we can upon suitable renormalization, find a constant C_s such that

$$\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_{s-1}\mathscr{L}_sS_{+s}=C_sS_{-s}. \tag{25}$$

Similarly one may prove the identity

$$
\begin{aligned}
&\mathscr{L}_{1-s}^\dagger\,\mathscr{L}_{2-s}^\dagger\cdots\mathscr{L}_{s-1}^\dagger\,\mathscr{L}_s^\dagger\\
&\quad\times[\mathscr{L}_{1-s}\,\mathscr{L}_s^\dagger-2(2s-1)\sigma a\cos\theta]\\
&\quad=[\mathscr{L}_{1-s}^\dagger\,\mathscr{L}_s+2(2s-1)\sigma a\cos\theta]\\
&\quad\times\mathscr{L}_{1-s}^\dagger\,\mathscr{L}_{2-s}^\dagger\cdots\mathscr{L}_{s-1}^\dagger\,\mathscr{L}_s^\dagger
\end{aligned}
\tag{26}
$$

from which we can write

$$\mathscr{L}_{1-s}^\dagger\,\mathscr{L}_{2-s}^\dagger\cdots\mathscr{L}_{s-1}^\dagger\,\mathscr{L}_s^\dagger\,S_{-s}=C_sS_{+s}. \tag{27}$$

These are the Teukolsky–Starobinsky identities known to be true for $s=\frac{1}{2},1,\frac{3}{2},2$. The question we now ask is what if any significance do the Teukolsky functions have for general s. Before giving a covariant equation that works for general s let us recapitulate how things work in the case of the Rarita–Schwinger field. The Rarita–Schwinger equation written in spinor notation is

$$\nabla^{AA'}F_{AB}{}^{B'}=0, \tag{28}$$

where $F_{ABB'}=F_{(AB)B'}$. We can construct a coupled system of equations as follows. Let

$$h_{ABC}=\nabla_{(AA'}F_{BC)}{}^{A'}. \tag{29}$$

Then h_{ABC} satisfies a first-order equation as follows:

$$
\begin{aligned}
\nabla^{AA'}h_{ABC}&=\nabla^{AA'}\nabla_{(AB'}F_{BC)}{}^{B'}\\
&=\tfrac{1}{3}\nabla^{AA'}\nabla_{AB'}F_{BC}{}^{B'}+\tfrac{2}{3}\nabla^{AA'}\nabla_{(BB'}F_{C)A}{}^{B'}.
\end{aligned}
\tag{30}
$$

Using the Rarita–Schwinger equation and the symmetry in the indices B and C we write

$$
\begin{aligned}
\tfrac{1}{3}\nabla^{AA'}\nabla_{AB'}F_{BC}{}^{B'}&=\tfrac{1}{3}\nabla^{AA'}\nabla_{BB'}F_{AC}{}^{B'}+\tfrac{1}{3}\nabla_B{}^{A'}\nabla_{AB'}F^A{}_C{}^{B'}\\
&=\tfrac{1}{3}\nabla^{AA'}\nabla_{(BB'}F_{C)A}{}^{B'}.
\end{aligned}
\tag{31}
$$

Consequently (30) becomes

$$
\begin{aligned}
\nabla^{AA'}h_{ABC}&=\nabla^{AA'}\nabla_{(BB'}F_{C)A}{}^{B'}\\
&=\nabla_{(BB'}\nabla^{AA'}F_{C)A}{}^{B'}+[\nabla^{AA'},\nabla_{(BB'}]F_{C)A}{}^{B'}\\
&=[\nabla^{AA'},\nabla_{(BB'}]F_{C)A}{}^{B'}\\
&=-\epsilon^{A'}{}_{B'}\Psi^A{}_{BC}{}^MF_{MA}{}^{B'}\\
&=\Psi_{BC}{}^{AM}F_{AM}{}^{A'}.
\end{aligned}
\tag{32}
$$

The pair of equations (29), (32) when written in Newman–Penrose notation become

$$(D-\rho)h_{111}-(\delta^*+3\pi+\alpha)h_{110}=-\Psi_2F_{110'}, \tag{33}$$

$$(D-2\rho)h_{110}-(\delta^*+2\pi-\alpha)h_{100}=\Psi_2F_{100'}, \tag{34}$$

$$(D-3\rho)h_{100}-(\delta^*+\pi-3\alpha)h_{000}=-\Psi_2F_{000'}, \tag{35}$$

$$(\delta+3\beta-\tau)h_{111}-(\mathbf{\Delta}+\gamma+3\mu)h_{110}=-\Psi_2F_{111'}, \tag{36}$$

$$(\delta+\beta-2\tau)h_{110}-(\mathbf{\Delta}-\gamma+2\mu)h_{100}=\Psi_2F_{101'}, \tag{37}$$

$$(\delta-\beta-3\tau)h_{100}-(\mathbf{\Delta}-3\gamma+\mu)h_{000}=-\Psi_2F_{001'}, \tag{38}$$

and

$$(D-\rho^*)F_{00}{}^{0'}+(\delta-\alpha^*-2\beta+\pi^*)F_{00}{}^{1'}=h_{000}, \tag{39}$$

$$
\begin{aligned}
&(\delta^*+2\alpha+\beta^*-\tau^*)F_{11}{}^{0'}\\
&\quad+(\mathbf{\Delta}+2\gamma-\gamma^*+\mu^*)F_{11}{}^{1'}=h_{111},
\end{aligned}
\tag{40}
$$

$$
\begin{aligned}
&2[(D-\rho^*+\rho)F_{10}{}^{0'}+(\delta+\pi^*-\alpha^*+\tau)F_{10}{}^{1'}]\\
&\quad+[(\delta^*+\beta^*-2\alpha-\tau^*-2\pi)F_{00}{}^{0'}\\
&\quad+(\mathbf{\Delta}+\mu^*-\gamma^*-2\gamma-2\mu)F_{00}{}^{1'}]=3h_{100},
\end{aligned}
\tag{41}
$$

$$2[(\delta^* + \beta^* - \tau^* - \pi)F_{10}{}^{0'} + (\mathbf{\Delta} - \gamma^* + \mu^* - \mu)F_{10}{}^{1'}] + [(D - \rho^* + 2\rho)F_{11}{}^{0'} + (\delta - \alpha^* + 2\beta + \pi^* + 2\tau)F_{11}{}^{1'}] = 3h_{110}. \quad (42)$$

Considering Eqs. (35), (38), and (39) and putting

$$h_{100} = (1/\tilde{\rho}^*)H_1, \quad h_{000} = H_0,$$
$$F_{000} = (1/\sqrt{2}\tilde{\rho}^*)G_{000'}, \quad F_{001'} = (1/\sqrt{2}\rho^2)G_{001'}, \quad (43)$$

we obtain

$$\left(\mathscr{L}^\dagger_{-1/2} + \frac{2ia\sin\theta}{\tilde{\rho}^*}\right)H_1 + \Delta\left(\mathscr{D}^\dagger_{3/2} - \frac{2}{\tilde{\rho}^*}\right)H_0 = -\Psi_2 G_{001'},$$
$$\left(\mathscr{D}_0 + \frac{2}{\tilde{\rho}^*}\right)H_1 - \left(\mathscr{L}_{3/2} - \frac{2ia\sin\theta}{\tilde{\rho}^*}\right)H_0 = -\Psi_2 G_{000'},$$
$$\left(\mathscr{D}_0 - \frac{1}{\tilde{\rho}^*}\right)G_{001'} - \left(\mathscr{L}^\dagger_{-1/2} - \frac{ia\sin\theta}{\tilde{\rho}^*}\right)G_{000'} = 2\rho^2 H_0. \quad (44)$$

These equations imply that H_0 satisfies the separable equation

$$(\Delta\mathscr{D}_1\mathscr{D}^\dagger_{3/2} + \mathscr{L}^\dagger_{-1/2}\mathscr{L}_{3/2} - 4i\sigma\tilde{\rho})H_0 = 0 \quad (45)$$

admitting solutions $H_0 = \Delta^{-3/2}P_{+3/2}S_{+3/2}$. Similarly if Eqs. (33), (36), and (40) are considered then putting

$$h_{110} = (1/\tilde{\rho}^{*2})H_2, \quad h_{111} = (1/\tilde{\rho}^{*3})H_3,$$
$$F_{110'} = (1/\sqrt{2}\tilde{\rho}^{*3})G_{110'}, \quad F_{111'} = (1/\sqrt{2}\rho^2\tilde{\rho}^{*2})G_{111'}, \quad (46)$$

we obtain

$$\left(\mathscr{D}_0 - \frac{2}{\tilde{\rho}^*}\right)H_3 - \left(\mathscr{L}_{-1/2} + \frac{2ia\sin\theta}{\tilde{\rho}^*}\right)H_2 = -\Psi_2 G_{110'},$$
$$\left(\mathscr{L}^\dagger_{3/2} - \frac{2ia\sin\theta}{\tilde{\rho}^*}\right)H_3 + \Delta\left(\mathscr{D}^\dagger_{-1/2} + \frac{2}{\tilde{\rho}^*}\right)H_2 = -\Psi_2 G_{111'},$$
$$\left(\mathscr{L}_{-1/2} - \frac{ia\sin\theta}{\tilde{\rho}^*}\right)G_{111'} + \Delta\left(\mathscr{D}^\dagger_{-1/2} - \frac{1}{\tilde{\rho}^*}\right)G_{110'} = 2\rho^2 H_3. \quad (47)$$

The functions H_3 satisfies the separable equation

$$(\Delta\mathscr{D}^\dagger_{-1/2}\mathscr{D}_0 + \mathscr{L}_{-1/2}\mathscr{L}^\dagger_{3/2} + 4i\sigma\tilde{\rho})H_3 = 0 \quad (48)$$

admitting solutions $H_3 = P_{-3/2}S_{-3/2}$. Two solutions to Eqs. (44) and (47) can be found

(1) $h_{000} = \Delta^{-3/2}P_{+3/2}S_{+3/2}$,

$$F_{000'} = \frac{1}{\sqrt{2}\tilde{\rho}^*\Psi_2}\left(\mathscr{L}_{3/2} - \frac{2ia\sin\theta}{\tilde{\rho}^*}\right)\Delta^{-3/2}P_{+3/2}S_{+3/2},$$
$$F_{001'} = \frac{-1}{\sqrt{2}\rho^2\Psi_2}\Delta\left(\mathscr{D}^\dagger_{3/2} - \frac{2}{\tilde{\rho}^*}\right)\Delta^{-3/2}P_{+3/2}S_{+3/2},$$

and

(2) $h_{111} = \frac{1}{\tilde{\rho}^{*3}}P_{-3/2}S_{-3/2}$,

$$F_{110'} = \frac{-1}{\sqrt{2}\tilde{\rho}^{*3}\Psi_2}\left(\mathscr{D}_0 - \frac{2}{\tilde{\rho}^*}\right)P_{-3/2}S_{-3/2},$$
$$F_{111'} = \frac{-\Delta}{\sqrt{2}\rho^2\tilde{\rho}^{*2}\Psi_2}\left(\mathscr{L}^\dagger_{3/2} - \frac{2ia\sin\theta}{\tilde{\rho}^*}\right)P_{-3/2}S_{-3/2}. \quad (49)$$

In the case of each solution we have given only the nonzero components. It is interesting to note that the spinor $F_{ABA'}$ satisfies the equation

$$\nabla^A{}_{C'}\nabla_{(AA'}F_{BC)}{}^{A'} = \Psi^{AM}{}_{BC}F_{AMC'}. \quad (50)$$

However, the above choices do not satisfy the Rarita–Schwinger equation (28). It is possible to extend these equations to a set which has solutions in terms of Teukolsky functions for general s. If we consider the equations

$$\phi\nabla_{(A_1A'}F^{A'}{}_{A_2\cdots A_{2s})} - \tfrac{1}{6}(2s-3)(\nabla_{(A_1A'}\phi)F^{A'}{}_{A_2\cdots A_{2s})} = \phi h_{A_1\cdots A_{2s}}, \quad (51)$$
$$\nabla^{AA'}h_{AA_2\cdots A_{2s}} = (2s-1)(s-1)\Psi_{(A_2A_3}{}^{BC}F^{A'}{}_{A_4\cdots A_{2s})BC},$$

where $\phi = 2I = \Psi_{ABCD}\Psi^{ABCD}$ (Ref. 8) then these equations admit analogous solutions, viz.

(1) $h_{0\cdots 0} = \Delta^{-s}P_{+s}S_{+s}$,

$$F_{0\cdots 01'} = \frac{-1}{\sqrt{2}\rho^2(2s-1)(s-1)\Psi_2} \times\Delta\left(\mathscr{D}^\dagger_s - \frac{(2s-1)}{\tilde{\rho}^*}\right)\Delta^{-s}P_{+s}S_{+s},$$
$$F_{0\cdots 00'} = \frac{1}{\sqrt{2}\tilde{\rho}^*(2s-1)(s-1)\Psi_2} \times\left(\mathscr{L}_s - \frac{(2s-1)ia\sin\theta}{\tilde{\rho}^*}\right)\Delta^{-s}P_{+s}S_{+s},$$

and

(2) $h_{1\cdots 1} = \frac{1}{(\tilde{\rho}^*)^{2s}}P_{-s}S_{-s}$,

$$F_{1\cdots 10'} = \frac{-1}{\sqrt{2}(\tilde{\rho}^*)^{2s}(2s-1)(s-1)\Psi_2} \times\left(\mathscr{D}_0 - \frac{(2s-1)}{\tilde{\rho}^*}\right)P_{-s}S_{-s},$$
$$F_{1\cdots 11'} = \frac{-\Delta}{\sqrt{2}\tilde{\rho}(\tilde{\rho}^*)^{2s}(2s-1)(s-1)\Psi_2} \times\left(\mathscr{L}^\dagger_s - \frac{(2s-1)ia\sin\theta}{\tilde{\rho}^*}\right)P_{-s}S_{-s}, \quad (52)$$

where $P_{+s}S_{+s}$ are separable solutions of

$$(\Delta\mathscr{D}_1\mathscr{D}^\dagger_s + \mathscr{L}^\dagger_{1-s}\mathscr{L}_s - 2(2s-1)i\sigma\tilde{\rho})\Delta^{-s}P_{+s}S_{+s} = 0$$

and $P_{-s}S_{-s}$ separable solutions of

$$(\Delta\mathscr{D}^\dagger_1\mathscr{D}_s + \mathscr{L}_{1-s}\mathscr{L}^\dagger_s + 2(2s-1)i\sigma\tilde{\rho})\Delta^{-s}P_{-s}S_{-s} = 0.$$

Equation (51) is a generalization of the Rarita–Schwinger equation although it does not in itself have obvious physical significance for $s > \frac{3}{2}$. We also note that the method of proof for the Teukolsky–Starobinsky identities can be successfully used in the general context of Torres del Castillo.[5] Indeed we have the following result.

Theorem 3: If the operators $\mathscr{D}_n$ and $\mathscr{D}^\dagger_n$ are defined by

$$\mathscr{D}_n = \frac{\partial}{\partial r} + i\frac{q}{Q} + n\frac{Q^{(1)}}{Q} = Q^{-n}\mathscr{D}_0 Q^n$$

and

$$\mathscr{D}^\dagger_n = \frac{\partial}{\partial r} - i\frac{q}{Q} + n\frac{Q^{(1)}}{Q} = Q^{-n}\mathscr{D}^\dagger_0 Q^n, \quad (53)$$

with the functions q and Q polynomials such that $q^{(3)} = 0$ and $Q^{(5)} = 0$, then for all integer s

$$Q^s \mathscr{D}_0^{\dagger 2s}[Q\mathscr{D}_{1-s}\mathscr{D}_0^{\dagger} - (2s-1)iq^{(1)} + \tfrac{1}{6}(s-1)(2s-1)Q^{(2)}] = [Q\mathscr{D}_{1-s}^{\dagger}\mathscr{D}_0 + (2s-1)iq^{(1)} + \tfrac{1}{6}(s-1)(2s-1)Q^{(2)}]Q^s\mathscr{D}_0^{\dagger 2s}, \tag{54}$$

with a similar complex conjugate identity also holding. This theorem applies to all non-null orbit, type-D vacuum metrics given as for example by Torres del Castillo.[5]

ACKNOWLEDGEMENT

The work of W. M. was supported in part by the National Science Foundation under grant DMS 86-00372.

APPENDIX

Here we list the value of the Teukolsky–Starobinsky constant $|D_s|^2$ for a number of values of s, where $\tilde{\alpha}^2 = \alpha^2 + ma/\sigma$:

$$|D_{1/2}|^2 = \lambda,$$

$$|D_1|^2 = \lambda^2 - 4\sigma^2\tilde{\alpha}^2,$$

$$|D_{3/2}|^2 = \lambda^2(\lambda+1) - 16\sigma^2(\lambda\tilde{\alpha}^2 - a^2),$$

$$|D_2|^2 = \lambda^2(\lambda+2)^2 - 8\sigma^2\tilde{\alpha}^2\lambda(5\lambda+6) + 96\sigma^2a^2\lambda + 144\sigma^4\tilde{\alpha}^4 + 144\sigma^2M^2,$$

$$|D_{5/2}|^2 = \lambda^2(\lambda+3)^2(\lambda+4) - 16\sigma^2\tilde{\alpha}^2\lambda(\lambda+3)(5\lambda+8) + 48\sigma^2a^2\lambda(7\lambda+12) + 1024\sigma^4\tilde{\alpha}^4(\lambda+1)$$
$$- 3072\sigma^4\tilde{\alpha}^2a^2 + 1152\sigma^2M^2(\lambda+2),$$

$$|D_3|^2 = \lambda^2(\lambda+4)^2(\lambda+6)^2 - 4\sigma^2\tilde{\alpha}^2\lambda(\lambda+4)(35\lambda^2+252\lambda+360) + 128\sigma^2a^2\lambda(\lambda+4)(7\lambda+15)$$
$$+ 16\sigma^4\tilde{\alpha}^4(259\lambda^2+1140\lambda+900) - 2560\sigma^4\tilde{\alpha}^2a^2(11\lambda+15) + 25600\sigma^4a^4 - 14400\sigma^6\tilde{\alpha}^6$$
$$+ 576\sigma^2M^2((3\lambda+10)^2 - 100\sigma^2\tilde{\alpha}^2),$$

$$|D_{7/2}|^2 = \lambda^2(\lambda+5)^2(\lambda+8)^2(\lambda+9) - 32\sigma^2\tilde{\alpha}^2\lambda(\lambda+5)(\lambda+8)(7\lambda^2+63\lambda+108)$$
$$+ 288\sigma^2a^2\lambda(\lambda+5)(7\lambda^2+65\lambda+120) + 256\sigma^4\tilde{\alpha}^4(49\lambda^3+549\lambda^2+1728\lambda+1296)$$
$$- 4608\sigma^4\tilde{\alpha}^2a^2(31\lambda^2+175\lambda+180) + 57600\sigma^4a^4(5\lambda+9) - 147456\sigma^6\tilde{\alpha}^6(\lambda+3)$$
$$+ 884736\sigma^6\tilde{\alpha}^4a^2 - 92160\sigma^4M^2(7\tilde{\alpha}^2\lambda + 30\tilde{\alpha}^2 - 15a^2) + 5760\sigma^2M^2(3\lambda^3+45\lambda^2+220\lambda+360),$$

$$|D_4|^2 = \lambda^2(\lambda+6)^2(\lambda+10)^2(\lambda+12)^2 - 48\sigma^2\tilde{\alpha}^2\lambda(\lambda+6)(\lambda+10)(7\lambda^3+154\lambda^2+996\lambda+1680)$$
$$+ 576\sigma^2a^2\lambda(\lambda+6)(\lambda+10)(7\lambda^2+78\lambda+168) + 96\sigma^4\tilde{\alpha}^4(329\lambda^4+7372\lambda^3+55484\lambda^2+156240\lambda+117600)$$
$$- 4608\sigma^4\tilde{\alpha}^2a^2(115\lambda^3+1592\lambda^2+6216\lambda+5880) + 9216\sigma^4a^4(191\lambda^2+1344\lambda+1764)$$
$$- 256\sigma^6\tilde{\alpha}^6(3229\lambda^2+31010\lambda+63700) + 64512\sigma^6\tilde{\alpha}^4a^4(169\lambda+630) - 25288704\sigma^6\tilde{\alpha}^2a^4$$
$$+ 28224000\sigma^8\tilde{\alpha}^8 + 28224000\sigma^6M^2\tilde{\alpha}^4 + 25401600\sigma^4M^4$$
$$- 11520\sigma^4M^2(341\tilde{\alpha}^2\lambda^2 + 4242\tilde{\alpha}^2\lambda + 12740\tilde{\alpha}^2 - 1596a^2\lambda - 8232a^2)$$
$$+ 630\sigma^2M^2(75\lambda^4+2112\lambda^3+21568\lambda^2+96000\lambda+161280).$$

[1] S. A. Teukolsky, Phys. Rev. Lett. **29**, 1114 (1972).
[2] E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).
[3] R. Güven, Proc. R. Soc. London Ser. A **356**, 465 (1977).
[4] R. Güven, Phys. Rev. D **10**, 2327 (1980).
[5] G. F. Torres del Castillo, J. Math. Phys. **29**, 2078 (1988).
[6] R. Penrose and W. Rindler, *Spinors and Spacetime Vol. 1 Two Spinor Calculus and Relativistic Fields* (Cambridge U.P., Cambridge, 1984).
[7] S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford U.P., London, 1984).
[8] D. Kramer, H. Stephani, E. Herlt, and M. MacCallum, *Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980).

Complex manifold methods in quantum field theory in curved space-time

A. Meister
Institut für Theoretische Physik, Sidlerstrasse 5, 3012 Bern, Switzerland

(Received 17 March 1989; accepted for publication 12 July 1989)

A real analytic space-time $\mathscr{N}$ can be embedded as a real slice in a complex Riemannian manifold (M,g). A theorem due to Woodhouse [Int. J. Theor. Phys. **16**, 671 (1977)], stating that real slices are necessarily totally geodesic submanifolds of (M,g), is used to find all real slices of the complexified sphere ${}^cS^n_r$ of dimension n and radius r, and of complexified Robertson–Walker space-time ${}^c\mathscr{R}$. The real slices $\mathscr{N}$ of ${}^cS^n_r$ are n-dimensional spaces of constant curvature $(\pm r^{-2})$ of all possible signatures. ${}^c\mathscr{R}$ is determined by a holomorphic radius function $r(x)$, the x-constant hypersurfaces being isometric to ${}^cS^n_{r(x)}$. In general, all real slices of ${}^c\mathscr{R}$ are Robertson–Walker-type spaces (of various signatures). The interesting case that a Robertson–Walker space-time with radius function $\mathscr{r}(t)$ intersects a Euclidean real slice in an n-dimensional hypersurface, can only occur if $\mathscr{r}(t)$ is time reflection symmetric. Finally, propagators to the scalar wave operator $(\Box - m^2)$ on real slices $\mathscr{N}_1$ of ${}^cS^4_r$ are considered and their analytic continuations to other real slices $\mathscr{N}_2$ are discussed. There are useful continuations only if $\mathscr{N}_1$, $\mathscr{N}_2$ are isometric either to the sphere and de Sitter space-time, or to the hyperbolic plane and anti-de Sitter space-time. In this case the continuation does not depend on the relative position of $\mathscr{N}_1$ and $\mathscr{N}_2$ in ${}^cS^4_r$. For instance, it also works if the intersection $\mathscr{N}_1 \cap \mathscr{N}_2$ is one dimensional.

I. INTRODUCTION AND RESULTS

The subject of this work is the generalization to curved space-time of the Wick rotation $t \mapsto -i\tau$ known in flat space-time. An application of this procedure to a curved space-time was used by Dowker and Critchley[1] to derive the Feynman propagator of the massive scalar field in de Sitter space-time for the de Sitter invariant vacuum state: The Wick rotation is done in the time variable of the embedding Lorentzian $\mathbb{R}^5$. Under this Wick rotation de Sitter space-time is analytically continued to the sphere S^4 embedded in Euclidean $\mathbb{R}^5$. The analytic continuation of the Dirichlet propagator to the operator $\Delta - m^2$ on S^4 back to de Sitter space-time yields the Feynman propagator mentioned above. Analytic continuation in a parameter related to t was used by Candelas and Raine[2] to derive the Feynman propagator in static space-times. A systematic treatment of the Wick rotation for the case of globally static space-times was given by Wald[3] and Fulling and Ruijsenaars[4]: It is shown in Wald[3] that, if the spacelike hypersurfaces orthogonal to the timelike Killing vector field are labelled by the Killing parameter t, the Wick rotation $t \mapsto -i\tau$ yields a Riemannian (i.e., positive definite) metric and the wave operator gets elliptic. The Dirichlet propagator to this operator can be analytically continued back to the Lorentzian space-time $\tau \mapsto it$ and yields the Feynman propagator one also obtains by using directly the full translation symmetry in the static Lorentzian space-time. In Fulling and Ruijsenaars[4] the treatment is generalized to include thermal states on the static space-time, and the relations are studied between these thermal states, periodicity of the two-point functions in the Euclidean time $\tau = -it$, and the existence of Killing horizons in the static space-time. The mathematical tools to generalize this procedure to arbitrary curved space-times in a geometric, coordinate-free way were introduced by Woodhouse[5]: A real analytic Lorentzian space-time L is embedded in a four-dimensional complex Riemannian manifold (a complex manifold with a nondegenerate, holomorphic tensor field) as a real slice. Such a complex Riemannian manifold can be obtained by starting from a real analytic atlas of L and allowing the coordinates to assume complex values. Since the coordinate transformations are real analytic functions, they can be continued slightly into the complex and can be used to "piece" the atlas together to form a complex manifold. The metric coefficients which are analytic functions of the coordinates can also be continued to the complex and can be interpreted as the coefficients of a second rank, covariant, holomorphic tensor field, a complex Riemannian metric. In Woodhouse[5] a reference is given for more detail on this point. The question is now whether there are other real slices in this complex Riemannian manifold, in particular whether there are Euclidean real slices, restricted to which the complex Riemannian metric is positive (or negative) definite. One important result of Woodhouse is that real slices are necessarily totally geodesic submanifolds of the complex Riemannian manifold. In particular, this implies that the intersection of two real slices is a totally geodesic submanifold of both real slices. So, for instance, if a Lorentzian space-time L has no totally geodesic three-dimensional submanifold, there are no Euclidean real slices that intersect L in a three-dimensional submanifold. This indicates that the Euclideanization procedure might fail in many cases, just because there are no Euclidean real slices.

This work is meant to shed some light on the following two questions which quite naturally arise here. First, one should get some better understanding under which conditions, or "how often" real slices do exist in a given complex Riemannian manifold. Second, provided there are some real slices, what is the use of possible analytic continuations from

 0022-2488/89/122930-13$02.50

one real slice to another. These two points are studied here by means of some specific examples for which the questions can be posed and answered explicitly. Point one is treated in Sec. III and point two in Sec. IV.

As to the existence of real slices the first example considered is the complexified sphere ${}^c S^n_r$, all real slices of which are found. ${}^c S^n_r$ is the submanifold of $\mathbb{C}^{n+1}$ described by the equation $\Sigma^n_{A=0} (\xi^A)^2 = r^2$. The result is that for $r^2 \notin \mathbb{R}$ there are no real slices at all. For $r^2 \in \mathbb{R}$, however, real slices exist and they are spaces of constant curvature of all possible signatures, the sphere S^n, signature $(+ \ + \cdots +)$, de Sitter space-time, $(- \ + \ + \cdots +)$, anti-de Sitter space-time, $(- \ - \cdots - \ +)$, the hyperbolic plane H^n, $(- \ - \cdots -)$ (i.e., with negative definite metric), and other spaces of constant curvature for the remaining signatures. All real slices of the same signature are isometric to each other. Finally, for the case $n = 4$, all possible intersections of two real slices are described.

The second class of examples are complexified Robertson–Walker space-times (RWS) of dimension higher than two. These manifolds are determined by a holomorphic radius function r of a complex "proper time" x. The x-constant hypersurfaces are complexified spheres isometric to ${}^c S^n_{r(x)}$. Again for a "generic" radius function $r(x)$ there are no real slices at all. Real slices exist if and only if for some $x = x_0$, $r^2(x_0 + t)$ or $r^2(x_0 + it)$ is real for all $t \in \mathbb{R}$. All real slices are then real RWS with proper time t and radius function $r(x_0 + t)$, $r(x_0 + it)$, respectively, differing just in the t-constant hypersurfaces, which can be any of the spaces of constant curvature contained as real slices in ${}^c S^n_r$. "Interesting" interesections of real slices can only occur if there is some $x = x_0$ for which both $r^2(x_0 + t)$ and $r^2(x_0 + it)$ are real for all real t. This then implies that the radius function $r(x_0 + t)$ of the real slices is symmetric under time-reflection: $r(x_0 + t) = r(x_0 - t)$ $\forall t$. This provides an example which shows that in a four-dimensional space-time L the existence of a three-dimensional totally geodesic submanifold N does not imply the existence of a Euclidean real slice which intersects L in N: In a real RWS with radius function $r(t)$ the t-constant surfaces at $t = t_0$ with $r'(t_0) = 0$ are totally geodesic submanifolds, but $r'(t_0) = 0$ does not imply the reflection symmetry of r about t_0 necessary for a Euclidean real slice to pass through the $t = t_0$ hypersurface.

To discuss the use of analytic continuation from one real slice to another, Green's functions to the wave-operator $\square - m^2$ of a massive scalar field are considered on the real slices of the four-dimensional complexified sphere ${}^c S^4_r$. To start with, all possible analytic continuations of the Dirichlet propagator on a real slice $\mathscr{N}_1$, isometric to the sphere S^4, are described. One of these continuations, when restricted to a real slice $\mathscr{N}_2$, isometric to de Sitter space-time, yields the Feynman propagator of the de Sitter invariant vacuum state. None of these continuations restricted to a real slice isometric to anti-de Sitter space-time (adS) yields a physically useful propagator; there is however a linear combination of two such continuations which yields the Feynman propagator of an adS-invariant vacuum state discussed in Avis, Isham, and Storey.[6] (This is not astonishing since all these propagators are solutions to an ordinary second-order differential equation in the squared geodesic distance and thus they are essentially the linear combination of two fundamental solutions.) In the same way as the Feynman propagator on de Sitter is obtained from the Dirichlet propagator on the sphere, this adS-invariant Feynman propagator can be obtained from the Dirichlet propagator on a real slice isometric to H^4.

These results seem to indicate that the continuation of propagators gives only useful results if the signatures of the two real slices differ only in one sign (as is the case for H^4 and adS or for S^4 and de Sitter). The result of the continuation from $\mathscr{N}_1$ to $\mathscr{N}_2$ discussed above does not depend on the specific choice of $\mathscr{N}_1$ and $\mathscr{N}_2$, the only condition on them is that they both contain a common point p (the two-point function to be continued is considered as a function of one variable by keeping one argument fixed, equal to p) and one of them is isometric to S^4, the other to de Sitter space-time. There are many such $\mathscr{N}_1$ and $\mathscr{N}_2$, in particular there are such $\mathscr{N}_1$, $\mathscr{N}_2$ that $\mathscr{N}_1 \cap \mathscr{N}_2$ is three-dimensional, and others such that $\mathscr{N}_1 \cap \mathscr{N}_2$ is one-dimensional, in which case the analytic continuation cannot be interpreted as a Wick rotation since the continuation is necessarily done in more than one coordinate. This might suggest that the dimension of the intersection of the two real slices is not essential, provided their signatures differ only in one sign. It might, however, hold only in this special case, because of the high symmetry of ${}^c S^4_r$ which also implies, for instance, that all real slices of the same signature are isometric to each other.

The plan of the paper is as follows: In Sec. II, in order to fix the notation and to keep the paper self-contained, the main definitions on complex manifolds and the main results of Woodhouse[5] are briefly reviewed and some corollaries are added which will facilitate the work in the subsequent sections. This section relies heavily on Chap. IX in Kobayashi and Nomizu[7] on complex manifolds. Then the possible tangent spaces to real slices and their mutual intersections are discussed. These tangent spaces can alternatively be interpreted as the real slices of complexified Minkowski space-time (i.e., flat $\mathbb{C}^n$) passing through the origin of $\mathbb{C}^n$. Most of the proofs in this section are kept short or are omitted. They are given in detail in Ref. 8.

In Sec. III the real slices of the complexified sphere and the complexified Robertson–Walker space-times are found and in Sec. IV the propagators on the real slices of ${}^c S^4_r$ are treated.

II. COMPLEX MANIFOLDS

We will first very briefly review the main definitions and statements about complex manifolds. This is found in more detail in Kobayashi and Nomizu,[7] Chap. IX. The definition of a complex manifold M is essentially the same as the definition of a real differentiable manifold, just that local coordinates (U, φ) are homeomorphisms of an open subset of M into $\mathbb{C}^n$ instead of $\mathbb{R}^n$ and coordinate transformations $\varphi \circ \varphi'^{-1}$ are required to be holomorphic (i.e., complex differentiable) functions.

It is convenient to consider a complex manifold as a real C^∞ manifold of dimension $2n$. This is done by identifying $\mathbb{C}^n$ with $\mathbb{R}^{2n}$ as follows:

$$\begin{array}{ccc} \mathbb{C}^n & \leftrightarrow & \mathbb{R}^{2n}, \\ (z^1,z^2,...,z^n) & \leftrightarrow & (x^1,x^2,...,x^{2n}), \quad z^k = x^k + ix^{k+n}. \end{array} \tag{2.1}$$

(Here and in the following, indices $i,j,k,...$ will run from 1 to n, indices $a,b,c,...$ will run from 1 to $2n$, and Einstein summation convention will be used.)

By means of the correspondence (2.1) the differentiable structure of a complex manifold M of dimension n is interpreted as an atlas of a $2n$-dimensional real C^∞ manifold also denoted by M. The coordinate systems of this atlas are called holomorphic coordinates. These holomorphic coordinates define a $\binom{1}{1}$ tensor field, the complex structure J:

$$J|_U = \frac{\partial}{\partial x^{i+n}} \otimes dx^i - \frac{\partial}{\partial x^i} \otimes dx^{i+n} \tag{2.2}$$

[where $(U,(x^1,x^2,...,x^{2n}))$ is any holomorphic coordinate system]. J will usually be interpreted as a mapping of the tangent bundle $T(M)$ onto itself or alternatively of the cotangent bundle $T^*(M)$ onto itself. Equation (2.2) shows that in either case it satisfies

$$J \circ J = -\mathbf{1}, \tag{2.3}$$

where $\mathbf{1}$ is the identity mapping on the tangent bundle (cotangent bundle).

The complexified tangent space $T_p^c(M)$ at a point $p \in M$ is the complexification of the tangent space $T_p(M)$. J is extended to act on $T_p^c(M)$ by complex linearity. $T_p^c(M)$ may be decomposed into the eigenspaces $T_p^{1,0}(M)$ and $T_p^{0,1}(M)$ to the eigenvalues $+i$ and $-i$, respectively,

$$T_p^c(M) = T_p^{0,1}(M) \oplus T_p^{0,1}(M) \tag{2.4}$$

($\oplus$ denotes the direct sum). The dual space $T_p^{c*}(M)$ is decomposed in the same way:

$$T_p^{c*}(M) = T_{p1,0}(M) \oplus T_{p0,1}(M), \tag{2.5}$$

where $T_{p1,0}(M)$ and $T_{p0,1}(M)$ are the eigenspaces of J to the eigenvalues $+i$ and $-i$, respectively. The decomposition (2.4) and (2.5) then extends to the full tensor algebra $\mathfrak{T}_p(M)$ generated by $T_p^c(M)$. The two spaces $T_p^{1,0}(M)$ and $T_{p1,0}(M)$ generate a subalgebra $S_p(M)$ of $\mathfrak{T}_p(M)$: Let $S^r_{ps}(M)$ denote the space:

$$S^r_{ps}(M) = \underbrace{T_p^{1,0}(M) \otimes T_p^{1,0}(M) \otimes \cdots \otimes T_p^{1,0}(M)}_{r \text{ times}} \otimes \underbrace{T_{p1,0}(M) \otimes T_{p1,0}(M) \otimes \cdots \otimes T_{p1,0}(M)}_{s \text{ times}} \tag{2.6}$$

and define $S_p(M) := \Sigma_{r,s} S^r_{ps}(M)$ (direct sum). Let $T^c(M)$, $T^{1,0}(M)$, $S_2^0(M)$, etc., denote the corresponding vector bundles and $\Gamma(B)$ denote the set of C^∞ cross sections of the bundle B (tensor fields). By complex linearity one can extend such notions, defined on the real bundles and their cross sections, as contractions, Lie differentiation, exterior differentiation, covariant differentiation, etc., to the complexified bundles and their cross sections.

In holomorphic local coordinates (U,φ) one introduces basic vector fields $(\partial/\partial z^k) \in \Gamma(T^{1,0}(U))$, $(\partial/\partial \bar z^k) \in \Gamma(T^{0,1}(U))$ and covector fields $dz^k \in \Gamma(T_{1,0}(U))$, $d\bar z^k \in \Gamma(T_{0,1}(U))$ adapted to this decomposition of the tensor algebra as follows:

$$\frac{\partial}{\partial z^k} := \frac{1}{2}\left(\frac{\partial}{\partial x^k} - i\frac{\partial}{\partial x^{k+n}}\right), \quad \frac{\partial}{\partial \bar z^k} := \frac{1}{2}\left(\frac{\partial}{\partial x^k} + i\frac{\partial}{\partial x^{k+n}}\right),$$
$$dz^k := dx^k + i\,dx^{k+n}, \quad d\bar z^k := dx^k - i\,dx^{k+n}. \tag{2.7}$$

Between these fields the following duality relations hold:

$$dz^k\left(\frac{\partial}{\partial z^j}\right) = d\bar z^k\left(\frac{\partial}{\partial \bar z^j}\right) = \delta^k_j$$

and

$$dz^k\left(\frac{\partial}{\partial \bar z^j}\right) = d\bar z^k\left(\frac{\partial}{\partial z^j}\right) = 0. \tag{2.8}$$

A tensor field $v \in \Gamma(S^r_s(M))$ is called holomorphic if its components with respect to the basic fields (2.7) are holomorphic functions. This definition makes sense due to the Cauchy–Riemann equations of the holomorphic coordinate transformations. Now we can define a complex Reimannian manifold.

Definition 2.1: A complex Riemannian manifold (M,g) is a complex manifold M endowed with a holomorphic covariant second-rank tensor field $g \in \Gamma(S_2^0(M))$ which is symmetric and nowhere degenerate, that is in every point $p \in M$, $g(Z_1,Z_2) = 0\ \forall Z_2 \in T_p^c(M)$ implies $Z_1 = 0$. (Woodhouse[5]).

Note that in contrast to the real case, g cannot be classified by a signature, since all nondegenerate, symmetric bilinear forms on a complex vectorspace are equivalent.

The metric g, regarded as a cross section of $T_2^{c0}(M)$, can be split into its real and imaginary part:

$$g = h + ik, \text{ where } h,k \in \Gamma(T_2^0(M)). \tag{2.9}$$

This splitting is given by $h(X,Y) = \mathrm{Re}(g(X,Y))$, $k(X,Y) = \mathrm{Im}(g(X,Y))$ for X,Y real vectors, i.e., $X,Y \in T(M)$. h,k are pseudo-Riemannian metrics [i.e., symmetric, nondegenerate $\binom{0}{2}$ tensor fields] of signature (n,n) and they satisfy (Woodhouse[5]):

$$\begin{aligned} h(X,Y) &= k(JX,Y), \\ k(X,Y) &= -h(JX,Y) \\ \rightarrow h(X,Y) &= -h(JX,JY), \\ k(X,Y) &= -k(JX,JY), \quad \forall X,Y \in T(M). \end{aligned} \tag{2.10}$$

The two pseudo-Riemannian metrics h and k define two metric connections ${}^h\nabla$ and ${}^k\nabla$. In Woodhouse[5] the following proposition is proven.

Proposition 2.2: The metric connections ${}^h\nabla$ of h and ${}^k\nabla$ of k coincide and the connection $\nabla :\equiv {}^h\nabla \equiv {}^k\nabla$ satisfies $\nabla g \equiv 0$ and $\nabla J \equiv 0$.

Next we define real slices.

Definition 2.3: A real slice $\mathcal{N}$ of an n-dimensional complex Riemannian manifold $(M,g = h + ik)$ is an n-dimensional real submanifold of M such that $h|_{\mathcal{N}}$ is nondegenerate and $k|_{\mathcal{N}} \equiv 0$. (Woodhouse[5]).

The second important result proven in Woodhouse[5] reads as follows.

Proposition 2.4: A real slice $\mathcal{N}$ of $(M,g = h + ik)$ is necessarily totally geodesic, i.e., a geodesic that is tangent to $\mathcal{N}$ in one point lies entirely in $\mathcal{N}$.

Since the intersection of two totally geodesic submanifolds is a totally geodesic submanifold it follows that the intersection of two real slices is a totally geodesic submanifold of M (and of both real slices) (Woodhouse[5]).

We will use the geodesic equation expressed in terms of holomorphic local coordinates.

Proposition 2.5: Let $(U,\varphi = (z^1,z^2,...,z^n))$ be holomorphic local coordinates. Define

$$\Gamma^i_{jk} := \tfrac{1}{2} g^{il}(g_{jl,k} + g_{lk,j} - g_{jk,l}),$$

where

$$g_{kl} := g\left(\frac{\partial}{\partial z^k}, \frac{\partial}{\partial z^l}\right)$$

and

$$g^{kl} g_{lj} = \delta^k_j$$

and

$$g_{jl,k} := \frac{\partial}{\partial z^k}(g_{jl}).$$

Let γ: $[0,1] \mapsto U$ be a differentiable curve represented by $z^k(\gamma(t)) =: c^k(t)$. Then the tangent $\dot\gamma(t)$ to γ is given by

$$\dot\gamma = \dot c^k \frac{\partial}{\partial z^k}\Big|_\gamma + \dot{\bar c}^k \frac{\partial}{\partial \bar z^k}\Big|_\gamma$$

and γ is an affinely parametrized geodesic if and only if the functions $c^k(t)$ satisfy:

$$\ddot c^i + \Gamma^i_{jk}(\gamma) \dot c^j \dot c^k = 0. \tag{2.11}$$

Proof: It follows from $\nabla J \equiv 0$ in Proposition 2.2 that ∇ is type preserving, so

$$\nabla_{\partial/\partial z^k}\left(\frac{\partial}{\partial z^j}\right)$$

is of type (1,0) and hence can be written as

$$\nabla_{\partial/\partial z^k}\left(\frac{\partial}{\partial z^j}\right) =: \hat\Gamma^i_{kj} \frac{\partial}{\partial z^i}.$$

Similarly

$$\nabla_{\partial/\partial \bar z^k}\left(\frac{\partial}{\partial z^j}\right) = \nabla_{\partial/\partial z^j}\left(\frac{\partial}{\partial \bar z^k}\right) + \left[\frac{\partial}{\partial \bar z^k}, \frac{\partial}{\partial z^j}\right] = \nabla_{\partial/\partial z^j}\left(\frac{\partial}{\partial \bar z^k}\right)$$

must be both of type (1,0) (left-hand side) and of type (0,1) (right-hand side) and hence must vanish. Finally, purely algebraically

$$\nabla_{\partial/\partial \bar z^k}\left(\frac{\partial}{\partial \bar z^j}\right) = \overline{\nabla_{\partial/\partial z^k}\left(\frac{\partial}{\partial z^j}\right)}.$$

Write $c^i(t) =: a^i(t) + i a^{i+n}(t)$, where $a^b(t)$ are real. Then really

$$\dot c^k \frac{\partial}{\partial z^k}\Big|_\gamma + \dot{\bar c}^k \frac{\partial}{\partial \bar z^k}\Big|_\gamma = a^b \frac{\partial}{\partial x^b}\Big|_\gamma = \dot\gamma.$$

Using this and $\nabla_{\dot\gamma}(\dot c^k) = \ddot c^k$ and the above relations one gets the equation

$$\nabla_{\dot\gamma}(\dot\gamma) = \ddot c^k \frac{\partial}{\partial z^k} + \dot c^i \dot c^j \hat\Gamma^k_{ji} \frac{\partial}{\partial z^k} + \ddot{\bar c}^k \frac{\partial}{\partial \bar z^k} + \dot{\bar c}^i \dot{\bar c}^j \overline{\hat\Gamma^k_{ji}} \frac{\partial}{\partial \bar z^k},$$

which yields the equivalence of the geodesic equation $\nabla_{\dot\gamma}(\dot\gamma) = 0$ and (2.11). The expression for $\hat\Gamma^i_{jk}$ is obtained as in the real case:

$$g_{ij,k} = \nabla_{\partial/\partial z^k} g\left(\frac{\partial}{\partial z^i}, \frac{\partial}{\partial z^j}\right) = (\nabla g \equiv 0)$$

$$= g\left(\hat\Gamma^l_{ki} \frac{\partial}{\partial z^l}, \frac{\partial}{\partial z^j}\right) + g\left(\frac{\partial}{\partial z^i}, \hat\Gamma^l_{kj} \frac{\partial}{\partial z^l}\right).$$

Evaluating now $\Gamma^i_{jk} := \tfrac{1}{2} g^{il}(g_{jl,k} + g_{lk,j} - g_{jk,l})$ one gets $\Gamma^i_{jk} = \hat\Gamma^i_{jk}$. ∎

Definition 2.6: Choose some $p \in M$ and let $(e_1,e_2,...,e_n)$ be an orthornormal basis of $T^{1,0}_p(M)$, i.e., $g(e_i,e_k) = \delta_{ik}$. Define complex normal coordinates (W,χ) at p by

$$\chi\colon W \mapsto V \subset \mathbb{C}, \qquad \chi^{-1}\colon \quad V \mapsto W$$
$$p \mapsto (\eta^1,\eta^2,...,\eta^n) \quad (\eta^1,\eta^2,...,\eta^n) \mapsto \exp(\eta^i e_i + \bar\eta^i \bar e_i).$$

Proposition 2.7: Complex normal coordinates (W,χ) at p are holomorphic coordinates. They are uniquely characterized by their properties [define

$$g_{ik} = g\left(\frac{\partial}{\partial \eta^i}, \frac{\partial}{\partial \eta^k}\right)$$

and Γ^i_{jk} as in Proposition 2.5]: (W,χ) are holomorphic local coordinates, $(\partial/\partial\eta^i)|_p = e_i$ and

$$\Gamma^i_{jk}(\chi^{-1}(\eta^1,\eta^2,...,\eta^n))\eta^j\eta^k = 0, \quad \forall (\eta^1,\eta^2,...,\eta^n) \in V.$$

Sketch of the proof: We will only prove that complex normal coordinates are holomorphic. From the theory on the exponential mapping (e.g., Helgason,[9] Chap. 1) it is known that there is a neighborhood V' of 0 in $T_p(M)$ such that

exp: $V' \mapsto M$

$v \mapsto \exp(v) = \gamma(1)$, where γ is the

affinely parametrized geodesic with $\dot\gamma(0) = v$,

is a diffeomorphism onto a neighborhood $\exp(V')$ of p. Furthermore for $v \in V'$, $\gamma(t) = \exp(vt)$ is the affinely parametrized geodesic with $\dot\gamma(0) = v$. So let (U,φ) be a holomorphic chart covering p such that $(\partial/\partial z^i)|_p = e_i$ and define Γ^i_{jk} with respect to this chart as in Proposition 2.5. Define the functions

$$c^k(\xi^1,\xi^2,...,\xi^{2n};t) := z^k \circ \exp\left(t\xi^a \frac{\partial}{\partial x^a}\Big|_p\right).$$

From the properties of exp mentioned above and from Proposition 2.5 it follows that:

$$\begin{aligned} &\ddot c^i + \Gamma^i_{jk} \dot c^j \dot c^k = 0, \quad \dot c^k|_{t=0} = \xi^k + i\xi^{k+n}, \\ &c^k|_{t=0} = z^k(p). \end{aligned} \tag{2.12}$$

Now take the partial derivative of (2.12) with respect to ξ^l and add to it i times the derivative of (2.12) with respect to ξ^{l+n} to get (using also that Γ^i_{jk} given in Proposition 2.5 are holomorphic functions)

$$\begin{aligned} &\frac{d^2}{dt^2}\left(\frac{\partial c^i}{\partial \xi^l} + i\frac{\partial c^i}{\partial \xi^{l+n}}\right) \\ &\quad + \frac{\partial(\Gamma^i_{jk} \circ \tilde\varphi^{-1})}{\partial z^m}\Bigg|_{z^m = c^m} \frac{dc^j}{dt}\frac{dc^k}{dt}\left(\frac{\partial c^m}{\partial \xi^l} + i\frac{\partial c^m}{\partial \xi^{l+n}}\right) \\ &\quad + 2\Gamma^i_{jk} \frac{dc^j}{dt}\frac{d}{dt}\left(\frac{\partial c^k}{\partial \xi^l} + i\frac{\partial c^k}{\partial \xi^{l+n}}\right) = 0. \end{aligned} \tag{2.13}$$

Consider the following differential equation for functions $f^i(t)$:

$$\frac{d^2}{dt^2}f^i + \frac{\partial(\Gamma^i_{jk}\circ\tilde{\varphi}^{-1})}{\partial z^m}\Big|_{z^m=c^m}\frac{dc^j}{dt}\frac{dc^k}{dt}f^m + 2\Gamma^i_{jk}\frac{dc^j}{dt}\frac{d}{dt}f^k = 0. \tag{2.14}$$

This is a linear differential equation for f, so it has unique solutions to given initial data. In particular, $f^i \equiv 0$ is the unique solution to the initial data $f^i(0) = 0, \dot{f}^i(0) = 0$. On the other hand we see in (2.14) and (2.13) that $\partial c^i/\partial\xi^l + i(\partial c^i/\partial\xi^{l+n})$ is a solution to (2.14) with this initial data, so it is zero for all t and in particular it is zero for $t = 1$. But these are just the Cauchy–Riemann equations which show that the $c^i|_{t=1}$ are holomorphic functions of the variables $\eta^k := \xi^k + i\xi^{k+n}$. This shows that χ as defined by

$$\chi^{-1}: (\eta^1,\eta^2,\ldots,\eta^n)\mapsto\exp(\eta^k\hat{e}_k + \bar{\eta}^k\bar{\hat{e}}_k) = \exp\Big((\eta^k\frac{\partial}{\partial z^k}\Big|_p + \bar{\eta}^k\frac{\partial}{\partial\bar{z}^k}\Big|_p\Big) = \exp\Big(\xi^a\frac{\partial}{\partial x^a}\Big|_p\Big)$$

is a holomorphic chart satisfying $(\partial/\partial\eta^k)|_p = e_k$. ■

Definition 2.8: Let $p\in M$. A k-null plane N is an n-dimensional subspace of $T_p(M)$ satisfying $k|_N = 0$, i.e., $k(X,Y) = 0\ \forall X,Y\in N$, and $h|_N$ is nondegenerate, i.e., $X\in N$, $h(X,Y) = 0\ \forall Y\in N$ implies $X = 0$.

Let $\mathcal{N}$ be a real slice of M passing through p. Then by Definitions 2.3 and 2.8 $T_p(\mathcal{N})$ is a k-null plane, so k-null planes are the possible tangent spaces to real slices. Choose a basis $(\hat{e}_1,\hat{e}_2,\ldots,\hat{e}_n)$ of $T_p^{1,0}(M)$ such that $g(\hat{e}_i,\hat{e}_k) = i\delta_{ik}$. Define $e_1,e_2,\ldots,e_n$ by $e_i := \hat{e}_i + \bar{\hat{e}}_i$. Then $h(e_i,e_k) + ik(e_i,e_k) = g(e_i,e_k) = g(\hat{e}_i,\hat{e}_k) = i\delta_{ik}$. From this and (2.9) it follows that $(e_1,\ldots,e_n,Je_1,\ldots,Je_n)$ form a basis of $T_p(M)$ satisfying $k(e_i,e_k) = -k(Je_i,Je_k) = \delta_{ik}$ and $k(Je_i,e_k) = h(e_i,e_k) = 0$. Using this basis one can describe the k-null planes in $T_p(M)$.

Proposition 2.9: There is a one-to-one correspondence between the k-null planes N and orthogonal matrices $\alpha\in 0(n)$ with $\alpha + \alpha^T$ nondegenerate: N is spanned by the n vectors $f_i = e_i + \alpha^j_i Je_j$ and $h|_N$ is given by $h(f_i,f_j) = -(\alpha^j_i + \alpha^i_j)$.

Sketch of the proof: (for more details see Ref. 8): Set $P = \mathrm{span}\{e_1,e_2,\ldots,e_n\}$. P satisfies $T_p(M) = P\oplus JP$ and $JP = P^\perp$ ($P^\perp$ denotes the orthogonal complement with respect to k). Define $\pi: T_p(M)\mapsto P$ and $\pi^\perp: T_P(M)\mapsto P^\perp$ by $X = \pi X + \pi^\perp X, \pi X\in P, \pi^\perp X\in P^\perp$ for all $X\in T_p(M)$. Now given a k-null plane N it follows from the fact that $k|_P$ is positive definite and $k|_{P^\perp}$ is negative definite that $\pi|_N$ and $\pi^\perp|_N$ are isomorphisms and so one can define $\tilde{\alpha}: P\mapsto P$ as $\tilde{\alpha} := -J\circ(\pi^\perp|_N)\circ(\pi|_N)^{-1}$. This implies that

$$(\mathbf{1} + J\tilde{\alpha})|_P = ((\pi|_N)\circ(\pi|_N)^{-1} + J\tilde{\alpha})|_P = ((\pi|_N) + (\pi^\perp|_N))\circ(\pi|_N)^{-1}|_P = (\pi|_N)^{-1}|_P,$$

so $(\mathbf{1} + J\tilde{\alpha})$ is injective and $N = (\mathbf{1} + J\tilde{\alpha})P$. Then

$$0\equiv k((\mathbf{1} + J\tilde{\alpha})|_P\cdot,(\mathbf{1} + J\tilde{\alpha})|_P\cdot)\equiv k(\cdot,\cdot) - k(\tilde{\alpha}\cdot,\tilde{\alpha}\cdot)$$

implies that $\tilde{\alpha}$ is an orthogonal mapping. Define $\alpha\in 0(n)$ by $\tilde{\alpha}(e_i) = \alpha^j_i e_j$. The rest of the proof is straightforward verification. ■

Let $\mathscr{D}(n)$ denote the set $\mathscr{D}(n) = \{\alpha\in O(n)|\alpha + \alpha^T$ nondegenerate$\}$. $\mathscr{D}(n)$ is an open submanifold of $O(n)$ since the set $O(n)\setminus\mathscr{D}(n) = \{\alpha\in O(n)|\det(\alpha + \alpha^T) = 0\}$ is closed as $\det(\cdot + \cdot^T)$ is a continuous function on $O(n)$. Proposition 2.9 allows to consider the set of all k-null planes as a manifold diffeomorphic to $\mathscr{D}(n)$.

One can show (a proof is given in Ref. 8) that $\mathscr{D}(n)$ has $n + 1$ connection components, one to every possible signature $(p,n-p)$ of the symmetric bilinearform $-(\alpha + \alpha^T), \alpha\in\mathscr{D}(n)$.

Using these orthogonal matrices one can also get some restrictions on the possible intersection of two k-null planes.

Proposition 2.10: Let N_1,N_2 be two k-null planes and let $(p_1,n-p_1)$, $(p_2,n-p_2)$, $p_2\geqslant p_1$ denote the signatures of $h|_{N_1}$, $h|_{N_2}$, respectively.

Define m, l_1, l_2 by: $m := \dim(N_1\cap N_2)$, and (l_1,l_2) the signature of $h|_{N_1\cap N_2}$. m, l_1, l_2 satisfy the following relations: (a) $(-1)^m = (-1)^{n+p_1+p_2}$, (b) $m - l_2\leqslant p_1$, (c) $m - l_1\leqslant n - p_2$, (d) $m\leqslant n + p_1 - p_2$. Conversely, given p_1, p_2, l_1, l_2, m satisyfing the above relations, there exist k-null planes N_1, N_2 such that $(p_1,n-p_1)$, $(p_2,n-p_2)$ are the signatures of $h|_{N_1}$, $h|_{N_2}$, respectively, $\dim(N_1\cap N_2) = m$ and the signature of $h|_{N_1\cap N_2}$ is (l_1,l_2).

Sketch of the proof (for more details see Ref. 8)*:* As in the proof of Proposition 2.9 let $N_1 = (\mathbf{1} + J\tilde{\alpha}_1)P$ and $N_2 = (\mathbf{1} + J\tilde{\alpha}_2)P$, $\tilde{\alpha}_1,\tilde{\alpha}_2\in\mathscr{D}(n)$. It follows that $N_1\cap N_2 = (\mathbf{1} + J\tilde{\alpha}_1)E$, where E is the eigenspace of $\tilde{\alpha}_1^{-1}\tilde{\alpha}_2$ to the eigenvalue 1 and therefore $m := \dim(N_1\cap N_2) = \dim(E)$. Considering the characteristical polynomial of the orthogonal mapping $\tilde{\alpha}_1^{-1}\tilde{\alpha}_2$ one finds that

$$\det(\tilde{\alpha}_1)\det(\tilde{\alpha}_2) = \det(\tilde{\alpha}_1^{-1}\tilde{\alpha}_2) = (-1)^{n-m}.$$

On the other hand, one finds [using the fact that the set of all $\alpha\in\mathscr{D}(n)$ such that $-(\alpha + \alpha^T)$ has a given signature is a connected open subset of $O(n)$] that

$$\det(\tilde{\alpha}_1)\det(\tilde{\alpha}_2) = (-1)^{p_1}(-1)^{p_2} = (-1)^{p_1+p_2}.$$

These two equations prove (a). Relations (b), (c), and (d) follow from the restrictions on the signature of $h|_{N_1\cap N_2}$ one gets using that $h|_{N_1\cap N_2} = h|_{N_1}|_{N_1\cap N_2} = h|_{N_2}|_{N_1\cap N_2}$. As to the last part of the proposition one can explicitly construct matrices α_1, α_2 such that $\alpha_1^T\alpha_2$ has an eigenspace E to the eigenvalue 1 such that $\dim(E) = m$ and the signature of $-(\alpha_1 + \alpha_1^T)|_E$ is (l_1,l_2). ■

Proposition 2.4 implies that if there is a real slice $\mathcal{N}$ tangent to a given k-null plane $N\in T_p(M)$ then $\mathcal{N}$ is just the image under the exponential mapping of N, since any line passing through the origin in N gets mapped onto a geodesic in M, which by Proposition 2.4 lies entirely in $\mathcal{N}$. This also implies that the restrictions on the possible intersections of k-null planes in Proposition 2.10 are directly applicable to the intersections of real slices (of course this is not true for the existence part).

Combining the above remark with Proposition 2.9 one works out (see Ref. 8) the following corollary which will be useful in the subsequent applications.

Corollary 2.11: Let $(M,g = h + ik)$ be a complex Riemannian manifold, $p\in M$. Choose complex normal coordinates (W,χ) at p as in Definition 2.6. Let $\alpha\in\mathscr{D}(n)$ and define $\mathscr{A}^i_k := \frac{1}{2}(1+i)\delta^i_k - \frac{1}{2}(1-i)\alpha^i_k$. Introduce coordinates (W,Λ) by

$$\Lambda^{-1}: \quad \mathbb{C}^n \mapsto W$$
$$(\lambda^1,\lambda^2,...,\lambda^n) \mapsto \chi^{-1}(\mathscr{A}^1_k\lambda^k,\mathscr{A}^2_k\lambda^k,...,\mathscr{A}^n_k\lambda^k).$$

Then
$N=\{\xi^k(\mathscr{A}^i_k\hat{e}_i+\bar{\mathscr{A}}^i_k\bar{\hat{e}}_i)\,|\,(\xi^1,\xi^2,...,\xi^n)\in\mathbb{R}^n\}\subset T_p(M)$ is a k-null plane and there is a real slice $\mathscr{N}$ tangent to N if and only if metric coefficients $g((\partial/\partial\lambda)^i,(\partial/\partial\lambda)^k)|_{\Lambda^{-1}}$ restricted to $(\lambda^1,\lambda^2,...,\lambda^n)\in\mathbb{R}^n$ are real. Every real slice passing through p can be described in this way.

This corollary shows that every real slice can be obtained locally by restricting an appropriate holomorphic coordinate system to real values.

III. TWO APPLICATIONS

A. The complexified sphere ${}^cS^n_r$

In the following n-tuples $(z_1,z_2,...,z_n)\in\mathbb{C}^n$ will be denoted by $\mathbf{z}\in\mathbb{C}^n$ and z will denote a square root of $\mathbf{z}\cdot\mathbf{z}$, that is $z^2=\Sigma_i z^i z^i$.

For some $r\in\mathbb{C}\setminus\{0\}$ define the manifold ${}^cS^n_r$ as the submanifold of $\mathbb{C}^{n+1}$ determined by the equation:

$$(\xi^0)^2+(\xi^1)^2+\cdots+(\xi^n)^2=r^2. \tag{3.1}$$

Denote by N, S the points $N:=(r,0)$, $S:=(-r,0)$. In order to see that ${}^cS^n_r$ is really a complex manifold introduce an atlas on ${}^cS^n_r$ consisting of the two complex stereographical projections (U_N,φ_N), (U_s,φ_s) about the points N,S, respectively,

$$U_N:=\{(\xi^0,\boldsymbol{\xi})\in{}^cS^n_r\,|\,\xi^0\neq r\},$$
$$U_S:=\{(\xi^0,\boldsymbol{\xi})\in{}^cS^n_r\,|\,\xi^0\neq -r\},$$
$$V=\varphi_N(U_N)=\varphi_S(U_S)=\{\mathbf{z}\in\mathbb{C}^n\,|\,\mathbf{z}^2\neq -r^2\}$$
$$\varphi_N: U_N\mapsto V\in\mathbb{C}^n,\quad (\xi^0,\boldsymbol{\xi})\mapsto[r/(r-\xi^0)]\boldsymbol{\xi},$$
$$\varphi_S: U_S\mapsto V\in C^n,\quad (\xi^0,\boldsymbol{\xi})\mapsto[r/(r+\xi^0)]\boldsymbol{\xi}. \tag{3.2}$$

Clearly U_N and U_S are open and cover ${}^cS^n_r$. Calculating the coordinate transformations one finds:

$$\varphi_N\circ\varphi_S^{-1}=\varphi_S\circ\varphi_N^{-1}: V'\mapsto V',$$
$$\mathbf{z}\mapsto(r^2/z^2)\mathbf{z}$$
$$V'=\varphi_N(U_N\cap U_S)=\varphi_S(U_N\cap U_S)$$
$$=\{\mathbf{z}\in\mathbb{C}^n\,|\,z^2\neq 0, z^2\neq -r^2\}. \tag{3.3}$$

These transformations are holomorphic, so the atlas (U_N,φ_N), (U_S,φ_S) turns ${}^cS^n_r$ into a complex manifold.

On $\mathbb{C}^{n+1}$ there is a natural complex Riemannian metric δ given by $\delta=\delta_{AB}d\xi^A\otimes d\xi^B$. Let I: ${}^cS^n_r\mapsto\mathbb{C}^{n+1}$ denote the inclusion of ${}^cS^n_r$ into $\mathbb{C}^{n+1}$ and I_* its differential. The metric δ induces a complex Riemannian metric g on ${}^cS^n_r$ by $g(Z_1,Z_2):=g(I_*(Z_1),I_*(Z_2))$ for $Z_1,Z_2\in T^c_p({}^cS^n_r)$, $p\in{}^cS^n_r$. In stereographical coordinates g is given by

$$g|_U=\frac{4r^4}{(r^2+z^2)^2}\delta_{ij}\,dz^i\otimes dz^j, \tag{3.4}$$

where $(U,(z^1,...,z^n))$ is any one of the charts (U_N,φ_N), (U_S,φ_S). Equation (3.4) shows that g is holomorphic and nondegenerate and hence a complex Riemannian metric.

An isometry S between two complex Riemannian manifolds (M,g), (M',g') is a holomorphic diffeomorphism S: $M\mapsto M'$ such that $S^*g'=g$. The isometry group of ${}^cS^n_r$ is the subgroup $\mathscr{CO}(n+1)$ of $GL(n+1,\mathbb{C})$ given by $\mathscr{CO}(n+1)=\{A\in GL(n+1,\mathbb{C})\,|\,AA^T=1\}$. This is due to the following proposition.

Proposition 3.1: Given p, $q\in{}^cS^n_r$, $(e_1,e_2,...,e_n)$, $(f_1,f_2,...f_n)$ orthonormal bases of $T^{1,0}_p({}^cS^n_r, T^{1,0}_q({}^cS^n_r)$, respectively, there is exactly one isometry of ${}^cS^n_r$ which takes $(e_1,e_2,...,e_n)$ to $(f_1,f_2,...f_n)$.

Proof: Let again I: ${}^cS^n_r\mapsto\mathbb{C}^{n+1}$ denote the inclusion of ${}^cS^n_r$ into $\mathbb{C}^{n+1}$. Define $E_1,E_2,...,E_n,F_1,F_2,...,F_n\in T^{1,0}_0(\mathbb{C}^{n+1})$ as the vectors $I_*e_1,I_*e_2,...,I_*e_n$, $I_*f_1,I_*f_2,...,I_*f_n$ parallely transported to the origin $O\in\mathbb{C}^{n+1}$. Identifying $T^{1,0}_0(\mathbb{C}^{n+1})$ with $\mathbb{C}^{n+1}$ in the natural way one can define $E_0:=(1/r)I(p)$ and $F_0:=(1/r)I(q)$. $(E_0,E_1,...,E_n)$ and $(F_0,F_1,...,F_n)$ form orthonormal bases of $T^{1,0}_0(\mathbb{C}^{n+1})$. The isometry $\mathscr{S}$ of $T^{1,0}_0(\mathbb{C}^{n+1})$ which takes $(E_0,E_1,...,E_n)$ to $(F_0,F_1,...,F_n)$ can be interpreted as an isometry $\mathscr{S}'$ of $(\mathbb{C}^{n+1},\delta)$ whose restriction to $I({}^cS^n_r)$ induces an isometry S of ${}^cS^n_r$. This isometry takes $(e_1,e_2,...,e_n)$ to $(f_1,f_2,...,f_n)$ and it is the only such isometry since two isometries coincide if they agree on one basis of a tangent space. ■

Using these isometries we get the following corollary.

Corollary 3.2: Assume there is a real slice $\mathscr{N}$ in ${}^cS^n_r$ and the signature of $h|_{\mathscr{N}}$ is $(p,n-p)$. Then to every k-null plane N in any tangent space $T_q({}^cS^n_r)$ such that $h_{|N}$ has signature $(p,n-p)$ there is a real slice $\mathscr{M}$ of ${}^cS^n_r$ tangent to N and $\mathscr{M}$ and $\mathscr{N}$ are isometric.

Proof: Choose some point $q'\in\mathscr{N}$ and a basis $(E_1,E_2,...,E_n)$ of the k-null plane $T_{q'}(\mathscr{N})$ orthogonal with respect to h and normalized to $h(E_i,E_i)=1$ for $i\leqslant p$ and $h(E_i,E_i)=-1$ for $i>p$. Define the orthonormal basis $(e_1,e_2,...,e_n)$ of $T^{1,0}_{q'}({}^cS^n_r)$ by $e_k=\frac{1}{2}(1-iJ)E_k$ for $k\leqslant p$ and $e_k=(i/2)(1-iJ)E_k$ for $k>p$. In the same way construct an orthonormal basis $(f_1,f_2,...f_n)$ starting from an orthogonal basis of N. Let S be the isometry of ${}^cS^n_r$ which takes $(e_1,e_2,...,e_n)$ to $(f_1,f_2,...f_n)$. Then $S_*(T_{q'}(\mathscr{N}))=N$ and since an isometry clearly maps a real slice to a real slice, $\mathscr{M}:=S(\mathscr{N})$ is a real slice tangent to N and isometric to $\mathscr{N}$. ■

Corollary 3.2 shows that it is sufficient to find one real slice $\mathscr{N}_p$ of ${}^cS^n_r$ to every possible signature $(p,n-p)$ of $h|_{\mathscr{N}_p}$. This is easily done for $r\in\mathbb{R}$, $r>0$.

Proposition 3.3: Let $r\in\mathbb{R}$, $r>0$. To every k-null plane N of any tangent space there is a real slice $\mathscr{N}$ tangent to N, and, depending on the signature $(p,n-p)$ of $h|_{\mathscr{N}}$, $\mathscr{N}$ is isometric to

$p=n$:	the n-dimensional sphere S^n,
$p=n-1$:	de Sitter space-time,
$\vdots$	(some other spaces of constant curvature),
$p=1$:	anti-de Sitter space-time,
$p=0$:	the hyperbolic plane H^n.

Proof: Let ${}^p\eta_{ik}$ denote the diagonal matrix with the first p diagonal elements equal to 1 the remaining $n-p$ equal to -1. Consider the real slice R_p of $(\mathbb{C}^{n+1},\delta)$ given by $\xi^0=\lambda^0$, $\xi^1=\lambda^1,...,\xi^p=\lambda^p,\xi^{p+1}=i\lambda^{p+1},...,\xi^n=i\lambda^n$, $(\lambda^0,\lambda^1,...,\lambda^n)\in\mathbb{R}^{n+1}$. The metric $\delta|_{R_p}$ then reads

$$\delta|_{R_p}=d\lambda^0\otimes d\lambda^0+{}^p\eta_{ik}\,d\lambda^i\otimes d\lambda^k. \tag{3.5}$$

The intersection $\mathcal{N}_p$ of R_p with ${}^c S^n_r$ is an n-dimensional submanifold of R_p given by the equation

$$(\lambda^0)^2 + {}^p\eta_{ik}\lambda^i\lambda^k = r^2 \tag{3.6}$$

and clearly $g|_{\mathcal{N}_p} = \delta|_{R_p}|_{\mathcal{N}_p}$. This shows that $\mathcal{N}_p$ is a real slice of ${}^c S^n_r$ and (3.5) and (3.6) are just the usual embeddings of the spaces of constant curvature listed in the proposition in $\mathbb{R}^{n+1}$. The rest of the proposition follows from Corollary 3.2. ■

In order to answer the question whether there are any real slices in ${}^c S^n_r$ for $r \notin \mathbb{R}$ we can apply Corollary 2.11. Due to Corollary 3.2 it suffices to find the answer for the real slices passing through some given point, say the point S. Take the stereographical coordinates (U,φ_N) and define the orthonormal basis $e_i := \frac{1}{2}(\partial/\partial z^i)|_S$ of $T^{1,0}_S({}^c S^n_r)$. Define the coordinates (W,χ) by

$$\varphi_N \circ \chi^{-1}(\boldsymbol{\eta}) = r\frac{\boldsymbol{\eta}}{\eta}\frac{\sin(\eta/r)}{1+\cos(\eta/r)}. \tag{3.7}$$

One checks that $(\partial/\partial\eta)^i|_{\chi^{-1}(0)} = e_i$ and that χ are holomorphic coordinates in some neighborhood of S. The metric expressed in terms of χ is calculated to be

$$g|_W = \sum_{i,k}\left(\frac{r^2}{\eta^2}\sin^2\left(\frac{\eta}{r}\right)\left(\delta_{ik} - \frac{\eta^i\eta^k}{\eta^2}\right) + \frac{\eta^i\eta^k}{\eta^2}\right)d\eta^i \otimes d\eta^k. \tag{3.8}$$

Verifying now that $\Gamma^i_{jk}(\chi^{-1}(\eta^1,\eta^2,\ldots,\eta^n))\eta^j\eta^k = 0$ one gets by Proposition 2.7 that (W,χ) are complex normal coordinates.

Now calculate the metric in terms of the coordinates (W,Λ) introduced in Corollary 2.11,

$$g = \left(\frac{r^2}{d_{ij}\lambda^i\lambda^j}\sin^2\left(\sqrt{\frac{d_{ij}\lambda^i\lambda^j}{r^2}}\right)\left(d_{kl} - \frac{d_{ik}d_{jl}\lambda^i\lambda^j}{d_{ij}\lambda^i\lambda^j}\right) + \frac{d_{ik}d_{jl}\lambda^i\lambda^j}{d_{ij}\lambda^i\lambda^j}\right)d\lambda^k \otimes d\lambda^l, \tag{3.9}$$

where d_{ik} is given by $d_{ik} = -\frac{1}{2}(\alpha^i_k + \alpha^k_i)$. Now by Corollary 2.11 there is a real slice tangent to the k-null plane determined by α if and only if g is real for all $(\lambda^1,\lambda^2,\ldots,\lambda^n) \in \mathbb{R}^n$. This implies the following proposition.

Proposition 3.4: If $r^2 \in \mathbb{R}$ then to every k-null plane in a tangent space $T^{1,0}_p({}^c S^n_r)$ there exists exactly one real slice tangent to this k-null plane. For $r^2 \notin \mathbb{R}$, ${}^c S^n_r$ has no real slices at all.

Finally one checks that $({}^c S^n_{kr}, g)$ and $({}^c S^n_r, k^2 g)$ are isometric to each other for all $k \in \mathbb{C}\setminus\{0\}$. This shows that the real slices of ${}^c S^n_r$ with $r \in i\mathbb{R}$ are the same as those for $r \in \mathbb{R}$ discussed in Proposition 3.3 up to an overall change of sign in the signatures.

As a last point, as an application of Proposition 2.10, we write down (for $r \in \mathbb{R}$, $r > 0$) all possible intersections of two real slices $\mathcal{N}_1$, $\mathcal{N}_2$ of the four-dimensional complexified sphere ${}^c S^4_r$ for the physically interesting cases of $\mathcal{N}_1,\mathcal{N}_2$ being isometric to the sphere S^4, de Sitter space-time dS^4, anti-de Sitter space-time adS^4, and the hyperbolic plane H^4. The upper index will denote the dimension of the manifold. [Since to every k-null plane there is a real slice, Proposition 2.10 yields the possible dimensions m of the intersections and the possible signatures (l_1,l_2) of the metric restricted to the real slices. To identify the intersections one can consider the metric (3.9) restricted to the intersection.]

Corollary 3.5: Let $\mathcal{N}_1,\mathcal{N}_2$ be two real slices passing through a fixed point $p \in {}^c S^4_r$, isometric ($\simeq$) to either of the spaces S^4, dS^4, adS^4, or H^4. The possible intersections of $\mathcal{N}_1$ and $\mathcal{N}_2$ are

$\mathcal{N}_1 \simeq S^4$	, $\mathcal{N}_2 \simeq S^4$	: S^4, S^2, S^0,
$\mathcal{N}_1 \simeq S^4$	, $\mathcal{N}_2 \simeq dS^4$	: S^3, S^1,
$\mathcal{N}_1 \simeq S^4$	, $\mathcal{N}_2 \simeq adS^4$	: S^1,
$\mathcal{N}_1 \simeq S^4$	, $\mathcal{N}_2 \simeq H^4$	: $\{p\}$,
$\mathcal{N}_1 \simeq dS^4$	, $\mathcal{N}_2 \simeq dS^4$	: $dS^4, S^2, dS^2, S^1 \times \mathbb{R}$ with a degenerate metric, S^0,
$\mathcal{N}_1 \simeq dS^4$	, $\mathcal{N}_2 \simeq adS^4$	: dS^2, S^0,
$\mathcal{N}_1 \simeq dS^4$	, $\mathcal{N}_2 \simeq H^4$	: H^1,
$\mathcal{N}_1 \simeq adS^4$	, $\mathcal{N}_2 \simeq adS^4$	: $adS^4, H^2, adS^2,, \mathbb{R}^2$ with a degenerate metric, S^0,
$\mathcal{N}_1 \simeq adS^4$	, $\mathcal{N}_2 \simeq H^4$	: H^3, H^1,
$\mathcal{N}_1 \simeq H^4$	, $\mathcal{N}_2 \simeq H^4$	: $H^4, H^2, \{p\}$.

B. Complexified Robertson–Walker space-times ${}^c\mathcal{R}$

Define a complexified Robertson–Walker space-time (RWS) ${}^c\mathcal{R}$ as follows.

Let $\mathcal{D} \subset \mathbb{C}$ be open and connected and let $r: \mathcal{D} \mapsto \mathbb{C}$ be a holomorphic function such that $r(x) \neq 0\ \forall x \in \mathcal{D}$. Define the complex manifold ${}^c\mathcal{R}$ as ${}^c\mathcal{R} = \mathcal{D} \times {}^c S^n_1$ and introduce the two projections π_1, π_2,

$$\pi_1: {}^c\mathcal{R} \mapsto \mathcal{D}, \quad \pi_2: {}^c\mathcal{R} \mapsto {}^c S^n_1, \\ (x,p) \mapsto x, \quad (x,p) \mapsto p. \tag{3.10}$$

Choose a stereographic atlas (U_N,φ_N), (U_S,φ_S) on ${}^c S^n_1$. Define the atlas (V_N,ψ_N), (V_S,ψ_S) on ${}^c\mathcal{R}$ by

$$V_{N,S} := \mathcal{D} \times U_{N,S}, \\ \psi_{N,S}: V_{N,S} \mapsto \mathbb{C}^{n+1}, \\ p \mapsto (x,\mathbf{z}) = (\pi_1(p), \varphi_{N,S}(\pi_2(p))). \tag{3.11}$$

Introduce a complex Riemannian metric g on ${}^c\mathcal{R}$ by

$$g|_V = dx \otimes dx + \frac{4r^2(x)}{(1+z^2)^2}\delta_{ik}\, dz^i \otimes dz^k, \tag{3.12}$$

where $(V,(x,\mathbf{z}))$ is one of the charts of the atlas (V_N,ψ_N), (V_S,ψ_S). The $x = \text{const}$ hypersurfaces of ${}^c\mathcal{R}$ are isometric to $({}^c S^n_1, r^2(x)g)$ which is in turn isometric to ${}^c S^n_{r(x)}$.

Define the vector field $\hat{e}_0$ to be the field $\partial/\partial x$ in either of the coordinates (V_N,ψ_N) or (V_S,ψ_S). By (3.12) $\hat{e}_0|_p$ is a unit vector for all $p \in {}^c\mathcal{R}$. The symmetries of ${}^c S^n_r$ give rise to symmetries of ${}^c\mathcal{R}$.

Corollary 3.6: Let $p,q \in {}^c\mathcal{R}$ such that $\pi_1(p) = \pi_1(q)$ and let $(\hat{e}_0|_p,\hat{e}_1,\ldots,\hat{e}_n)$ and $(\hat{e}_0|_q,f_1,\ldots,f_n)$ be orthonormal bases of $T^{1,0}_p({}^c\mathcal{R})$, $T^{1,0}_q({}^c\mathcal{R})$, respectively. There is an isometry S of ${}^c\mathcal{R}$ which takes $(\hat{e}_0|_p,\hat{e}_1,\ldots,\hat{e}_n)$ to $(\hat{e}_0|_q,f_1,\ldots,f_n)$.

Next consider complex normal coordinates (W,χ) at $p \in {}^c\mathcal{R}$ with respect to an orthonormal basis of $T^{1,0}_p({}^c\mathcal{R})$ of the form $(\hat{e}_0|_p,\hat{e}_1,\ldots,\hat{e}_n)$. In the following capital latin letters $A,B,C,\ldots$ will denote indices running from 0 to n. By Corollary 3.6 the functional dependence on $(\eta^0,\boldsymbol{\eta})$ of the metric components

$$g\left(\frac{\partial}{\partial\eta^A},\frac{\partial}{\partial\eta^B}\right)\bigg|_{\chi^{-1}(\eta^0,\boldsymbol{\eta})}$$

can depend only on $x:=\pi_1(p)$ and not on the special choice of $p\in\pi_1^{-1}(x)$ and of the orthonormal family $(\hat{e}_1,\hat{e}_2,\ldots,\hat{e}_n)$. We may therefore assume that $p=\psi_N^{-1}(x,\mathbf{0})$ and

$$\hat{e}_i=\frac{1}{2r(x)}\frac{\partial}{\partial z^i}\bigg|_p .$$

Making for the affinely parametrized geodesic γ: $t\mapsto\chi^{-1}(\eta^0 t,\boldsymbol{\eta}t)$ the following ansatz [guided by (3.7)]

$$\psi_N\circ\chi^{-1}(\eta^0 t,\boldsymbol{\eta}t)=\left(\rho(t),\frac{\boldsymbol{\eta}}{\eta}\frac{\sin(f(t)\eta t)}{1+\cos(f(t)\eta t)}\right) \tag{3.13}$$

one can calculate the geodesic equation (2.11) and finds that the complex normal coordinates χ are given by

$$\psi_N\circ\chi^{-1}(\eta^0,\boldsymbol{\eta})=\left(\rho,\frac{\boldsymbol{\eta}}{\eta}\frac{\sin(f\eta)}{1+\cos(f\eta)}\right) \tag{3.14}$$

where $f=f(1)$ and $\rho=\rho(1)$ and $f(t)$ and $\rho(t)$ are solutions to the following differential equation with initial data:

$$\ddot{\rho}-\eta^2 r(\rho)r'(\rho)\dot{f}^2=0,\quad \rho(0)=x,\quad \dot{\rho}(0)=\eta^0,$$
$$\ddot{f}+2[r'(\rho)/r(\rho)]\dot{f}\dot{\rho}=0,\quad f(0)=0,\quad \dot{f}(0)=1/r(x). \tag{3.15}$$

Equation (3.15) shows that f and ρ are functions of η^2 and η^0.

From Proposition 2.7 we know that the complex normal coordinates χ are holomorphic coordinates, so the metric coefficients of the metric g in these coordinates are represented by a power series in $\eta^0,\eta^1,\ldots,\eta^n$. Having calculated the metric in terms of χ using (3.14) one can use (3.15) to calculate the coefficients of this power series. After a somewhat lengthy computation one finds that the expansion up to second order is given by [r,r',r'' stand for $r:=r(x)$, $r':=r'(x)$, $r'':=r''(x)$]

$$\begin{aligned}g&=\delta_{AB}\,d\eta^A\otimes d\eta^B\\&\quad+[(1-r'^2)/3r^2](\delta_{AB}\delta_{CD}\eta^A\eta^C\,d\eta^B\otimes d\eta^D\\&\quad-\delta_{AB}\eta^A\eta^B\delta_{CD}\,d\eta^C\otimes d\eta^D)\\&\quad+[(1-r'^2+rr'')/3r^2]\delta_{AB}(\eta^A\,d\eta^0-\eta^0\,d\eta^A)\\&\quad\otimes(\eta^B\,d\eta^0-\eta^0\,d\eta^B)+O(3).\end{aligned} \tag{3.16}$$

In order to apply Corollary 2.11, change to the coordinates $\Lambda=(\lambda^0,\lambda^1,\ldots,\lambda^n)$ as indicated in Corollary 2.11 [set again $d_{AB}:=-\frac{1}{2}(\alpha^A_B+\alpha^B_A)$],

$$\begin{aligned}g&=d_{AB}\,d\lambda^A\otimes d\lambda^B\\&\quad+[(1-r'^2)/3r^2](d_{AB}d_{CD}\lambda^A\lambda^C\,d\lambda^B\otimes d\lambda^D\\&\quad-d_{AB}\lambda^A\lambda^B d_{CD}\,d\lambda^C\otimes d\lambda^D)\\&\quad+[(1-r'^2+rr'')/3r^2]d_{AB}\mathscr{A}^0_C\mathscr{A}^0_D\\&\quad\times(\lambda^A\,d\lambda^C-\lambda^C\,d\lambda^A)\\&\quad\otimes(\lambda^B\,d\lambda^D-\lambda^D\,d\lambda^B)+O(3)\end{aligned} \tag{3.17}$$

Now consider the restriction of $(\lambda^0,\lambda^1,\ldots,\lambda^n)$ to $\mathbb{R}^{n+1}$. If this restriction describes a real slice, by Corollary 2.11 the coefficients of the metric must be real, so in particular their Taylor expansion up to any order must be real, so from the form (3.17) of the metric we can get necessary conditions which α has to satisfy for (3.17) to describe a real slice. The first possibility is that the coefficient of the third term vanishes identically.

Proposition 3.7: $(1-r'^2+rr'')/3r^2$ vanishes identically if and only if $r(x)=a\sin((1/a)x+b)$ for some $a,b\in\mathbb{C}$. In this case, ${}^c\mathscr{R}$ is locally isometric to ${}^cS^{n+1}_a$ and for the real slices of ${}^c\mathscr{R}$ the results of the previous section apply.

Proof: $r(x)=a\sin((1/a)x+b)$ is the general solution of $(1-r'^2+rr'')/3r^2=0$. To establish the local isometry to S_a introduce complex normal coordinates (W',χ') on ${}^cS^{n+1}_a$ as given in (3.7) and (3.8). Introduce new coordinates Ψ on ${}^cS^{n+1}_a$ by

$$\chi'\circ\Psi^{-1}(x,\mathbf{z})=(\eta^0,\eta^1,\ldots,\eta^n),$$

where

$$\eta^0=x[(z^2-1)/(z^2+1)],\quad \boldsymbol{\eta}=x[2\mathbf{z}/(z^2+1)].$$

Calculating the metric (3.8) in terms of the coordinates $\Psi=(x,\mathbf{z})$ one gets (3.12) with $r(x)=a\sin((1/a)x)$. This establishes the local isometry by identifying the coordinates Ψ with ψ_N on ${}^c\mathscr{R}$ for instance. An affine transformation of the coordinate x then yields the general form $r(x)=a\sin((1/a)x+b)$. ■

Otherwise we get the following result.

Proposition 3.8: Assume that ${}^c\mathscr{R}$ is not locally isometric to ${}^cS^{n+1}_a$ and let $n\geqslant 2$. There is a real slice passing through a point p with $\pi_1(p)=x$ if and only if $r^2(x+t)$ or $r^2(x+it)$ is real for all $t\in\mathbb{R}$.

Sketch of the proof (for more details see Ref. 8): Assume first that there is a real slice $\mathscr{N}$ passing through p. We may further assume that $[(1-r'^2+rr'')/3r^2]|_{\pi_1(p)}$ does not vanish (if it would vanish for all $p\in\mathscr{N}$, it would vanish identically, contradicting the assumption by Proposition 3.7). Let α be the orthogonal mapping describing the k-null plane $T_p(\mathscr{N})$ in (3.17).

Considering the third term in (3.17) define the function B on $(\mathbb{R}^{n+1})^3$ given by

$$B(\lambda,\xi,\zeta)=d_{AB}\mathscr{A}^0_C\mathscr{A}^0_D(\lambda^A\xi^C-\lambda^C\xi^A)(\lambda^B\zeta^D-\lambda^D\zeta^B). \tag{3.18}$$

If $n\geqslant 2$ and if α^0_0 is not equal to $+1$ or -1, then the values assumed by B on $(\mathbb{R}^{n+1})^3$ do not lie in a one-dimensional subspace of $\mathbb{C}$ (considering $\mathbb{C}$ as a vector space over $\mathbb{R}$). This is due to the following: $\mathscr{A}^0_C$ is given by $\mathscr{A}^0_C=\frac{1}{2}(1+i)\delta^0_C-\frac{1}{2}(1-i)\alpha^0_C$. If α^0_0 is not equal to $+1$ or -1, δ^0_C and α^0_C are linearly independent, so the complex-valued bilinear form $\mathscr{A}$ on $\mathbb{R}^{n+1}$ given by $\mathscr{A}(\xi)=\mathscr{A}^0_C\xi^C$ is surjective onto $\mathbb{C}$. One can then show that one can choose three linearly independent vectors of $\mathbb{R}^{n+1}$ (here we use $n\geqslant 2$) which yield two linearly independent complex values when plugged into (3.18) in different ways.

This shows that α^0_0 must be either $+1$ or -1 since otherwise the metric (3.17) cannot be real for all $(\lambda^0,\lambda^1,\ldots,\lambda^n)\in\mathbb{R}^{n+1}$. Taking this into account, we can write (3.17) as

$$g=d_{AB}\,d\lambda^A\otimes d\lambda^B+[(1-r'^2)/3r^2](d_{AB}d_{CD}\lambda^A\lambda^C\,d\lambda^B\otimes d\lambda^D-d_{AB}\lambda^A\lambda^B d_{CD}\,d\lambda^C\otimes d\lambda^D)$$

$$- \alpha_0^0[(1 - r'^2 + rr'')/3r^2]d_{AB}(\lambda^A\, d\lambda^0 - \lambda^0\, d\lambda^A)$$
$$\otimes(\lambda^B\, d\lambda^0 - \lambda^0\, d\lambda^B) + O(3).$$

Since in $n \geqslant 2$ the second and the third term may vary independently of each other, the reality of g implies the reality of both coefficients:

$$(1 - r'^2)/3r^2 \text{ and } (1 - r'^2 + rr'')/3r^2 \text{ must be real.} \quad (3.19)$$

On the other hand $\alpha_0^0 = 1$ or $\alpha_0^0 = -1$ implies that γ: $t \mapsto \Lambda^{-1}(t,0,...,0) = \chi^{-1}(vt,0,...,0)$, where $v = i$ if $\alpha_0^0 = 1$ and $v = 1$ if $\alpha_0^0 = -1$, gets mapped to $\psi_N \circ \gamma(t) = (x + vt,0,...,0)$ [see (3.15)]. So the real slice $\mathscr{N}$ contains points q with $\pi_1(q) = x + vt$ for all t. Since the arguments above did not depend on the special choice of x, (3.19) obtains

$$[(1 - r'^2)/3r^2]|_{x+vt} \text{ and } [(1 - r'^2 + rr'')/3r^2]|_{x+vt} \quad (3.20)$$

must be real for all $t \in \mathbb{R}$. Then we get

$$\frac{d}{dt}\left(\frac{1 - r'^2}{3r^2}\right)\Bigg|_{x+vt} = -\frac{2vr'}{r}\,\frac{1 - r'^2 + rr''}{3r^2}\Bigg|_{x+vt}$$

must be real for all t, so using (3.20) $v(r'/r)$ must be real, so r'^2/r^2 must be real, so finally, using once more the first term in (3.20), r^2 must be real for all t. This proves the "only if" part of the proposition.

Assume now that for $v = i$ or $v = 1$, $r^2(x + vt)$ is real for all t. Introduce coordinates $\Phi = (x^0,x^1,...,x^n)$ by $\tilde{\psi}_N \circ \Phi^{-1}(x^0,\mathbf{x}) = (x + vx^0,\mathbf{x})$ and restrict these coordinates to real values. Then (3.12) shows that the restricted metric is real, so the coordinates describe a real slice. ■

By an argument similar to that leading to Corollary 3.2 we get the following corollary.

Corollary 3.9: Let $n \geqslant 2$, $\mathscr{N}$ a real slice of ${}^c\mathscr{R}$, $p \in \mathscr{N}$, and $x = \pi_1(p)$. Then $\mathscr{N} \cap \pi_1^{-1}(x)$ is a real slice of ${}^cS^n_{r(x)}$ and conversely to every real slice $\mathscr{M}'$ of ${}^cS^n_{r(x)}$ there is a real slice $\mathscr{M}$ of ${}^c\mathscr{R}$ such that $\mathscr{M} \cap \pi_1^{-1}(x) = \mathscr{M}'$.

The different real slices in Corollary 3.9 reflect only the results on the real slices of ${}^cS^n_r$ and are somehow trivial with respect to the geometry of ${}^c\mathscr{R}$. We will say that all real slices obtained from one another in the way described in Corollary 3.9 belong to the same class of real slices. If two real slices belonging to two different classes intersect in a point p, then by Corollary 3.9 and Proposition 3.8 $r^2(\pi_1(p) + t)$ and $r^2(\pi_1(p) + it)$ must both be real for all t. Since $r(\pi_1(p)) \neq 0$ this implies that the Taylor expansion of $r(x)$ about the point $\pi_1(p)$ can contain only even powers of $(x - \pi_1(p))$. This leads to the following corollary.

Corollary 3.10: Let $n \geqslant 2$. Two real slices belonging to different classes can pass through some point $p \in {}^c\mathscr{R}$ if and only if $r^2(\pi_1(p))$ is real and the function r is symmetric about the point $\pi_1(p)$, i.e., $r(\pi_1(p) + x) = r(\pi_1(p) - x)$ for all x. Furthermore, the intersection of two such real slices is a real slice or a totally geodesic submanifold of a real slice of ${}^cS^n_{r(x)}$.

IV. PROPAGATORS ON THE COMPLEXIFIED SPHERE

Throughout this section r will be real and positive. We will discuss ${}^cS^4_r$-invariant Green's functions to the wave-operator $\Box$-m^2, in particular, analytic continuations of such functions from one real slice to another.

First we will introduce an auxiliary function $\mathfrak{p}$. (This function is the straightforward generalization of the function p used in Dowker and Critchley.[1]) As in Sec. III let I denote the embedding mapping of ${}^cS^n_r$ into $\mathbb{C}^{n+1}$, but this time interpret $\mathbb{C}^{n+1}$ not as a complex Riemannian manifold but as a complex vector space provided with a symmetric bilinear form δ given by $\delta(v,w) = \delta_{AB}v^Aw^B$ [so for instance we have $\delta(I(p),I(p)) = r^2$ for all $p \in {}^cS^n_r$].

Every point $p \in {}^cS^n_r$ can be assigned its antipodal point denoted by p^* and defined by $I(p^*) = -I(p)$.

Define the function $\mathfrak{p}$:

$$\mathfrak{p}: {}^cS^n_r \times {}^cS^n_r \mapsto \mathbb{C}$$
$$(p,q) \mapsto \mathfrak{p}(p,q) = (1/r^2)\delta(I(p),I(q)) \quad (4.1)$$

This $\mathfrak{p}$ is a ${}^cS^n_r$-invariant function, that is, $\mathfrak{p}(p,q) = \mathfrak{p}(Sp,Sq)$ for all $p,q \in {}^cS^n_r$ and for all isometries S of ${}^cS^n_r$. Conversely one can show (see Ref. 8) that every ${}^cS^n_r$-invariant function $\tilde{f}$ is essentially a function f of $\mathfrak{p}$, that is, f defined by $f(\mathfrak{p}(p,q)) = \tilde{f}(p,q)$ is well-defined (except maybe for $p = q$ or $p = q^*$). $\mathfrak{p}$ satisfies $\mathfrak{p}(p,q) = -\mathfrak{p}(p,q^*)$ as is seen in (4.1).

Furthermore one can show (see Ref. 8) that $\mathfrak{p}(p,q)$ is real if and only if there is a real slice $\mathscr{N}$ of ${}^cS^n_r$ containing p and q or p^* and q. In this case:

If $\mathfrak{p}(p,q) > 1$: p and q can be joined by a timelike geodesic;

if $\mathfrak{p}(p,q) = 1$: p and q can be joined by a null geodesic;

if $1 > \mathfrak{p}(p,q) > -1$: p and q can be joined by a spacelike geodesic;

if $\mathfrak{p}(p,q) = -1$: p^* and q can be joined by a null geodesic;

if $\mathfrak{p}(p,q) < -1$: p^* and q can be joined by a timelike geodesic;

where timelike, null, and spacelike mean that the tangent to the geodesic has everywhere, respectively, negative, zero, or positive squared length.

Next fix some point $p \in {}^cS^n_r$ and define the sets $S^+ := \{q \in {}^cS^n_r | \mathrm{Im}(\mathfrak{p}(p,q)) > 0\}$, $S^- := \{q \in {}^cS^n_r | \mathrm{Im}(\mathfrak{p}(p,q)) < 0\}$, $S^0 := \{q \in {}^cS^n_r | \mathrm{Im}(\mathfrak{p}(p,q)) = 0\}$. S^+, S^-, S^0 are connected and S^0 is the boundary of S^+ and S^-, $S^0 = \partial S^+ = \partial S^-$.

Now turn to the wave-operator $(\Box - m^2)$. Given a holomorphic function f on ${}^cS^n_r$, define the action of $(\Box - m^2)$ on f by

$$(\Box - m^2)f|_q = g^{ik}|_q\left(\frac{\partial^2(f \circ \tilde{\varphi}^{-1})}{\partial z^i\, \partial z^k} - \Gamma^j_{ik}|_q\, \frac{\partial(f \circ \tilde{\varphi}^{-1})}{\partial z^j}\right)\Bigg|_{\tilde{\varphi}(q)} - m^2 f(q) \quad (4.2)$$

in some holomorphic chart (U,φ). This operator clearly reduces to the usual wave operator on all real slices of ${}^cS^n_r$ as is seen by choosing holomorphic coordinates which describe the real slice when restricted to real values. By Proposition 3.3 one sees, considering the signatures of the restricted metric, that on real slices isometric to de Sitter space-time m^2 must be chosen positive in order that solutions of the wave equation describe physical particles, and on real slices isometric to anti-de Sitter space-time m^2 must be chosen negative.

The propagators we will consider will turn out to be invariant and can therefore be considered as functions of $\mathfrak{p}$. So let $f(z)$ be a holomorphic function. The action of $(\Box - m^2)$ on the function $f(\mathfrak{p}(p,\cdot))$ is given by

$$(\Box - m^2)f(\mathfrak{p}(p,\cdot))|_q = \frac{d^2f}{dz^2}\frac{1}{r^2}(1-z^2) - \frac{df}{dz}\frac{nz}{r^2} - m^2 f|_{z=\mathfrak{p}(p,q)}. \tag{4.3}$$

Now turn to the four-dimensional case. The homogeneous wave equation for an invariant function f reads by (4.3),

$$\left((1-z^2)\left(\frac{d}{dz}\right)^2 - 4z\frac{d}{dz} - m^2r^2\right)f = 0. \tag{4.4}$$

Two linearly independent solutions $\mathscr{F}_1, \mathscr{F}_2$ to this equation are given by

$$m^2r^2 \neq 2: \mathscr{F}_1 = \frac{d}{dz}P_\nu(z), \quad \mathscr{F}_2 = \frac{d}{dz}Q_\nu(z),$$

where $\nu = -\frac{1}{2} + \frac{1}{2}\sqrt{9 - 4m^2r^2}$,

$$m^2r^2 = 2: \mathscr{F}_1 = \frac{1}{1+z}, \quad \mathscr{F}_2 = \frac{d}{dz}Q_0(z) \tag{4.5}$$

where P_ν, Q_ν are the Legendre functions of the first, respectively, second kind (Gradshteyn–Ryzhik[10]). The case $m^2r^2 = 2$ has to be treated separately since in this case $\nu = 0$ and $(d/dz)P_0(z) \equiv 0$. In the following we will denote the derivatives of P_ν and Q_ν by P'_ν and Q'_ν respectively. Note also that if a function $f(z)$ solves (4.4) then $f(-z)$ is another solution of (4.4).

Let us collect some properties of the Legendre functions. Most of the following formulas are either directly taken from Gradshteyn–Ryzhik[10] or follow from asymptotic expansions about the singular points -1, 1, ∞ of the Legendre equation, as they are treated for example in Smirnov.[11]

For $\nu \notin \mathbb{Z}$, P_ν is singular at -1 and ∞ and is analytic on $\mathbb{C}|[-\infty,-1]$ and the continuation onto the branch cut from above or from below follows from

$$\begin{aligned} &P_\nu(-z) = e^{i\nu\pi}P_\nu(z) - (2/\pi)\sin(\nu\pi)Q_\nu(z) \\ &[\mathrm{Im}(z)<0], \\ &P_\nu(-z) = e^{-i\nu\pi}P_\nu(z) - (2/\pi)\sin(\nu\pi)Q_\nu(z) \\ &[\mathrm{Im}(z)>0]. \end{aligned} \tag{4.6}$$

The singular behavior at -1 is given by

$$\begin{aligned} P_\nu(z) &= (1/\pi)\sin(\nu\pi)\ln(z+1)P_\nu(-z) + h(z) \\ &= (1/\pi)\sin(\nu\pi)\ln(z+1)(1 - [\nu(\nu+1)/2](z+1) \\ &\quad + \cdots) + h(z), \quad |z+1|<2, \end{aligned} \tag{4.7}$$

where $h(z)$ is an analytic function and the cut for the logarithm is taken to be $[-\infty,0]$.

For $\nu \in \mathbb{Z}$, P_ν is a Legendre polynomial, in particular $P_0 \equiv 1$, $P_1(x) \equiv x$.

For $\nu \notin \mathbb{Z}$, Q_ν is singular at ∞, -1, 1 and is analytic on $\mathbb{C}\setminus[-\infty,1]$ and its continuation onto the branch cut from above and from below are given by (4.6) for $-1 < \mathrm{Re}(z) < 1$ and for $\mathrm{Re}(z) < -1$ by

$$\begin{aligned} Q_\nu(-z) &= -e^{-i\nu\pi}Q_\nu(z) \quad [\mathrm{Im}(z)<0], \\ Q_\nu(-z) &= -e^{i\nu\pi}Q_\nu(z) \quad [\mathrm{Im}(z)>0]. \end{aligned} \tag{4.8}$$

For $\nu = n \in \mathbb{Z}$, Q_ν is singular at -1 and 1 and analytic on $\mathbb{C}\setminus[-1,1]$. Q_n reads

$$Q_n(z) = \tfrac{1}{2}P_n(z)\ln((z+1)/(z-1)) - W_{n-1}(z), \tag{4.9}$$

where W_{n-1} is a polynomial,

$$W_{n-1}(z) = \sum_{k=1}^{n} \frac{1}{k} P_{k-1}(z)P_{n-k}(z).$$

For $|z| > 1$ Q_ν and P_ν are given by

$$\begin{aligned} Q_\nu(z) &= cz^{-\nu-1}\sum_{n=0}^{\infty} c_n z^{-2n} \\ &= cz^{-\nu-1}F\left(\frac{\nu+2}{2},\frac{\nu+1}{2},\frac{\nu+3}{3},\frac{1}{z^2}\right), \\ P_\nu(z) &= c'z^\nu \sum_{n=0}^{\infty} c'_n z^{-2n} + c''Q_\nu(z), \\ &c,c',c_0,c'_0 \neq 0 \end{aligned} \tag{4.10}$$

where $F(\alpha,\beta,\gamma,z)$ denotes the hypergeometric series.

In (4.5) it is seen that for $m^2r^2 > \frac{9}{4}$, ν is of the form $\nu = -\frac{1}{2} + i\gamma$, $\gamma \in \mathbb{R}$, for $m^2r^2 \leqslant \frac{9}{4}$, ν is real and for $m^2r^2 = 2 - n(n+1)$, ν is an integer and equals n.

Now consider the Dirichlet propagator $G(p,q)$ on a real slice $\mathscr{N}$ isometric to S^4, satisfying

$$(\Delta - m^2)G(p,\cdot)|_q = -\delta(p,q), \tag{4.11}$$

where δ is the Dirac delta function on $\mathscr{N}$ and Δ is the Laplacian on $\mathscr{N}$. Assuming that $G(p,q)$ is invariant under the symmetries of S^4, we can write it as a function f of $\mathfrak{p}(p,\cdot)|_{\mathscr{N}}$. This function must then solve (4.4) everywhere except at 1 (corresponding to $p = q$). It follows that f must be proportional to $\mathscr{F}_1(-z)$ in (4.5) since all other functions are singular at -1 (corresponding to $p = q^*$). The Dirichlet propagator $G(p,q)$ is derived for instance in Dowker and Critchley[1] and reads, for $m^2r^2 \neq 2$:

$$G(p,q) = [1/8\pi r^2 \sin(\nu\pi)]P'_\nu(x)|_{x=-\mathfrak{p}(p,q)},$$

for $m^2r^2 = 2$:

$$G(p,q) = (1/8\pi^2r^2)[1/1 - \mathfrak{p}(p,q)]. \tag{4.12}$$

(The case $m^2r^2 = 2$ can also be obtained directly by taking the limit $\nu \to 0$ in the general case; we will therefore not write it down explicitly any longer.) Equation (4.12) describes the Dirichlet propagator satisfying (4.11) for all values of m^2 except the values $m^2r^2 = 2 - n(n+1)$, $n \in \mathbb{N}$. For these values the polynomials $P'_n(x)|_{x=\mathfrak{p}(p,\cdot)}$ are eigenfunctions of the $\Box$ operator to the eigenvalue m^2, so $(\Box - m^2)$ has no inverse. For all other values of m^2 the kernel of $(\Box - m^2)$ is 0, so the inverse exists, is unique, and is represented by the Dirichlet propagator $G(p,q)$ in (4.12).

Now proceed to the analytic continuation of $G(p,q)$ to other real slices. To this end keep p fixed and continue the function $G(p,\cdot)$. First we note the following proposition.

Proposition 4.1: Any function F defined on an open connected subset U of ${}^{\mathbb{C}}S^n_r$ (we mean in particular that F has no singularities on U) and obtained by analytic continuation of the function $G(p,\cdot)|_{\mathscr{N}}$ can be written as a holomorphic function of $\mathfrak{p}(p,\cdot)$ and solves the homogeneous wave equation $(\Box - m^2)F|_U = 0$.

Proof: If F is obtained from $G(p,\cdot)|_{\mathscr{N}}$ by analytic continuation, there is a sequence $(U_1,F_1),(U_2,F_2),\ldots,(U_m,F_m)$ of open connected subsets U_i of ${}^cS^n_r$ such that $U_i \cap U_{i+1}$ is nonempty and holomorphic functions F_i defined on U_i, such that $F_i|_{U_i\cap U_{i+1}} = F_{i+1}|_{U_i \cap U_{i+1}}$ for all $1 \leqslant i \leqslant m$, $U_1 \cap \mathscr{N}$ is nonempty and $F_1|_{U_1\cap\mathscr{N}} = G(p,\cdot)|_{U_1\cap\mathscr{N}}$ and the last element (U_m,F_m) equals (U,F). On the open set $\widetilde{U}_1$ defined as $\widetilde{U}_1 = \{q\in U_1 | \mathrm{Re}(\mathfrak{p}(p,q)) < -1\}$ (4.12) defines a holomorphic function $\widetilde{G}(p,\cdot)$ satisfying $\widetilde{G}(p,\cdot)|_{\widetilde{U}_1\cap\mathscr{N}} = G(p,\cdot)|_{\widetilde{U}_1\cap\mathscr{N}}$. So F_1 and $\widetilde{G}(p,\cdot)$ coincide on $\widetilde{U}_1$, so the holomorphic function $(\Box - m^2)F_1 = (\Box - m^2)\widetilde{G}(p,\cdot)$ vanishes on $\widetilde{U}_1$ by (4.12), (4.4), and (4.3), and hence, as $\widetilde{U}_1$ is an open subset of U_1, $(\Box - m^2)F_1$ vanishes on U_1. Therefore the holomorphic functions $(\Box - m^2)F_i$ must all vanish since they coincide on the nonempty sets $U_i \cap U_{i+1}$. Analogously one infers that since $F_1|_{\widetilde{U}_1}$ is a holomorphic function of $\mathfrak{p}(p,\cdot)$, and therefore all F_i are holomorphic functions of $\mathfrak{p}(p,\cdot)$. ∎

We can now write down the general form of the analytic continuation.

Proposition 4.2: Let U be an open connected subset of S^+ (S^-) (see the beginning of the section) and F^+ (F^-) be a function on U obtained by analytic continuation from $G(p,\cdot)|_{\mathscr{N}}$. Then F^+ (F^-) can be written in the form

$$F^+(F^-)|_q = [1/8\pi r^2 \sin(\nu\pi)](T^{11}P'_\nu(z) + T^{21}Q'_\nu(z))|_{z=-\mathfrak{p}(p,q)}, \tag{4.13}$$

where the matrix T^{ik} is an element of the discrete subgroup $\mathscr{S}$ of $\mathrm{SL}(2,\mathbb{C})$ generated by the two matrices A and B:

$$A := \begin{pmatrix} 1 & -i\pi \\ 0 & 1 \end{pmatrix}, \quad B := \begin{pmatrix} e^{2i\nu\pi} & 0 \\ -(2i/\pi)\tan(\nu\pi)\sin(2\nu\pi) & e^{-2i\nu\pi} \end{pmatrix}. \tag{4.14}$$

Conversely to every element $T\in\mathscr{S}$ there is a continuation $F^+[T]$ $(F^-[T])$ of $G(p,\cdot)|_{\mathscr{N}}$ to S^+ (S^-) such that $F^+[T]$ $(F^-[T])$ is given by (4.13).

Proof: Consider again a sequence $(U_1,F_1)(U_2,F_2),\ldots,(U_m,F_m)$ as in the proof of Proposition 4.1. Denote by U_i^+, U_i^- the intersections $U_i \cap S^+$, $U_i\cap S^-$, respectively. By Proposition 4.1 $F_i|_{U_i^+}$, $F_i|_{U_i^-}$ can be written as

$$F_i|U_i+ = (a_i^+\mathscr{F}_1(z) + b_i^+\mathscr{F}_2(z))|_{z=-\mathfrak{p}(p,\cdot)},$$
$$F_i|U_i- = (a_i^-\mathscr{F}_1(z) + b_i^-\mathscr{F}_2(z))|_{z=-\mathfrak{p}(p,\cdot)}$$

where $a_i^+, b_i^+, a_i^-, b_i^- \in\mathbb{C}$ and $a_1^+ = a_1^- = 1/8\pi r^2\sin(\nu\pi)$, $b_1^+ = b_1^- = 0$. Then clearly $a_i^+ = a_i^+, b_{i+1}^+ = b_i^+$, $a_{i+1}^- = a_i^-$, $b_{i+1}^- = b_i^-$. On the other hand, from the continuity of F_i one gets by (4.8) and (4.6), $\binom{a_i^+}{b_i^+} = T_i\binom{a_i^-}{b_i^-}$, where T_i is either of the matrices $\mathbf{1}$, A, or B. Expressing now $a_m^+, b_m^+, a_m^-, b_m^-$ successively in terms of $a_1^+, b_1^+, a_1^- b_1^-$ one gets (4.13). Finally one easily convinces oneself that by appropriately choosing the open sets $U_2,\ldots,U_n$ one can realize every element of $\mathscr{S}$. ∎

The analytic continuations $F^+[T]$ to S^+ ($F^-[T]$ to S^-) defined in Proposition 4.2 can be continued to a function ${}^+F^0[T]$ $({}^-F^0[T])$ on the boundary S^0 of S^+ (S^-) which contains all real slices passing through p. These continuations to S^0 may be singular on the light cones of p and p^* [where $\mathfrak{p}(p,q)$ equals $+1$ or -1]. In order to identify the propagators it will be useful to know the singular behavior of these continuations at p expressed in terms of the squared geodesic distance η^2. Using (3.7), (3.2), and (4.1) one finds that $\mathfrak{p}(p,q) = \cos(\eta/r)$. Using (4.7) and (4.6) one finds for the singular behavior of ${}^+F^0[T]$ $({}^-F^0[T])$ at p

$$\mathrm{sing}({}^\pm F^0[T]) = \left(T^{11} + \frac{\pi e^{i\nu\pi}T^{21}}{2\sin(\nu\pi)}\right) \times\left(\frac{1}{4\pi^2\eta^2} + \frac{m^2r^2-2}{16\pi^2r^2}\ln\left(\left|\frac{\eta^2}{2r^2}\right|\right) \pm\frac{i}{4\pi}\delta(\eta^2) \pm \frac{i(m^2r^2-2)}{16\pi r^2}\theta(\eta^2)\right), \tag{4.15}$$

where θ is the step function. Since all real slices passing through p lie in S^0, (4.15) gives the singular behavior of the restriction of ${}^\pm F^0[T]$ to any real slice passing through p.

Now choose a real slice $\mathscr{M}$ passing through p isometric to de Sitter space-time. Comparing (4.15) to the singular behavior of Green's functions in flat space-time one finds that $i\,{}^-F^0[\mathbf{1}]$ has the singular behavior of a Feynman propagator. Setting $G_\mathrm{F}(p,q) := i\,{}^-F^0[\mathbf{1}](q)$ we get [using (4.6)]

$$G_\mathrm{F}(p,q) = \frac{i}{8\pi r^2\sin(\nu\pi)}\left.\begin{Bmatrix} x>-1: & P'_\nu(x) \\ x<-1: & -e^{+i\nu\pi}P'_\nu(-x) + (2/\pi)\sin(\nu\pi)Q'_\nu(-x)\end{Bmatrix}\right|_{x=-\mathfrak{p}(p,q)}. \tag{4.16}$$

$G_\mathrm{F}(p,q)$ is the Feynman propagator of the de Sitter-invariant vacuum state derived for instance in Dowker and Critchley[1] by analytic continuation in $\mathfrak{p}$ or in Tagirov[12] by summing over orthogonal modes. $-i\,{}^+F^0[\mathbf{1}]$ is the complex conjugate of $i\,{}^-F^0[\mathbf{1}]$, so it is the anti-Feynman propagator. Since the future and past lightcone of p in $\mathscr{M}$ are disconnected, we can also choose $-i\,{}^+F^0[\mathbf{1}]$ inside the future lightcone of p and $i\,{}^-F^0[\mathbf{1}]$ inside the past lightcone of p and vice versa which yields the Wightman two-point functions iG^+ and $-iG^-$, respectively. So the situation is similar to the Wick rotation in flat space-time.

Now let $\mathscr{M}$ be a real slice passing through p, isometric to anti-de Sitter space-time (adS). On $\mathscr{M}$ the signature of the metric is $(+\,-\,-\,-)$, so the physical value of the squared mass is $-m^2$. Taking this into account one sees in (4.15) that $-i\,{}^+F^0[\mathbf{1}]$ has the singular behavior of a Feynman propagator. But it follows from (4.10) that this function blows up at timelike infinity of p [which corresponds physically to spacelike infinity because of the reversed signature $(+\,-\,-\,-)$ on adS], so it cannot describe a good propagator. There are however other continuations which have the right singular behavior at p, for instance $-i^+F^0[B]$, as is checked using (4.15) and (4.14). $-i^+F^0[B]$ reads

$$-i^+F^0[B]=\frac{-i}{8\pi r^2\sin(\nu\pi)}\left\{\begin{array}{ll} x>1: & e^{+2i\nu\pi}P'_\nu(x)-(4i/\pi)\sin^2(\nu\pi)Q'_\nu(x)\\ -1<x<1: & P'_\nu(x)-2i\sin(\nu\pi)P'_\nu(-x)\\ x<-1: & -e^{+i\nu\pi}P'_\nu(-x)+(2/\pi)\sin(\nu\pi)Q'_\nu(-x)\end{array}\right\}\Bigg|_{x=-\mathfrak{p}(p,\cdot)}. \tag{4.17}$$

There is a linear combination of $-i^+F^0[\mathbf{1}]$ and $-i^+F^0[B]$ which falls of at physically spacelike infinity, namely ${}^LG(p,\cdot):=ce^{i\nu\pi}(-i^+F^0[\mathbf{1}])-ce^{-i\nu\pi}(-i^+F^0[B])$. If we want ${}^LG$ to have the correct singular behavior, we must choose $c=1/2i\sin(\nu\pi)$. ${}^LG$ then reads [using (4.17), and the complex conjugate of (4.16)]

$${}^LG(p,\cdot)=\frac{-i}{4\pi^2r^2}\left\{\begin{array}{ll} x>1: & e^{-i\nu\pi}Q'_\nu(x)\\ -1<x<1: & [\pi/2\sin(\nu\pi)](e^{-i\nu\pi}P'_\nu(-x)+P'_\nu(x))\\ x<-1: & Q'_\nu(-x)\end{array}\right\}\Bigg|_{x=-\mathfrak{p}(p,\cdot)}, \tag{4.18}$$

which can be written in a more elegant form:

$${}^LG(p,\cdot)=(-i/4\pi^2r^2)Q'_\nu(x+i0)|_{x=\mathfrak{p}(p,\cdot)}. \tag{4.19}$$

Here we note that although we started from $G(p,\cdot)|_{\mathscr{N}}$ which was defined only for $m^2r^2\neq 2-n(n+1)$, $n\in\mathbb{N}$, ${}^LG(p,\cdot)$ is now well-defined for all $m^2\leqslant 0$. So, by construction, ${}^LG$ has the singular behavior of a Feynman propagator, falls off at physically spacelike infinity of p, and is regular in physically spacelike distance to p. It is seen in (4.18) that for $\nu\in\mathbb{N}$, ν even, ${}^LG(p,\cdot)$ is even under the antipodal transformation $q\to q^*$ and for $\nu\in\mathbb{N}$, ν odd, ${}^LG$ is odd under this transformation. For these values of m^2, that is $m^2r^2=2-n(n+1)$, $n\in\mathbb{N}$, ${}^LG$ is the Feynman propagator to an adS-invariant vacuum state as derived and discussed in Avis, Isham, and Storey[6] [up to a sign due to the normalization of ${}^LG$: ${}^LG$ satisfies $(\Box+m^2){}^LG(p,\cdot)=\delta(p,\cdot)+(-1)^n\delta(p^*,\cdot)$ where we have replaced m^2 by its physical value $(-m^2)$].

Now let $\mathscr{M}$ be a connected real slice isometric to the hyperbolic plane H^4 containing p. As in the case above the continuations ${}^\pm F^0[\mathbf{1}]$ and ${}^\pm F^0[B]$ restricted to $\mathscr{M}$ blow up at infinity, but again the linear combination ${}^LG$ restricted to $\mathscr{M}$ falls off at infinity. The operator $-\Box$ on $\mathscr{M}$ is just the Laplacian. The Laplacian on H^4 acts on the space $C_0^\infty(H^4)$ of C^∞ functions with compact support and is essentially self-adjoint, that is, it has a unique self-adjoint extension in the Hilbert space $\mathscr{L}^2(H^4)$ of square integrable functions. This is stated for instance in Wald[3] and is due to a theorem by Gaffney[13] which says that the Laplacian on manifolds with "negligible boundary" is essentially self-adjoint. The class of manifolds with "negligible boundary" contains in particular the complete Riemannian manifolds. It is shown in Wald[3] that for $m^2<0$ the Dirichlet propagator to $\Box-m^2$ can be given in the form of a two-point function. For $m^2r^2<\frac{3}{2}$ $|{}^LG|^2$ falls off at infinity as $\mathfrak{p}^\beta$ with $\beta<-3$ as is seen using (4.10). $\mathfrak{p}^\beta$ is integrable for $\beta<-3$ and not integrable for $\beta\geqslant -3$. So $\mathscr{G}:=-i{}^LG$ describes the Dirichlet propagator satisfying (the factor $-i$ adjusts the singular behavior)

$$(\Box-m^2)\mathscr{G}(p,\cdot)=\delta(p,\cdot)$$

for $m^2<0$:

$$\mathscr{G}(p,\cdot)=(-1/4\pi^2r^2)Q'_\nu(x)|_{x=\mathfrak{p}(p,\cdot)}. \tag{4.20}$$

This shows that in the cases where it exists, the Feynman propagator on adS is obtained by analytic continuation from the Dirichlet propagator on H^4 in the same way as the Feynman propagator on de Sitter is obtained from the Dirichlet propagator on S^4, both are boundary values of a holomorphic function defined on S^+ (H^4,adS), S^-, (S^4, de Sitter), respectively. The question which arises naturally is: does there exist an analytic continuation of say the Dirichlet propagator on the sphere which would yield the Dirichlet propagator on the hyperbolic plane? The following proposition shows that this is not the case, which implies also that there is no analytic continuation from the sphere S^4 to the Feynman propagator on adS or from H^4 to the sphere S^4 or to de Sitter.

Proposition 4.3: Let $m^2r^2<0$, $m^2r^2\neq 2-n(n+1)$, $n\in\mathbb{N}$. There is no analytic continuation of the Dirichlet propagator $G(p,\cdot)|_{\mathscr{N}}$ in (4.12) which yields the Dirichlet propagator on a real slice $\mathscr{M}$ isometric to H^4 passing through p.

Proof: It is seen from (4.20) and using Propositions 4.1 and 4.2 that it is sufficient to show that there is no $T\in\mathscr{T}$ such that $T^{11}=0$.

An element T of $\mathscr{T}$ is of the form $T=A^{n_1}B^{k_1}A^{n_2}B^{k_2}\cdots A^{n_r}B^{k_r}$ for $n_1,k_1,n_2,k_2,\ldots,n_r,k_r\in\mathbb{Z}$. Using induction one easily proves that

$$A^n:=\begin{pmatrix}1 & -in\pi\\ 0 & 1\end{pmatrix},$$
$$B^k:=\begin{pmatrix}e^{2ik\nu\pi} & 0\\ -(2i/\pi)\tan(\nu\pi)\sin(2k\nu\pi) & e^{-2ik\nu\pi}\end{pmatrix}. \tag{4.21}$$

Using (4.21) we find that

$$B^kA^n+A^{-n}B^{-k}=F(n,k)\mathbf{1}, \tag{4.22}$$

with

$$F(n,k):=2\cos(2k\nu\pi)-2n\tan(\nu\pi)\sin(2k\nu\pi).$$

Note that for $m^2r^2<0$, ν and hence $F(n,k)$ is real. Using the above relation we can rewrite the element T as follows:

$$\begin{aligned}T&=A^{n_1}B^{k_1}A^{n_2}B^{k_2}\cdots A^{n_r}B^{k_r}\\ &=-A^{n_1-n_2}B^{k_2-k_1}A^{n_3}B^{k_3}\cdots A^{n_r}B^{k_r}\\ &\quad+F(n_2,k_1)A^{n_1}B^{k_2}A^{n_3}B^{k_3}\cdots A^{n_r}B^{k_r}.\end{aligned} \tag{4.23}$$

Using this relation repeatedly we can express T in the form:

$$T=\sum_{n,k} l(n,k)A^{n}B^{k}, \tag{4.24}$$

where the coefficients $l(n,k)$ depend on $n_1,k_1,n_2,k_2,\ldots,n_r,k_r$, only finitely many of them are different from zero and they all are real [since $F(n,k)$ is real].

Now assume that there is a $T\in\mathscr{T}$ with $T^{11}=0$. Express T in the form (4.24). Since $T^{11}=0$, its real and imaginary part must vanish. Since the $l(n,k)$ are real, $\mathrm{Im}(T^{11})$ is given by [using (4.21) and (4.24)]

$$0=\mathrm{Im}(T^{11})=\sum_{n,k} l(n,k)\sin(2k\nu\pi). \tag{4.25}$$

On the other hand T^{21} is given by [using (4.21) and (4.24)]

$$T^{21}=-\frac{2i}{\pi}\tan(\nu\pi)\sum_{n,k} l(n,k)\sin(2k\nu\pi). \tag{4.26}$$

So by (4.25) we get $T^{21}=0$. But then $\det(T)=0$ which contradicts $\det(T)=1\ \forall\ T\in\mathscr{T}$, so there is no $T\in\mathscr{T}$ with $T^{11}=0$. ■

ACKNOWLEDGMENTS

I am greatly indebted to Petr Hajicek who supervised, guided, and encouraged this work. I also wish to thank Alan Held and Bruce Jensen for carefully reading the manuscript and for helpful discussions.

[1] J. S. Dowker and R. Critchley, Phys. Rev. D **13**, 224 (1976).
[2] P. Candelas and D. J. Raine, J. Math. Phys. **17**, 2101 (1976).
[3] R. Wald, Commun. Math. Phys. **70**, 221 (1979).
[4] S. A. Fulling and S. N. M. Ruijsenaars, Phys. Rep. **152**, 135 (1987).
[5] N. Woodhouse, Int. J. Theor. Phys. **16**, 671 (1977).
[6] S. J. Avis, C. J. Isham, and D. Storey, Phys. Rev. D **18**, 3565 (1978).
[7] S. Kobayashi and K. Nomizu, *Foundations of Differential Geometry* (Wiley, New York, 1969), Vol. II.
[8] A. Meister, Lizentiatsarbeit 1989, Institut für Theoretische Physik, Sidlerstrasse 5, CH-3012 Bern.
[9] S. Helgason, *Differential Geometry, Lie Groups, and Symmetric Spaces* (Academic, New York, 1978).
[10] I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1980).
[11] W. I. Smirnov, *Lehrgang der höheren Mathematik, III,2* (Deutscher Verlag der Wissenschaften, Berlin, 1955).
[12] E. A. Tagirov, Ann. Phys. (NY) **76**, 561 (1973).
[13] M. P. Gaffney, Proc. Natl. Acad. Sci. **37**, 48 (1951); Ann. Math. **60**, 140 (1954).

A simplified variational characterization of Schrödinger processes

A. Wakolbinger
Institut für Mathematik, Johannes Kepler Universität, A-4040 Linz, Austria

(Received 24 January 1989; accepted for publication 26 July 1989)

The variational problem $E\left[\frac{1}{2}\int_0^T(\beta_t - A(t,X_t))^2\,dt - \int_0^T c(t,X_t)\,dt\right] = \min$, where β_t is the drift process of a diffusion process with unit diffusion coefficient and given initial and final distributions, A is a given vector field, and c is a given scalar field, is considered. It is shown that the solution is given by a certain Markovian diffusion process, which (in the case $A = 0$, $c = 0$) first was investigated by Schrödinger (Sitzungsber. Preuss. Akad. Wiss. Phys. Math. Kl. **1931**, 144).

I. INTRODUCTION

In 1931/32, Schrödinger[1,2] treated the following problem: A large number N of particles in Euclidean space $\mathbb{R}^d$ with initial distribution (approximately) $\rho_0(x)dx$ performs standard Brownian motion, i.e., the particles move independently of each other according to the probability transition density

$$p_W(s,x;t,y):$$
$$= \exp(-(x-y)^2/2(t-s))/(2\pi(t-s))^{d/2}.$$

Assume that at some later time T the particle distribution is observed to be (approximately) $\rho_T(z)dz$, where $\rho_T(z)$ "largely deviates from the law of large numbers" in the sense that it is different from

$$\int \rho_0(x)p_W(0,x;T,z)dx:$$

then which way that leads from ρ_0 to ρ_T is still the most probable? Schrödinger showed that the most probable distribution densities at intermediary time points are of the product form

$$\rho_t(y) = \hat{\varphi}(t,y)\varphi(t,y), \tag{1}$$

where

$$\hat{\varphi}(t,y) = \int \hat{\varphi}(x)p_W(0,x;t,y)dx,$$
$$\varphi(t,y) = \int \varphi(z)p_W(t,y;T,z)dz, \tag{2}$$

and the pair $\hat{\varphi}(x),\varphi(z)$ solves the integral equations

$$\rho_0(x) = \hat{\varphi}(x)\int \varphi(z)p_W(0,x;T,z)dz,$$
$$\rho_T(z) = \varphi(z)\int \hat{\varphi}(x)p_W(0,x;T,z)dx. \tag{3}$$

The crucial step in Schrödinger's derivation was to determine the most probable joint distribution $\nu(dx,dz)$ of initial and final positions, which turned out to be

$$\nu(dx,dz) = \hat{\varphi}(x)p_W(0,x;T,z)\varphi(z)dx\,dz. \tag{4}$$

Since the joint distribution densities at times $t_1,\dots,t_n$, conditioned upon the initial and final positions x and z are given by

$$[p_W(0,x;t_1,y_1)\cdots p_W(t_n,y_n;T,z)]/p_W(0,x;T,z),$$

the unconditioned joint distribution densities at times $0<t_1<\cdots<t_n<T$ are

$$\hat{\varphi}(x)p_W(0,x;t_1,y_1)\cdots p_W(t_n,y_n;T,z)\varphi(z). \tag{5}$$

Hence the distribution Q of the corresponding stochastic process is absolutely continuous with respect to the stationary Wiener measure, with Radon–Nikodym derivative $\hat{\varphi}(X_0)\varphi(X_T)$. Here and below, $X = (X_t)$ stands for an element of the "path space" Ω, the space of continuous functions from $[0,T]$ into $\mathbb{R}^d$.

Föllmer[3] gives a rigorous treatment of Schrödinger's problem, proving the above-mentioned results by methods from the theory of large deviations. Using his approach, one can also show (cf. the result in Sec. II for the special case $A=0$, $c=0$) that Schrödinger's process satisfies a simple least action principle. The fact that Schrödinger's processes should be accessible to a variational characterization was already conjectured by Bernstein[4] in 1932, cf. remarks in the paper of Zambrini.[5]

Now let p_W be replaced by a more general transition density p, say, the fundamental solution of the differential equation

$$\left(\frac{\partial}{\partial s} + \frac{1}{2}\frac{\partial^2}{\partial x^2} + A\cdot\frac{\partial}{\partial x} + c\right)p = 0,$$
$$\lim_{s\uparrow t} p(s,\cdot,t,y) = \delta_y, \tag{6}$$

for a given vector field $A(s,x)$ and a given scalar field $c(s,x)$. If the field c does not vanish, p is no longer a probability transition density, but rather corresponds to a stochastic process with creation and annihilation.[6] For any given initial density ρ_0 and final density ρ_T it is, however, still possible (a) to construct a Markovian diffusion process (with conservation of mass) via the system of integral equations (2) and the joint distribution densities (5) (see Jamison,[7,8] Zambrini,[5,9,10] and Nagasawa[11]); and (b) to formulate a least action principle generalizing the above-mentioned one for the case $A=0$, $c=0$.

There have been various attempts to relate these two approaches: Zambrini,[5,9,10] who revived Schrödinger's ideas for contemporary physics and gave connections to Nelson's[12] stochastic mechanics, shows that among a certain class of "neighboring diffusions" with drift processes (β_t), the above-mentioned Markovian diffusion process (which we continue to call Schrödinger's process) minimizes the functionals

$$E_{(s,x)}[L_{s,T} - \ln\varphi(T,X_T)], \quad 0\leqslant s\leqslant T, \quad x\in\mathbb{R}^d,$$

 0022-2488/89/122943-04$02.50

where the "stochastic Lagrangian" is given by

$$L_{s,T}:=\int_s^T \frac{1}{2}(\beta_t - A(t,X_t))^2\,dt - \int_s^T c(t,X_t)dt. \quad (7)$$

(In fact, Zambrini considers the case $A=0$ and gives a reverse time formulation.)

Nagasawa[13] gives a similar variational characterization of what he calls a Schrödinger process prescribed by a pair of functions $\varphi(t,y)$, $\hat{\varphi}(t,y)$: among all diffusion processes with unit diffusion coefficient and square integrable drift processes, it minimizes the functional

$$E\left[L_{0,T}^{\text{symm}} + \frac{1}{2}\log\left(\frac{\varphi}{\hat{\varphi}}\right)(0,X_0) - \frac{1}{2}\log\left(\frac{\varphi}{\hat{\varphi}}\right)(T,X_T)\right],$$

where $L_{0,T}^{\text{symm}}$ is a certain time symmetric stochastic Lagrangian first considered by Yasue.[14] For related results, see Blanchard *et al.*[15] and also Zhao.[16]

We will show that, among all diffusion processes with unit diffusion coefficient, square integrable drift process, and prescribed initial and final densities ρ_0 and ρ_T, Schrödinger's process is the one that minimizes the functional $E[L_{0,T}]$, where the stochastic Lagrangian $L_{0,T}$ is defined in (7). This reveals, in particular, that one can dispose in the objective functional of initial and final valuations invoking φ and $\hat{\varphi}$, if one restricts the variational problem to processes with prescribed initial and final distributions—a restriction on the line of Schrödinger's original program. The methods used for the proof are purely probabilistic; they are close to the approach of Föllmer[3] and different from those of Zambrini[5,10] and Nagasawa.[13]

II. FORMULATION AND PROOF OF THE RESULT

Let D be the set of distributions of diffusion processes with unit diffusion coefficient and square integrable drift process, i.e., D consists of the probability measures Q on path space $\Omega:=C([0,T],\mathbb{R}^d)$ for which there exists a family of measurable mappings β_t: $\Omega\to\mathbb{R}$ with β_t depending only on X_s, $0\leqslant s\leqslant t$, such that

$$E_Q\left[\int_0^T \beta_t^2\,dt\right]<\infty$$

and

$$W_t:=X_t-\int_0^t \beta_s(X)ds$$

is a standard Wiener process defined on (Ω,Q).

For any given probability densities ρ_0, ρ_T on $\mathbb{R}^d$, let $D(\rho_0,\rho_T)$ denote the set of all distributions in D having initial density ρ_0 and final density ρ_T, i.e.,

$$D(\rho_0,\rho_T):=\{Q\in D:\quad Q[X_0\in dx]=\rho_0(x)dx,$$
$$Q[X_T\in dz]=\rho_T(z)dz\}.$$

Let A: $[0,T]\times\mathbb{R}^d\to\mathbb{R}^d$ be a given vector field, and c: $[0,T]\times\mathbb{R}^d\to\mathbb{R}$ be a given scalar field. To avoid technical difficulties, we will assume that A is continuous and bounded, and c is continuous and bounded from above (though presumably the following theorem should hold for a larger class of fields A and c).

Theorem: Let ρ_0 and ρ_T be probability densities on $\mathbb{R}^d$, with $\rho_0>0$. Suppose that $D(\rho_0,\rho_T)$ is nonvoid. Then the unique solution of the variational problem

$$I:=E_Q\left[\frac{1}{2}\int_0^T(\beta_t - A(t,X_t))^2\,dt - \int_0^T c(t,X_t)dt\right]=\min,$$
$$Q\in D(\rho_0,\rho_T), \quad (8)$$

is given by $Q^*=\hat{\varphi}(X_0)\varphi(X_T)P_1$, where P_1 is the measure on Ω starting with Lebesgue measure at time zero and having transition density p given by (6), and $(\varphi,\hat{\varphi})$ is the unique solution of the system of integral equations

$$\rho_0(x)=\hat{\varphi}(x)\int p(0,x;T,z)\varphi(z)dz,$$
$$\rho_T(z)=\varphi(z)\int p(0,x;T,z)\hat{\varphi}(x)dx. \quad (9)$$

Moreover, Q^* is the distribution of a Markovian diffusion with drift process $\beta_t(X)=A(t,X_t)+\nabla\ln\varphi(t,X_t)$, where

$$\varphi(t,y):=\int p(t,y;T,z)\varphi(z)dz. \quad (10)$$

Proof: A combination of the formulas of Girsanov and Feynman–Kac[17] shows that P_1 has Radon–Nikodym derivative

$$N:=\exp\left(\int_0^T A(t,X_t)dX_t - \frac{1}{2}\int_0^T A^2(t,X_t)dt + \int_0^T c(t,X_t)dt\right),$$

with respect to stationary Wiener measure P_1^{W} on Ω. Writing $P^{\mathrm{W}}:=\rho_0(X_0)P_1^{\mathrm{W}}$, $P:=\rho_0(X_0)P_1$, one has

$$P=N\cdot P^{\mathrm{W}}.$$

Any Q in $D(\rho_0,\rho_T)$ with drift process (β_t) has, by Girsanov's formula, a Radon–Nikodym derivative

$$\frac{dQ}{dP^{\mathrm{W}}}(X)=\exp\left(\int_0^T\beta_t(X)dX_t - \frac{1}{2}\int_0^T\beta_t^2(X)dt\right),$$
$$Q \text{ almost surely.} \quad (11)$$

Since the martingale expectations

$$E_Q\left[\int_0^T \beta_t\,dW_t\right]$$

and

$$E_Q\left[\int_0^T A(t,X_t)dW_t\right]$$

vanish, one can rewrite the functional I as

$$I=E_Q\left[\int_0^T\beta_t(X)dW_t - \int_0^T A(t,X_t)dW_t + \frac{1}{2}\int_0^T\beta_t^2(X)dt - \int_0^T A(t,X_t)\beta_t(X)dt + \frac{1}{2}\int_0^T A^2(t,X_t)dt - \int_0^T c(t,X_t)dt\right]$$
$$=E_Q\left[\int_0^T\beta_t(X)dX_t - \frac{1}{2}\int_0^T\beta_t^2(X)dt - \int_0^T A(t,X_t)dX_t + \frac{1}{2}\int_0^T A^2(t,X_t)dt - \int_0^T c(t,X_t)dt\right]$$
$$=E_Q\left[\ln\frac{dQ}{dP_W} - \ln N\right]=E_Q\left[\ln\frac{dQ}{dP}\right].$$

[If the finite measure P would have total mass 1 (this is the case, e.g., if $c\equiv 0$), the latter expression would just be the relative entropy of Q with respect to P.] We now consider the disintegrations of P and Q with respect to initial and final positions,

$$P_x^z := P[\,\cdot\,|X_0=x,\quad X_T=z],$$
$$Q_x^z := Q[\,\cdot\,|X_0=x,\quad X_T=z],$$

and the projections of P and Q under initial and final positions,

$$\mu := P[(X_0,X_T)\in(\cdot)],$$
$$\nu := Q[(X_0,X_T)\in(\cdot)].$$

Note that P_x^z and Q_x^z are probability measures on Ω, and

$$P=\int P_x^z(\cdot)\mu(dx,dz),\quad Q=\int Q_x^z(\cdot)\nu(dx,dz).$$

In view of the multiplication formula

$$\frac{dQ}{dP}(X)=\frac{d\nu}{d\mu}(X_0,X_T)\,\frac{dQ_{X_0}^{X_T}}{dP_{X_0}^{X_T}}(X),\quad Q \text{ almost surely,} \tag{12}$$

one has

$$I=E_Q\left[\ln\frac{dQ}{dP}\right]=E_Q\left[\ln\frac{d\nu}{d\mu}(X_0,X_T)\right] + E_Q\left[\ln\frac{dQ_{X_0}^{X_T}}{dP_{X_0}^{X_T}}(X)\right]=\int\left(\ln\frac{d\nu}{d\mu}\right)d\nu + \int\int\int\left(\ln\frac{dQ_x^z}{dP_x^z}\right)dQ_x^z\,\nu(dx,dz).$$

Because of Jensen's inequality, the second summand on the right-hand side becomes minimal (namely zero) if and only if

$$Q_x^z=P_x^z,\quad \text{for } \nu \text{ almost all } (x,z). \tag{13}$$

Thus the problem (8) boils down to

$$\int\left(\ln\frac{d\nu}{d\mu}\right)d\nu=\min \tag{14}$$

under the side conditions

$$\nu(dx\times\mathbb{R}^d)=\rho_0(x)dx,\quad \nu(\mathbb{R}^d\times dz)=\rho_T(z)dz. \tag{15}$$

Denoting the total mass of μ (which equals that of P and hence is finite) by k and writing $m:=\mu/k$, we observe that

$$\ln\frac{d\nu}{d\mu}=-\ln k+\ln\frac{d\nu}{dm}.$$

Hence (14) is equivalent to

$$\int\left(\ln\frac{d\nu}{dm}\right)d\nu=\min. \tag{16}$$

A result of Föllmer[3] shows that the minimization problem [(16) and (15)] has a unique solution ν^*, which is determined by the following conditions:

$$\nu^*(dx,dz)=f(x)g(z)m(dx,dz), \tag{17}$$
$$\log f(x)\in L^1(\rho_0(x)dx),\quad \log g(z)\in L^1(\rho_T(z)dz), \tag{18}$$
$$\rho_0(x)=f(x)\int k^{-1}\rho_0(x)p(0,x;T,z)g(z)dz,$$
$$\rho_T(z)=g(z)\int k^{-1}\rho_0(x)p(0,x;T,z)f(x)dx. \tag{19}$$

Writing $\hat\varphi(x):=k^{-1}f(x)\rho_0(x)$, $\varphi(z):=g(z)$, we thus note that (14) and (15) have a unique solution ν^*, which is determined by

$$\nu^*(dx,dz)=\hat\varphi(x)\varphi(z)p(0,x;T,z)dx\,dz, \tag{20}$$

together with (18) and the system (9). From (12), (13), and (20) one has, for the distribution Q^*, which minimizes the functional I,

$$\frac{dQ^*}{dP}(X)=\frac{d\nu^*}{d\mu}(X_0,X_T)=\frac{\hat\varphi(X_0)\varphi(X_T)p(0,X_0;T,X_T)}{\rho_0(X_0)p(0,X_0;T,X_T)}; \tag{21}$$

hence

$$Q^*=\hat\varphi(X_0)\varphi(X_T)[1/\rho_0(X_0)]P=\hat\varphi(X_0)\varphi(X_T)P_1.$$

With $\varphi(t,y)$ defined by (10), one infers from (21) and the first equation of (9) that

$$\frac{dQ^*}{dP}(X)=\frac{\hat\varphi(X_0)\varphi(X_T)}{\hat\varphi(X_0)\varphi(0,X_0)},$$

and hence one obtains, applying Itô's formula to $\nabla\ln\varphi(t,X_t)$,

$$\begin{aligned}\frac{dQ^*}{dP_W}(X)&=\frac{dQ^*}{dP}(X)\frac{dP}{dP_W}(X)\\
&=\exp(\ln\varphi(T,X_T)-\ln\varphi(0,X_0))\exp\left(\int_0^T A(t,X_t)dX_t-\frac{1}{2}\int_0^T A^2(t,X_t)dt+\int_0^T c(t,X_t)dt\right)\\
&=\exp\left(\int_0^T\nabla\ln\varphi(t,X_t)dX_t-\frac{1}{2}\int_0^T(\nabla\ln\varphi)^2(t,X_t)dt\right.\\
&\quad\left.+\int_0^T\left[\frac{1}{\varphi}\left(\frac{1}{2}\Delta\varphi+\frac{\partial\varphi}{\partial t}\right)\right](t,X_t)dt\right)\exp\left(\int_0^T A(t,X_t)dX_t-\frac{1}{2}\int_0^T A^2(t,X_t)dt+\int_0^T c(t,X_t)dt\right)\\
&=\exp\left(\int_0^T(A+\nabla\ln\varphi)(t,X_t)-\frac{1}{2}\int_0^T(A+\nabla\ln\varphi)^2(t,X_t)dt\right).\end{aligned}$$

This reveals, by Girsanov's formula, that Q^* belongs to D, with drift process

$$\beta_t(X)=A(t,X_t)+\nabla\ln\varphi(t,X_t).$$

The Markov property of Q^* now is an easy consequence of the fact that β_t depends only on the present state X_t of the process X. Thus the proof of the theorem is complete.

ACKNOWLEDGMENTS

Most of the studies that lead to the completion of this paper were carried through during a research visit at the University of Passau, FRG, in July 1988. I would like to thank G. Ritter and G. Leha for their hospitality and for stimulating discussions. Also, I am grateful to M. Nagasawa, J. C. Zambrini, and W. Stummer for valuable hints.

[1]E. Schrödinger, Sitzungber. Preuss. Akad. Wiss. Phys. Math. Kl. **1931**, 144.

[2]E. Schrödinger, Ann. Inst. H. Poincáre **2**, 269 (1932).

[3]H. Föllmer, "Random fields and diffusion processes," in *Ecole d'Eté de Saint Flour XV-XVII (1985–1987), Lecture Notes in Mathematics,* Vol. 1362 (Springer, Berlin, 1988).

[4]S. Bernstein, Verh. Int. Math. Kongress, Zürich, **I**, 288 (1932).

[5]J. C. Zambrini, J. Math. Phys. **27**, 2307 (1986).

[6]M. Nagasawa, Z. Wahrscheinlichkeitstheorie verw. Gebiete **14**, 49 (1969).

[7]B. Jamison, Z. Wahrscheinlichkeitstheorie verw. Gebiete **30**, 65 (1974).

[8]B. Jamison, Z. Wahrscheinlichkeitstheorie verw. Gebiete **32**, 323 (1975).

[9]J. C. Zambrini, Phys. Rev. A **33**, 1532 (1986).

[10]J. C. Zambrini, "New probabilistic approach to the classical heat equation," in *Stochastic Mechanics and Stochastic Processes,* edited by A. Truman and I. M. Davies, *Lecture Notes in Mathematics,* Vol. 1325 (Springer, Berlin, 1988), p. 205.

[11]M. Nagasawa, Probab. Th. Rel. Fields **82**, 109 (1989).

[12]E. Nelson, *Quantum Fluctuations* (Princeton, U.P., Princeton, NJ, 1985).

[13]M. Nagasawa, "Stochastic variational principle of Schrödinger processes," Preprint, Universität Zürich, 1988.

[14]K. Yasue, J. Funct. Anal. **41**, 327 (1981).

[15]Ph. Blanchard, Ph. Combe, and W. Zheng, *Mathematical and Physical Aspects of Stochastic Mechanics, Lecture Notes in Physics,* Vol. 281 (Springer, Berlin, 1897).

[16]Z. Zhao, Stochastics **18**, 1 (1986).

[17]I. Karatzas and S. E. Shreve, *Brownian Motion and Stochastic Calculus* (Springer, New York, 1988).

Disproof of a conjectured inequality arising in the theory of magnetic flux diffusion

David L. Book
Laboratory for Computational Physics, Code 4405, Naval Research Laboratory, Washington, DC 20375
Todd A. Brun[a)]
Science Applications International Corporation, Mc Lean Virginia 22102

(Received 22 November 1988; accepted for publication 9 August 1989)

The conjecture that the magnetic field energy introduced into an imperfectly conducting medium by diffusion is less than the thermal energy produced by the associated Joule heating is shown to imply the positivity of a certain integral operator with an indefinite kernel. Counterexamples are found by means of a variational calculation.

I. INTRODUCTION

In a number of experiments carried out during the past several decades, multimegagauss magnetic fields have been generated by rapid compression of magnetic flux confined within a metal chamber or liner (see, e.g., the review by Herlach[1]. The limit to the peak fields attained in these attempts is imposed in practice either by the compressibility of the liner or by losses resulting from finite electrical conductivity. In the latter case flux diffuses through the collapsing walls at a rate equal to the speed with which the liner moves, and no further increase in the field strength results.

Attempts to predict the maximum obtainable fields[1,2] begin by treating the simplest problem, that of magnetic diffusion into a static homogeneous medium. When a magnetic field is produced in a cavity or dielectric medium adjoining a region of large but finite conductivity, it immediately begins to diffuse into the latter. This process is described by Ampére's law, Faraday's law, Ohm's law, and a prescription for the time history of the field strength **B** outside the conducting region. If we assume that the permeability μ and conductivity σ are constant and that the displacement current and the motion of the medium can be neglected, we can determine the evolution of **B** within the conducting region in planar geometry by solving

$$\frac{\partial B}{\partial t}=\kappa\frac{\partial^2 B}{\partial x^2}, \qquad (1)$$

where κ is the (constant) magnetic diffusion coefficient or diffusivity, given in cgs units by

$$\kappa=c^2/4\pi\sigma\mu. \qquad (2)$$

The magnetic energy density is $B^2/8\pi\mu$, so the total magnetic energy per unit area introduced into the conducting region as a result of diffusion is

$$W_M(t)=\frac{1}{8\pi\mu}\int dx\, B^2(x,t). \qquad (3)$$

The diffusion induces currents in the conducting medium. In terms of the current density $\mathbf{J}=(c/4\pi\mu)\nabla\times\mathbf{B}$ the rate of Joule heating is J^2/σ, so the thermal energy per unit area evolved by Joule heating in the course of the process is

$$W_J(t)=\frac{\kappa}{4\pi\mu}\int_0^t dt'\int dx[\nabla\times\mathbf{B}(x,t')]^2. \qquad (4)$$

In Eqs. (3) and (4) the spatial integrals are carried out over the entire conducting region.

Assuming that the conducting region occupies the half-space $0<x<\infty$ (see Fig. 1) and that B vanishes within this region at $t=0$, we can write the solution of Eq. (1) in the familiar form[3]

$$B(x,t)=\frac{x}{2(\pi\kappa)^{1/2}}\int_0^t\frac{dt'\,B_0(t')}{(t-t')^{3/2}}\exp\left[-\frac{x^2}{4\kappa(t-t')}\right], \qquad (5)$$

for $x>0$, $t>0$. Here $B_0(t)$, the boundary or "driving" magnetic field, is the value of B applied at the interface. Knoepfel[2] has evaluated Eq. (5) for a number of different forms of $B_0(t)$ and used the resulting solution $B(x,t)$ to determine W_M and W_J. In every case these quantities are found to satisfy the inequality

$$W_M(t)\leqslant W_J(t). \qquad (6)$$

For example, when $B_0(t)\propto t^n$ one finds $W_M/W_J=0.414$, 0.811, 0.891, 0.924, 0.942, and 0.953 for $n=0,1,2,3,4$, and 5, respectively. This pattern has given rise to the conjecture that Eq. (6) is a general property of magnetic flux diffusion.

There is no obvious way to derive the conjecture (6),

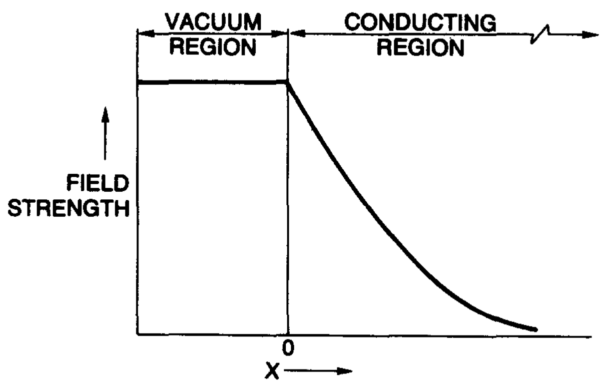


FIG. 1. Typical plot of magnetic field strength $B(x,t)$ vs x, superposed on sketch indicating locations of vacuum and conducting regions in slab geometry.

[a)] Current address: Physics Department, California Institute of Technology, Pasadena, CA 91125.

nor does there appear to be any thermodynamic reason why it should hold. In this paper we show that in fact it is sometimes false. To do this, we proceed as follows. In the next section we express $\Delta W = W_J - W_M$ for an arbitrary $B_0(t)$ as an integral with respect to its two arguments of the product of a symmetric real kernel $K(\tau,\tau')$ and $f(\tau)f(\tau')$, where $f(\tau)$ is a normalized form of the derivative of $B_0(t)$—in other words, as a quadratic form on a real Hilbert space. Then in Sec. III we show by means of a variational calculation that this form can become negative, i.e., that the operator K has negative eigenvalues. In Sec. IV we discuss these results and summarize our conclusions.

II. DERIVATION OF THE INTEGRAL OPERATOR

To find an expression for the magnetic energy W_M in terms of B_0, we substitute Eq. (1) in the one-dimensional form of Eq. (3). Interchanging orders of integration and integrating over x yields

$$W_M = \frac{\kappa^{1/2}}{16\pi^{3/2}\mu} \int_0^t dt' \int_0^t dt'' \frac{B_0(t')B_0(t'')}{(2t - t' - t'')^{3/2}}. \tag{7}$$

An equivalent expression for the Joule heat W_J is most easily found if we first derive a relation for the value of $\partial B_0(x,t)/\partial t$ at $x = 0$. This is done by differentiating Eq. (1) with respect to x and integrating by parts:

$$\begin{aligned}\frac{\partial B(x,t)}{\partial x} &= \frac{1}{2(\pi\kappa)^{1/2}} \int_0^t \frac{dt'\, B_0(t')}{(t-t')^{3/2}}\left[1 - \frac{x^2}{2\kappa(t-t')}\right]\exp\left[-\frac{x^2}{4\kappa(t-t')}\right] \\ &= \frac{1}{2(\pi\kappa)^{1/2}} \int_0^t \frac{dt'\, B_0(t')}{(t-t')^{3/2}}\left\{\exp\left[-\frac{x^2}{4\kappa(t-t')}\right] + 2(t-t')\frac{\partial}{\partial t'}\exp\left[-\frac{x^2}{4\kappa(t-t')}\right]\right\} \\ &= -\frac{B_0(0)}{(\pi\kappa t)^{1/2}}\exp\left[-\frac{x^2}{4\kappa t}\right] - \frac{1}{(\pi\kappa)^{1/2}}\int_0^t \frac{dt'}{(t-t')^{1/2}}\frac{dB_0(t')}{dt'}\exp\left[-\frac{x^2}{4\kappa(t-t')}\right].\end{aligned} \tag{8}$$

The desired expression is obtained in the limit $x \to 0$.

Making use of this result, we integrate by parts with respect to x in the one-dimensional form of Eq. (4) and substitute using Eq. (1):

$$\begin{aligned}W_J &= \frac{\kappa}{4\pi\mu}\int_0^t dt' \int_0^\infty dx \left[\frac{\partial B(x,t')}{\partial x}\right]^2 \\ &= \frac{\kappa}{4\pi\mu}\int_0^t dt'\left[-B(0,t')\frac{\partial B(0,t')}{\partial x} - \frac{1}{\kappa}\int_0^\infty dx\, B(x,t')\frac{\partial B(x,t')}{\partial t'}\right] \\ &= \frac{\kappa^{1/2}}{4\pi^{3/2}\mu}\int_0^t dt'\, B_0(t')\left[\frac{B_0(0)}{t'^{1/2}} + \int_0^{t'} \frac{dt''}{(t''-t'')^{1/2}}\frac{dB_0(t'')}{dt''}\right] - W_M.\end{aligned} \tag{9}$$

We now subtract Eq. (7) from Eq. (9). After several further integrations by parts and interchanges of orders of integrations, the energy difference ΔW can be written in the form

$$\begin{aligned}\Delta W = \frac{\kappa^{1/2}}{2\pi^{3/2}\mu}&\int_0^t dt' \frac{dB_0(t')}{dt'}\int_0^t dt'' \frac{dB_0(t'')}{dt''} \\ &\times\left[(2t-t'-t'')^{1/2} - \frac{1}{2}|t'-t''|^{1/2} - (t-t')^{1/2}\right] \\ &+ B_0(0)\left\{\int_0^t dt' \frac{dB_0(t')}{dt'}\,[2(2t-t')^{1/2} \right. \\ &\left. - t'^{1/2} - (t-t')^{1/2}] - B_0(t)t^{1/2} + B_0(0)(2t)^{1/2}\right\}.\end{aligned} \tag{10}$$

We take $B_0(0) = 0$. This entails no loss of generality, because $B(x,t)$ can always be approximated to any desired accuracy in the case $B_0(0) \neq 0$ by letting the driving field rise from 0 to a finite value in a time short compared with the diffusion time $4x^2/\kappa$, making a negligible contribution to both W_M and W_J at distances greater than x. As a result, we have a relation for ΔW of the form

$$\Delta W = \int_0^1 d\tau' \int_0^1 d\tau''\, K(\tau',\tau'')f(\tau')f(\tau''), \tag{11}$$

where $\tau' = 1 - t'/t$ and $\tau'' = 1 - t''/t$, the kernel is

$$K(\tau',\tau'') = 2(\tau' + \tau'')^{1/2} - |\tau' - \tau''|^{1/2} - \tau'^{1/2} - \tau''^{1/2}$$

in symmetrized form, and we have written

$$f(\tau') = \frac{\kappa^{1/4}t^{5/4}}{2\pi^{3/4}\mu^{1/2}}\frac{dB_0(t-t')}{dt'} = -\frac{(\kappa t)^{1/4}}{2\pi^{3/4}\mu^{1/2}}\frac{dB_0(t\tau')}{d\tau'}. \tag{12}$$

The only restriction on the real function $f(\tau)$ is that it be integrable.

III. VARIATIONAL CALCULATION

Equation (6) would be valid if the diagonal elements of K, which are positive, were sufficiently dominant. It would obviously be false if $K(\tau',\tau'')^2 > K(\tau',\tau')K(\tau'',\tau'')$, held for some choice of τ' and τ''. But direct calculation (see Fig. 2) shows that this is not the case.

To resolve the question, we adopt a computational approach. We seek to minimize the value of the functional

$$\lambda = \frac{\int_0^1 d\tau \int_0^1 d\tau'\, K(\tau,\tau')f(\tau)f(\tau')}{\int_0^1 d\tau [f(\tau)]^2}. \tag{13}$$

This is equivalent to finding the lowest eigenvalue of the Euler–Lagrange equation, which determines λ and the corresponding eigenfunctions f_λ for the operator K.

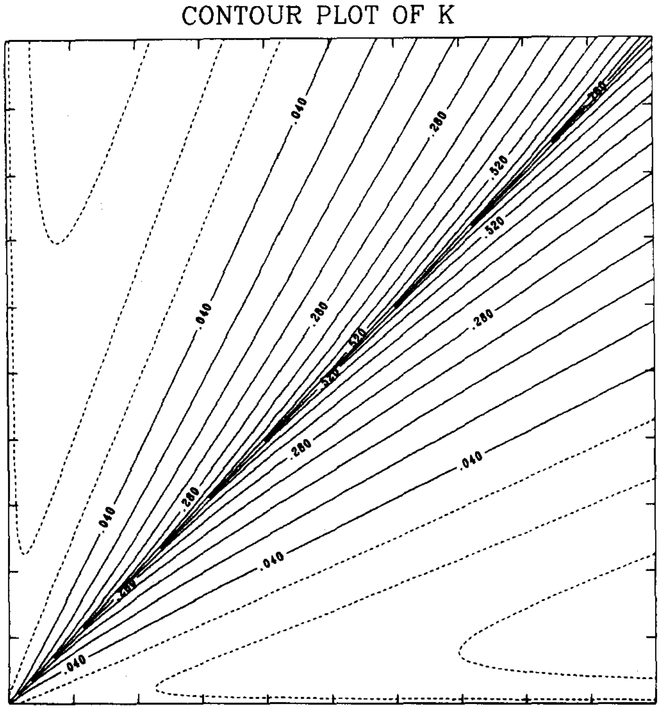


FIG. 2. Contours of constant $K(\tau',\tau'')$ for arguments between 0 and 1. Shown are levels from -0.14 to 0.76 at intervals of 0.06. Broken lines represent negative levels.

We replace f with a piecewise constant approximation:

$$f(\tau) = \sum_{n=0}^{N} c_n \theta(\tau_{n+1} - \tau)\theta(\tau - \tau_n), \tag{14}$$

where θ is the Heaviside step function, $\tau_n = n/(N+1)$, and the $\{c_n\}$ are constants. Substituting Eq. (14) in Eq. (13), integrating over τ' and τ'', differentiating with respect to c_m, and equating the result to zero, we obtain a system of linear equations that can be written as

$$\sum_{n=0}^{N} K_{mn} c_n = \lambda \sum_{n=0}^{N} L_{mn} c_n, \tag{15}$$

$m = 0,1,\ldots,N$. The condition that the system (15) be soluble is the vanishing of the determinant:

$$\det|\mathsf{K} - \lambda \mathsf{L}| = 0. \tag{16}$$

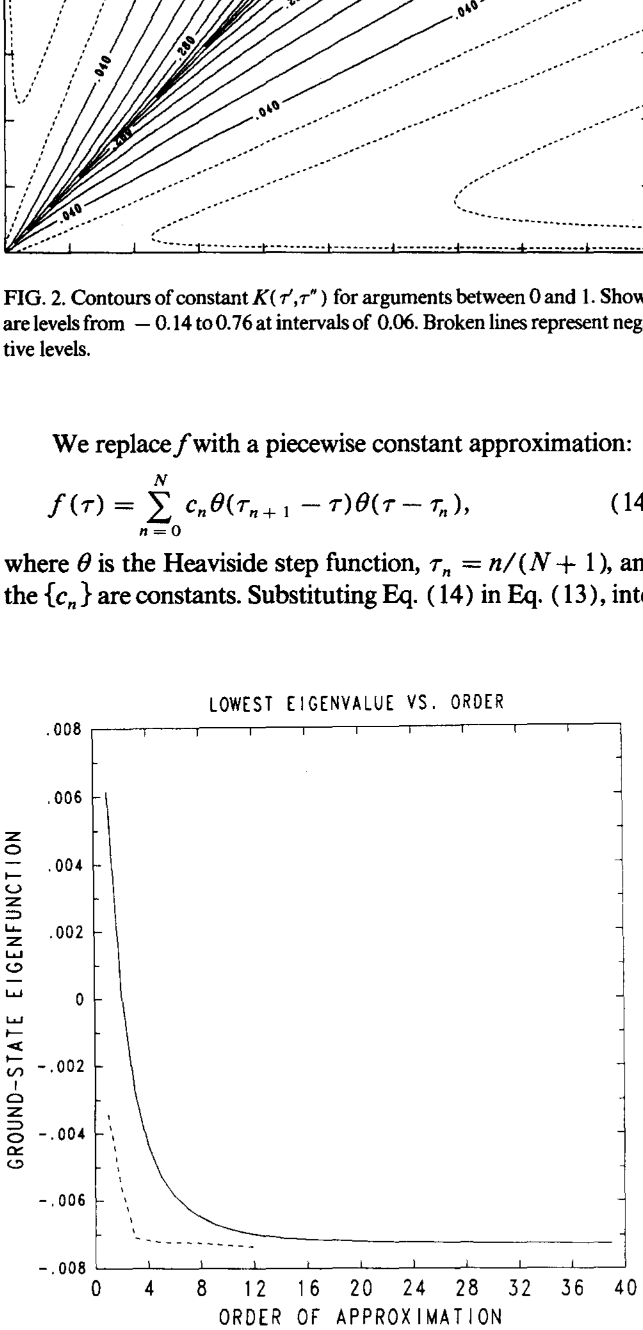


FIG. 3. Ground-state eigenvalue λ_0 obtained from variational calculation of order N, plotted versus N. Solid trace represents results of the square-wave approximation (14), broken trace represents results of the power-series approximation (17).

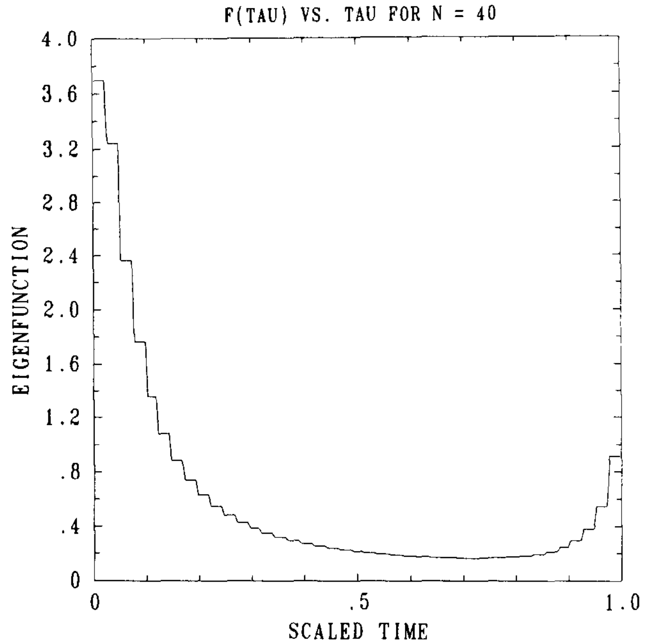


FIG. 4. Ground-state eigenfunction for square-wave approximation of order 40, plotted versus τ.

We solve Eq. (16), a standard problem in linear algebra, using the routine RSG from the EISPACK library.[4] The lowest eigenvalue λ_0 is found to be negative for $N \geqslant 3$. All other eigenvalues are positive. In Fig. 3 $\lambda_0(N)$ for $1 \leqslant N \leqslant 40$ is plotted. It is evident that λ_0 approaches a limiting value of approximately -7.3×10^{-3}. In Fig. 4 the ground-state eigenfunction for $N = 40$ is displayed. It is peaked at $\tau = 0$ and $\tau = 1$, which indicates that the most negative values of ΔW correspond to driving fields that rise fairly steeply at first, then slowly for a time, and then steeply again. [Remember that $f(\tau)$ is essentially the derivative of $B_0(t)$.] This behavior is seen in Fig. 5, which displays the corresponding driving field B_0 normalized with respect to its maximum value, plotted against the normalized time.

As a check, we carry out a second variational calculation using a power-series test function,

$$f(\tau) = \sum_{n=0}^{N} c_n \tau^n. \tag{17}$$

Again, it is necessary to solve Eq. (16). This problem is extremely ill conditioned, as usually happens when power series are employed in a variational calculation. By using 64-bit arithmetic it is nevertheless possible to find the first zero of the characteristic function (the lowest eigenvalue) iteratively for $N \lesssim 10$. The results, which are plotted as the broken curve in

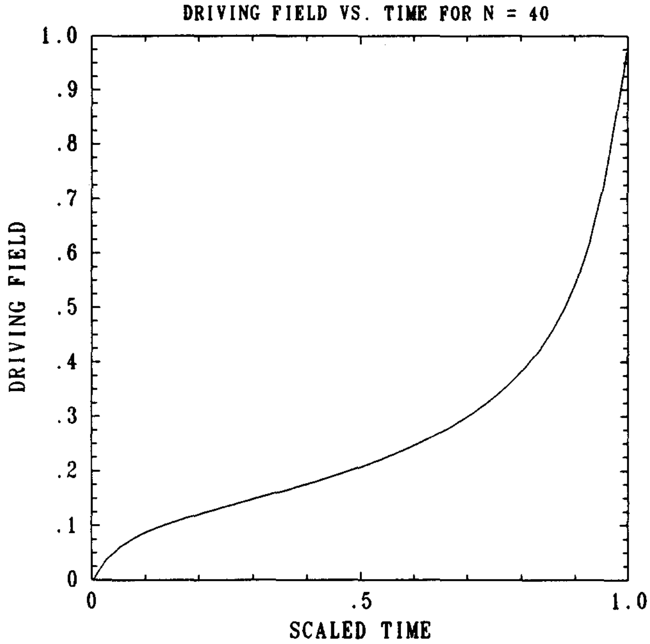


FIG. 5. Driving magnetic field corresponding to the eigenfunction shown in Fig. 4, plotted versus the normalized time.

Fig. 3, are consistent with those obtained using piecewise-constant test functions.

IV. CONCLUSIONS

In this paper we have reduced the problem of proving or disproving the positivity of ΔW to a spectral problem in terms of the operator K. The eigenvalues and eigenfunctions of this operator have been determined by means of a variational calculation. The ground state has been shown to have a negative eigenvalue. In addition to proving that ΔW can be negative and providing concrete examples, our calculation determines the form of B_0 that minimizes ΔW and the corresponding eigenvalue. It is found that the eigenfunction of this ground state has sharp peaks at $\tau = 0$ and $\tau = 1$ ($t' = t$ and $t' = 0$, respectively).

Although the present work is formulated in electromagnetic terms, the results are applicable elsewhere, e.g., in hydrodynamics. Specifically, consider a viscous incompressible fluid flowing past a solid wall with time-dependent velocity $-\mathbf{v}_0(t)$. In the rest frame of the fluid the wall moves with velocity $\mathbf{v}_0$ and the fluid becomes entrained as the velocity diffuses into it. Assuming that the motion is rectilinear and pressure gradients are negligible, we can describe the diffusion process by Eq. (1), where $\mathbf{B}$ is replaced by $\mathbf{v}$ and κ is replaced by the kinematic viscosity ν. The kinetic energy in the entrained fluid is

$$W_K = \frac{\rho}{2} \int d^3x\, v^2. \tag{18}$$

where ρ is the mass density. The total thermal energy generated by viscous dissipation in the fluid as a function of time is

$$W_V = \rho\nu \int_0^t dt' \int d^3x [\nabla\mathbf{v}(\mathbf{x},t') : \nabla\mathbf{v}(\mathbf{x},t')]. \tag{19}$$

In slab geometry these relations are formally identical with Eqs. (3) and (4). Based on the preceding analysis we can therefore conclude that, although $W_K \leqslant W_V$ holds most of the time (at least in slab geometry), it fails in some cases.

ACKNOWLEDGMENTS

We wish to thank Professor Richard Beals and Professor Ira Bernstein of Yale University for useful suggestions.

This work was supported in part by the U. S. Office of Naval Research.

[1]F. Herlach, in *Megagauss Physics and Technology*, Proceedings of the 2nd International Conference on Megagauss Magnetic Field Generation and Related Topics, edited by Peter J. Turchi (Plenum, New York, 1980).

[2]H. Knoepfel, *Pulsed High Magnetic Fields* (North-Holland, Amsterdam, 1970).

[3]H. S. Carslaw and J. C. Jaeger, *Conduction of Heat in Solids* (Oxford U.P., London, 1959), 2nd ed.

[4]B. T. Smith, J. M. Boyle, J. J. Dongarra, B. S. Garbow, Y. Ikebe, V. C. Klema, and C. B. Moler, *Matrix Eigenvalue Routines—EISPACK Guide*, 2nd ed. (Springer, New York, 1976); B. S. Garbow, J. M. Boyle, J. J. Dongarra, and C. B. Moler, *Matrix Eigenvalue Routines—EISPACK Guide Extension* (Springer, New York, 1977).

Hamiltonian methods for nonlinear sigma models

N. K. Pak
Middle East Technical University, Ankara, Turkey
R. Percacci
International School for Advanced Studies, Trieste, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Trieste, Italy

(Received 15 March 1989; accepted for publication 12 July 1989)

Nonlinear sigma models are studied in n space dimensions with values in a coset space G/H as infinite dimensional Hamiltonian systems. An "intrinsic" formulation is discussed in terms of coordinates on G/H, an "embedded" formulation in terms of fields satisfying a constraint and a "lifted" formulation in terms of fields having values in $G/\overline{H}$, where $\overline{H}$ is a normal subgroup of H. The coupling of the sigma model to Yang–Mills fields with structure group G is then considered, and it is shown that this system is equivalent to a massive Yang–Mills theory.

I. INTRODUCTION

The nonlinear sigma models (NSM) have long been used as phenomenological theories in elementary particle physics. From this point of view, the most interesting models are the so-called chiral models, which have values in $G_L \times G_R/G_V$, where G_L and G_R are isomorphic flavor groups and G_V is the vector subgroup (diagonal subgroup).[1] This model is supposed to represent the low energy approximation of an underlying fermionic theory (QCD) in which the left- and right-handed fermions transform under the groups G_L and G_R, respectively.

Other kinds of NSM [O(N) models, $\mathbf{CP}^N$ models, Grassmannian models] have been studied in great detail, in particular, in two dimensions. Although not of much use for phenomenology, these models are of great theoretical interest and have become an important subject in modern mathematical physics, because they are the simplest examples of nonlinear field theories.[2] More recently, sigma models with values in an arbitrary Riemannian manifold N have become interesting because of their applications to supergravity and superstring theories.[3]

The dynamical variable of an NSM is a map $\varphi: M \to N$, where M is space and N is an "internal" space. Throughout this work, we shall assume that M is $\mathbf{R}^n$, eventually compactified to S^n by adding one point at infinity, and N is a coset space G/H. The main reason for restricting our attention to homogeneous spaces is that we will study in some detail the gauging of the group G. It is clear, however, that many results can be generalized to the case of an N-valued model, with N an arbitrary Riemannian manifold.

The most "intrinsic" formulation of the NSM makes use of a set of real scalar fields, whose number equals the dimension of N. Every field gives a coordinate on N of the point $\varphi(x)$ as x varies in M. We will call this the "intrinsic" or "minimal" formulation of the NSM. In the case $N = G/H$, the NSM is often presented in different formulations, which make use of a nonminimal set of fields. In order to maintain equivalence to the intrinsic formulation, some of these fields have to be made unphysical, and this can be achieved in two alternative ways: by means of constraints at the Lagrangian level or by means of gauge invariances. The first possibility amounts to considering G/H as a submanifold of another (usually linear) manifold V, the second to considering G/H as the quotient of $G/\overline{H}$ by the action of K, where $\overline{H}$ is a normal subgroup of H and $K = H/\overline{H}$. Following Ref. 4, we call the first an "injective" or "embedded" formulation and the latter a "projective" or "lifted" formulation. Due to general theorems, every G/H-valued NSM admits an injective formulation (see Sec. III) and clearly, if H is nontrivial, also a projective formulation (with $K = H$ and $\overline{H}$ trivial). Although we shall not discuss this here, the injective and projective formulations exist also in the case of an arbitrary internal manifold N: In the former, N is embedded in another manifold V, in the latter it is regarded as the base manifold of some principal fiber bundle.

It should be observed that the two procedures can be used simultaneously in the same model. For instance, in the standard formulation of the $\mathbf{CP}^N$ model, where one starts with $N+1$ complex linear fields z^a, one degree of freedom is eliminated by means of the Lagrangian constraint $\Sigma_{a=1}^N |z^a|^2 = 1$, and another one by declaring the configurations z^a and λz^a to be gauge equivalent when λ is a complex function of modulus one. In order not to complicate the discussion too much, we shall discuss only the "purely injective" and the "purely projective" cases separately.

The embedded formulations are physically interesting because they make clear the relation existing between the nonlinear and the linear sigma models. At the classical level, every NSM is the limit of a linear sigma model (or Higgs model), when the potential becomes peaked on one orbit. The lifted formulations are interesting in a more abstract sense. The NSM is the only nonlinear field theory for which it is possible to work directly with the physical degrees of freedom. This is practically impossible, e.g., in the Yang–Mills (YM) case: An intrinsic formulation of YM theory would make use of the gauge equivalence classes of connections as fundamental variables. In this sense, the standard formulation of YM theory, in which the gauge potentials are the fundamental variables, is analogous to the lifted formulation of the NSM. In both cases, there is a gauge group and some of the dynamical variables are unphysical (the gauge degrees of freedom). One may hope that studying the relation between the intrinsic and lifted formulations of the

 0022-2488/89/122951-12$02.50

NSM may shed some light on the analogous problem for other nonlinear field theories.

In the Hamiltonian approach, the intrinsic formulation is completely straightforward and presents no constraints, whereas the nonminimal formulations require using Dirac's theory of constrained dynamical systems.[5] As we shall see, the embedded formulations lead to second class constraints at the Hamiltonian level, whereas the lifted formulations lead to first class constraints at the Hamiltonian level.

The second part of this paper will be devoted to the coupling of the G/H-valued NSM to a dynamical YM field for the group G. The system obtained in this way will be called a gauged NSM (GNSM). The motivation for studying this system is threefold. First, in a realistic phenomenological model gauging the flavor group (or a subgroup thereof) means taking into account the electroweak interactions of the underlying fermionic theory. More generally, due to the embedding theorems quoted above, every GNSM may be regarded, at least classically, as the strong coupling limit of a gauged Higgs model. This is also of some interest for phenomenology.[6] Second, the GNSM features prominently in recent attempts at a consistent quantization of anomalous YM theories. A GNSM with Wess–Zumino term may be regarded as the low energy approximation of a fermionic theory in which the fermions carry an anomalous representation of the flavor group. Thus studying the GNSM may give useful information on the physics of anomalous gauge theories.[7] As a third independent motivation, it can be shown that general relativity, and, more generally, any "metric" theory of gravity, contains dynamical variables corresponding to a GL(4)/O(4)-valued GNSM. Therefore, studying the GNSM also means studying a particular subsector of a theory of gravity.[4,8]

In this paper, the dimension of space n and the internal manifold G/H are left completely arbitrary. In Secs. II–IV, we study the intrinsic, embedded, and lifted formulations of the NSM, respectively, and show their equivalence. In Sec. V, we study the intrinsic formulation of the GNSM and show that it is equivalent to a massive YM theory. In Sec. VI, we briefly discuss the embedded and lifted formulations of the GNSM. The Appendix is devoted to massive YM theory.

II. THE MINIMAL FORMULATION

We begin by describing the intrinsic, or minimal, formulation of the NSM. The Hamiltonian formalism is entirely straightforward in this case, and a good deal of this section will actually be devoted to the geometry of homogeneous spaces and to setting up the notation.

Throughout this paper, G will denote a Lie group, not necessarily compact; the Lie algebra of G will be denoted $\mathscr{L}(G)$. We assume that in $\mathscr{L}(G)$ there is given an inner product, not necessarily Ad(G) invariant, and $\{T_a\}$, with $a=1,\ldots,\dim G$ will be an orthonormal basis in $\mathscr{L}(G)$. When the generators T_a are represented by matrices, we will assume that they are normalized so that $\mathrm{Tr}(T_aT_b) = -\frac{1}{2}\delta_{ab}$. The structure constants $f_{ab}{}^c$ are defined by

$$[T_a,T_b]=f_{ab}{}^cT_c; \tag{2.1}$$

if the inner product in $\mathscr{L}(G)$ is Ad(G) invariant, then the structure constants are totally antisymmetric [note that since the metric in $\mathscr{L}(G)$ is δ_{ab}, the distinction between upper and lower indices is immaterial].

Now, let H be a closed subgroup of G. We will assume that the coset space G/H is reductive, i.e., there exists an Ad(H)-invariant subspace $\mathscr{P}$ of $\mathscr{L}(G)$ such that

$$\mathscr{L}(G)=\mathscr{L}(H)\oplus\mathscr{P}. \tag{2.2}$$

The space $\mathscr{P}$ can be identified with the tangent space to G/H at the coset $o=eH$ (we will call o the "origin" of G/H). The group G acts on G/H from the left by $g(g'H)=(gg')H$; the restriction of this action to the subgroup H leaves the origin fixed and, hence, maps the tangent space to the origin to itself. This defines a linear representation of the group H, which is called the linear isotropy representation. When G/H is reductive, the linear isotropy representation can be identified with the restriction of the adjoint representation of H to $\mathscr{P}$. Note that if the basis is chosen in such a way that $\{T_{\hat a}\}$ with $\hat a=1,\ldots,\dim H$ is a basis in $\mathscr{L}(H)$ and $\{T_{\bar a}\}$ with $\bar a=1,\ldots,d$ (with $d=\dim G/H$) is a basis in $\mathscr{P}$, then

$$f_{\hat a\hat b}{}^{\bar c}=0;\quad f_{\hat a\bar b}{}^{\hat c}=0. \tag{2.3}$$

Let $\{y^\alpha\}$ with $\alpha=1,\ldots,d$ be local coordinates on G/H. Without loss of generality, we assume that the coordinates of the origin are $y^\alpha=0$. In the following, the components of all tensor fields on G/H will be referred to the natural bases $\{\partial_\alpha\}$ and $\{dy^\alpha\}$. The left action of G on G/H is generated by vector fields $K_a=K_a{}^\alpha\partial_\alpha$, which, under Lie brackets, form an algebra anti-isomorphic to $\mathscr{L}(G)$:

$$[K_a,K_b]^\beta=K_a{}^\alpha\partial_\alpha K_b{}^\beta-K_b{}^\alpha\partial_\alpha K_a{}^\beta=-f_{ab}{}^cK_c{}^\beta. \tag{2.4}$$

We assume that the restriction to $\mathscr{P}$ of the inner product in $\mathscr{L}(G)$ is Ad(H) invariant; via standard theorems, this gives rise to a G-invariant metric $h=h_{\alpha\beta}\,dy^\alpha\otimes dy^\beta$ on G/H. The vectors K_a are Killing vectors for this metric. That is, if $\mathscr{L}_v$ denotes the Lie derivative along v,

$$(\mathscr{L}_{K_a}h)_{\beta\gamma}=K_a{}^\alpha\partial_\alpha h_{\beta\gamma}+h_{\beta\delta}\partial_\gamma K_a{}^\delta+h_{\gamma\delta}\partial_\beta K_a{}^\delta=0. \tag{2.5}$$

The canonical configuration space for the NSM is the space $\mathscr{Q}_{\mathrm{NSM}}=\Gamma(\mathbf{R}^n,G/H)$ of maps from $\mathbf{R}^n$ to G/H satisfying certain regularity conditions and with prescribed behavior at spatial infinity. In order to guarantee finiteness of the energy, we will assume that $\varphi\in\mathscr{Q}_{\mathrm{NSM}}$ tends to a constant at infinity, so that $\mathbf{R}^n$ can be effectively replaced by S^n. The tangent and cotangent spaces to $\Gamma(S^n,G/H)$ at φ are the spaces of sections of the pullbacks by φ of the tangent and cotangent bundles of G/H, $\varphi^*T(G/H)$, and $\varphi^*T^*(G/H)$. The phase space $T^*\mathscr{Q}_{\mathrm{NSM}}$ is endowed with a natural symplectic structure which gives rise to the usual Poisson brackets

$$\{F,G\}=\int d\mathbf{x}\left(\frac{\delta F}{\delta\varphi^\alpha(\mathbf{x})}\frac{\delta G}{\delta\pi_\alpha(\mathbf{x})}-\frac{\partial F}{\delta\pi_\alpha(\mathbf{x})}\frac{\delta G}{\delta\varphi^\alpha(\mathbf{x})}\right), \tag{2.6}$$

where F and G are functionals of φ^α and their conjugate momenta π_α, i.e., functions on phase space. The canonical coordinates and momenta can be regarded as locally defined functions on $T^*\mathscr{Q}_{\rm NSM}$; they satisfy the canonical Poisson brackets

$$\{\varphi^\alpha(\mathbf{x}),\varphi^\beta(\mathbf{y})\}=0, \tag{2.7a}$$

$$\{\varphi^\alpha(\mathbf{x}),\pi_\beta(\mathbf{y})\}=\delta^\alpha_\beta\delta^{(n)}(\mathbf{x}-\mathbf{y}), \tag{2.7b}$$

$$\{\pi_\alpha(\mathbf{x}),\pi_\beta(\mathbf{y})\}=0. \tag{2.7c}$$

The Lagrangian of the NSM is

$$\mathscr{L}=-(1/2f^2)h_{\alpha\beta}(\varphi)\partial^\mu\varphi^\alpha\partial_\mu\varphi^\beta, \tag{2.8}$$

where $\mu, \nu ... = 0,1,...,n$ are space-time indices and the metric is $\eta_{\mu\nu}=\operatorname{diag}(-1,1,...,1)$. The space indices will be denoted $i,j... = 1,...,n$. The momentum conjugate to φ^α is

$$\pi_\alpha=\frac{\delta\mathscr{L}}{\delta\partial_0\varphi^\alpha}=\frac{1}{f^2}h_{\alpha\beta}(\varphi)\partial_0\varphi^\beta. \tag{2.9}$$

Since the matrix $h_{\alpha\beta}$ is nonsingular, the relation between velocities and momenta is invertible ($\partial_0\varphi^\alpha=f^2h^{\alpha\beta}\pi_\beta$), so in this formulation the NSM is a regular dynamical system. The canonical Hamiltonian is

$$H=\int d\mathbf{x}\left[\frac{f^2}{2}h^{\alpha\beta}(\varphi)\pi_\alpha\pi_\beta+\frac{1}{2f^2}h_{\alpha\beta}(\varphi)\partial_i\varphi^\alpha\partial_i\varphi^\beta\right]. \tag{2.10}$$

From the canonical Poisson brackets in (2.7), we get Hamilton's equations:

$$\frac{d\varphi^\alpha}{dt}=\{\varphi^\alpha,H\}=f^2h^{\alpha\beta}\pi_\beta, \tag{2.11a}$$

$$\frac{d\pi_\alpha}{dt}=\{\pi_\alpha,H\}=\frac{1}{f^2}\partial_i(h_{\alpha\beta}\partial_i\varphi^\beta)-\frac{f^2}{2}\partial_\alpha h^{\beta\gamma}\pi_\beta\pi_\gamma-\frac{1}{2f^2}\partial_\alpha h_{\beta\gamma}\partial_i\varphi^\beta\partial_i\varphi^\gamma. \tag{2.11b}$$

The first one reproduces the relation (2.9) between velocities and momenta; the two together give rise to the covariant equation

$$\partial_\mu\partial^\mu\varphi^\alpha+\Gamma^\alpha_{\beta\gamma}\partial_\mu\varphi^\beta\partial^\mu\varphi^\gamma=0, \tag{2.12}$$

where $\Gamma^\alpha_{\beta\gamma}$ are the Christoffel symbols of the metric $h_{\alpha\beta}$. This is just the Euler–Lagrange equation that one gets from varying (2.8).

For every $u\in\mathscr{L}(G)$, we define a function ρ_u on $T^*\mathscr{Q}_{\rm NSM}$ by

$$\rho_u(\varphi,\pi)=\int d\mathbf{x}\, u^a K_a{}^\alpha(\varphi)\pi_\alpha.$$

We have

$$\{\rho_u,\varphi^\alpha\}=-u^aK_a{}^\alpha, \tag{2.13a}$$

$$\{\rho_u,\pi_\alpha\}=u^a\partial_\alpha K_a{}^\beta\pi_\beta, \tag{2.13b}$$

so the functions ρ_u generate the action of G on $T^*\mathscr{Q}_{\rm NSM}$. They satisfy the algebra

$$\{\rho_u,\rho_v\}=\rho_{[u,v]}. \tag{2.14}$$

The G invariance of the theory follows from

$$\{\rho_u,H\}=0. \tag{2.15}$$

The energy-momentum tensor corresponding to the Lagrangian (2.8) is

$$T_{\mu\nu}=(1/f^2)h_{\alpha\beta}(\varphi)[\partial_\mu\varphi^\alpha\partial_\nu\varphi^\beta-\tfrac{1}{2}g_{\mu\nu}g^{\rho\sigma}\partial_\rho\varphi^\alpha\partial_\sigma\varphi^\beta]. \tag{2.16}$$

In the rest of this section, we consider the special case of an NSM with values in G. If we regard this as the special case $H=\{e\}$ of the situation considered above, then the model is called a "principal" NSM. In this case, one only requires invariance under the left action of G on itself, which is given by left multiplication. On the other hand, we can also think of G as the coset space $G\times G/\Delta G$, where $G\times G$ acts on G from the left by $(a,b)g=agb^{-1}$ and ΔG is the diagonal subgroup, consisting of couples of the form (a,a). In this case, one requires that the metric h on G be invariant under the action of $G\times G$, i.e., that it be bi-invariant. The corresponding model is called a "chiral" NSM. We will consider here the second case.

The left and right actions of G on itself are generated by vector fields $R_a=R_a{}^\alpha\partial_\alpha$ and $L_a=L_a{}^\alpha\partial_\alpha$, respectively. If we identify $\mathscr{L}(G)$ with the tangent space to G at the identity e, we have $R_a(e)=L_a(e)=T_a$. These vector fields obey the following algebra:

$$[R_a,R_b]^\beta=R_a{}^\alpha\partial_\alpha R_b{}^\beta-R_b{}^\alpha\partial_\alpha R_a{}^\beta=-f_{ab}{}^cR_c{}^\beta; \tag{2.17a}$$

$$[L_a,L_b]^\beta=L_a{}^\alpha\partial_\alpha L_b{}^\beta-L_b{}^\alpha\partial_\alpha L_a{}^\beta=f_{ab}{}^cL_c{}^\beta; \tag{2.17b}$$

$$[R_a,L_b]^\beta=R_a{}^\alpha\partial_\alpha L_b{}^\beta-L_b{}^\alpha\partial_\alpha R_a{}^\beta=0. \tag{2.17c}$$

In particular, (2.17c) shows that the vector fields R_a are right invariant and the vector fields L_a are left invariant (this is the reason for the notation). We assume that the inner product in $\mathscr{L}(G)$ is $\operatorname{Ad}(G)$ invariant; from standard theorems, we then have a bi-invariant metric h on G. The vector fields L_a and R_a are Killing vectors for this metric:

$$(\mathscr{L}_{R_a}h)_{\beta\gamma}=R_a{}^\alpha\partial_\alpha h_{\beta\gamma}+h_{\beta\delta}\partial_\gamma R_a{}^\delta+h_{\gamma\delta}\partial_\beta R_a{}^\delta=0; \tag{2.18a}$$

$$(\mathscr{L}_{L_a}h)_{\beta\gamma}=L_a{}^\alpha\partial_\alpha h_{\beta\gamma}+h_{\beta\delta}\partial_\gamma L_a{}^\delta+h_{\gamma\delta}\partial_\beta L_a{}^\delta=0. \tag{2.18b}$$

It follows from these definitions that the vector fields R_a and L_a are orthonormal with respect to h:

$$h_{\alpha\beta}R_a{}^\alpha R_b{}^\beta=h_{\alpha\beta}L_a{}^\alpha L_b{}^\beta=\delta_{ab}. \tag{2.19}$$

The right- and left-invariant [$\mathscr{L}(G)$-valued] Maurer–Cartan forms $\theta_R=\theta_R{}^aT_a$ and $\theta^L=\theta_L{}^aT_a$ are defined by $\theta_R{}^a(R_b)=\delta^a_b$ and $\theta_L{}^a(L_b)=\delta^a_b$ and, therefore, have components $\theta_R{}^a=(R^{-1})_\alpha{}^a dy^\alpha$ and $\theta_L{}^a=(L^{-1})_\alpha{}^a dy^\alpha$, where $(R^{-1})_\alpha{}^a$ and $(L^{-1})_\alpha{}^a$ are the matrix inverses of $R_a{}^\alpha$ and $L_a{}^\alpha$, respectively. From (2.19), we see that the matrices $(R^{-1})_\alpha{}^a$ and $(L^{-1})_\alpha{}^a$ can be obtained from $R_a{}^\alpha$ and $L_a{}^\alpha$ by raising and lowering indices and transposing:

$$(R^{-1})_\alpha{}^a=\delta^{ab}R_b{}^\beta h_{\beta\alpha}, \tag{2.20a}$$

$$(L^{-1})_\alpha{}^a=\delta^{ab}L_b{}^\beta h_{\beta\alpha}. \tag{2.20b}$$

In accordance with the standard rules of tensor calculus, we will, therefore, always write $R^a{}_\alpha$ and $L^a{}_\alpha$ instead of $(R^{-1})_\alpha{}^a$ and $(L^{-1})_\alpha{}^a$. With these conventions, the Maurer–Cartan forms can be written $\theta_R = R^a{}_\alpha\, dy^\alpha \otimes T_a$ and $\theta_L = L^a{}_\alpha\, dy^\alpha \otimes T_a$. The Maurer–Cartan forms satisfy the following Mauer–Cartan equations, which are equivalent to Eqs. (2.17a) and (2.17b):

$$\partial_\beta R^a{}_\gamma - \partial_\gamma R^a_\beta = f_{bc}{}^a R^b{}_\beta R^c{}_\gamma, \tag{2.21a}$$

$$\partial_\beta L^a{}_\gamma - \partial_\gamma L^a_\beta = -f_{bc}{}^a L^b{}_\beta L^c{}_\gamma. \tag{2.21b}$$

Given a map φ: $\mathbf{R}^{n+1} \to G$, the pullbacks of the Maurer–Cartan forms on space-time are

$$\varphi^*\theta_R = d\varphi\, \varphi^{-1} = R^a{}_\mu\, dx^\mu \otimes T_a, \tag{2.22a}$$

$$\varphi^*\theta_L = \varphi^{-1}\, d\varphi = L^a{}_\mu\, dx^\mu \otimes T_a, \tag{2.22b}$$

where $R^a{}_\mu = \partial_\mu \varphi^\alpha R^a{}_\alpha$ and $L^a{}_\mu = \partial_\mu \varphi^\alpha L^a{}_\alpha$. Note that the spatial components $R^a{}_i$ and $L^a{}_i$ depend only on the canonical coordinates φ^α, whereas the time components depend on the momenta:

$$R^a{}_0 = f^2 R_a{}^\alpha \pi_\alpha, \tag{2.23a}$$

$$L^a{}_0 = f^2 L_a{}^\alpha \pi_\alpha. \tag{2.23b}$$

If we define the vector and axial vector currents $V^a{}_\mu = R^a{}_\mu - L^a{}_\mu$ and $A^a{}_\mu = R^a{}_\mu + L^a{}_\mu$, we find the following Poisson brackets, which are well known in current algebra:

$$\{V^a{}_0(\mathbf{x}),\ V^b{}_0(\mathbf{y})\} = -f_{abc} V^c{}_0(\mathbf{x})\delta^{(n)}(\mathbf{x}-\mathbf{y}), \tag{2.24a}$$

$$\{V^a{}_0(\mathbf{x}),\ A^b{}_0(\mathbf{y})\} = f_{abc} A^c{}_0(\mathbf{x})\delta^{(n)}(\mathbf{x}-\mathbf{y}), \tag{2.24b}$$

$$\{A^a{}_0(\mathbf{x}),\ A^b{}_0(\mathbf{y})\} = -f_{abc} V^c{}_0(\mathbf{x})\delta^{(n)}(\mathbf{x}-\mathbf{y}). \tag{2.24c}$$

The Poisson brackets between the time and space components of the currents depend explicitly upon φ and do not reproduce the standard forms of current algebra. Using (2.19), the energy-momentum tensor (2.16) can be rewritten in the Sugawara form[9]:

$$T_{\mu\nu} = -\frac{1}{4f^2}\Big[V^a{}_\mu V^a{}_\nu - \frac{1}{2} g_{\mu\nu}\, g^{\rho\sigma} V^a{}_\rho V^a{}_\sigma + A^a{}_\mu A^a{}_\nu - \frac{1}{2} g_{\mu\nu}\, g^{\rho\sigma} A^a{}_\rho A^a{}_\sigma \Big]. \tag{2.25}$$

In particular, the Hamiltonian (2.10) is

$$H = \int d\mathbf{x}\, T_{00} = -\frac{1}{8f^2}\int d\mathbf{x}\,[V^a{}_0 V^a{}_0 + V^a{}_i V^a{}_i + A^a{}_0 A^a{}_0 + A^a{}_i A^a{}_i]. \tag{2.26}$$

III. THE EMBEDDED FORMULATION

A general theorem states that every compact Riemannian homogeneous space can be embedded isometrically and equivariantly in some vector space V carrying a linear representation of G.[10] Here, we will identify V with $\mathbf{R}^N$, for N sufficiently large. Choose an orthonormal basis in $\mathbf{R}^N$ and denote Y^m, with $m = 1,...,N$, the components of a vector Y relative to this basis. The embedding of the theorem will be denoted j: $G/H \to \mathbf{R}^N$; it is represented locally by a set of functions $j^m = j^m(y^\alpha)$. The condition that j be isometric means that

$$h_{\alpha\beta} = \frac{\partial j^m}{\partial y^\alpha}\frac{\partial j^n}{\partial y^\beta}\delta_{mn}. \tag{3.1}$$

The condition that j be equivariant means that the image of the vector fields K_a in $\mathbf{R}^N$ coincide with the generators of the action of G on $\mathbf{R}^N$; if we denote $(T_a)^m{}_n$ the matrix that represents $T_a \in \mathscr{L}(G)$, we must have

$$K_a{}^\alpha \frac{\partial j^m}{\partial y^\alpha} = (T_a)^m{}_n Y^n. \tag{3.2}$$

It is possible to choose locally a set of curvilinear coordinates $\{Z^m\}$ on $\mathbf{R}^N$ such that the surfaces $Z^m =$ constant for $m = d+1,...,N$ are the orbits of the group G and $\{Z^m\}$ for $m = 1,...,d$ are coordinates in the orbits (recall that $d = \dim G/H$). In particular, we can choose the coordinates $\{Z^m\}$ in such a way that the embedding j is represented by

$$j^\alpha(y^\beta) = y^\alpha \quad \text{for} \quad \alpha = 1,...,d \tag{3.3a}$$

$$j^a(y^\beta) = 0 \quad \text{for} \quad a = m - d = 1,...,N-d. \tag{3.3b}$$

The functions $Z^m = Z^m(Y^n)$ define locally a coordinate transformation on $\mathbf{R}^N$ and, therefore, the Jacobian matrix $\partial Z^m/\partial Y^n$ is nondegenerate. Equation (3.3b) shows that we can take $Z^a(Y) = 0$, for $a = m - d = 1,...,N-d$ as the equations defining the embedding of G/H in $\mathbf{R}^N$. Note that the functions $Z^a(Y)$ are G invariant, in the sense that

$$(T_b)^m{}_n Y^n \frac{\partial Z^a}{\partial Y^m} = 0. \tag{3.4}$$

The NSM can be described in terms of N linear fields $\phi^m(x)$ (with $m = 1,...,N$) satisfying the constraints $Z^a(\phi) = 0$ for $a = 1,...,N-d$. The constraints can be incorporated in the dynamics by introducing $N-d$ Lagrange multipliers λ^a. Therefore, the canonical configuration space is $\mathscr{Q}' = \Gamma(\mathbf{R}^n,\mathbf{R}^N) \times \Gamma(\mathbf{R}^n,\mathbf{R}^{N-d})$. The Lagrangian is

$$\mathscr{L} = -\frac{1}{2f^2}\partial^\mu\phi^m\partial_\mu\phi_m + \sum_{a=1}^{N-d}\lambda^a Z^a(\phi). \tag{3.5}$$

This Lagrangian is equivalent to the Lagrangian (2.10) and can be obtained as a limiting case of a Higgs Lagrangian.

We shall now develop the Hamiltonian form of the embedded formulation and prove the equivalence with the intrinsic formulation. The momentum conjugate to ϕ^m is

$$\Pi_m = \frac{\delta\mathscr{L}}{\delta\partial_0\phi^m} = \frac{1}{f^2}\partial_0\phi^m, \tag{3.6}$$

whereas the momentum conjugate to λ^a vanishes:

$$\Lambda_a = \frac{\delta\mathscr{L}}{\delta\partial_0\lambda^a} = 0. \tag{3.7}$$

This defines the primary constraints of the theory: $0 \approx \psi_1^a = \Lambda^a$, where $\approx$ means "weak" equality. The canonical Hamiltonian is

$$H_C = \int d\mathbf{x}\left[\frac{f^2}{2}\Pi_m\Pi_m + \frac{1}{2f^2}\partial_i\phi^m\partial_i\phi^m - \sum_{a=1}^{N-d}\lambda^a Z^a\right], \tag{3.8}$$

and we define the primary Hamiltonian by

$$H_P = H_C + \int d\mathbf{x}\sum_{a=1}^{N-d} u^a\psi_1^a, \tag{3.9}$$

where u^a are other Lagrange multipliers. Consistency of the primary constraints under the time evolution generated by H_P gives secondary constraints

$$0 \approx \psi_2^a = \{\psi_1^a, H_P\} = Z^a(\phi). \tag{3.10}$$

Repeating this process, we get the following "tertiary" and "quaternary" constraints:

$$0 \approx \psi_3^a = \{\psi_2^a, H_P\} = f^2\frac{\partial Z^a}{\partial\phi^m}\Pi_m, \tag{3.11}$$

$$\begin{aligned}0 \approx \psi_4^a &= \{\psi_3^a, H_P\}\\ &= f^4\frac{\partial^2 Z^a}{\partial\phi^m\,\partial\phi^n}\Pi_m\Pi_n + \frac{\partial Z^a}{\partial\phi^n}\Big(\partial_i\,\partial_i\phi^n\\ &\quad + f^2\sum_{b=1}^{N-d}\lambda^b\frac{\partial Z^b}{\partial\phi^n}\Big).\end{aligned} \tag{3.12}$$

No further constraints arise. In fact, consider the equation

$$0 \approx \{\psi_4^a, H_P\} = \sum_{b=1}^{N-d}\frac{\partial Z^a}{\partial\phi^m}\frac{\partial Z^b}{\partial\phi^m}u^b + \{\psi_4^a, H_C\}. \tag{3.13}$$

The matrix $M^{ab} = (\partial Z^a/\partial\phi^m)(\partial Z^b/\partial\phi^m)$ is nondegenerate because $\partial Z^a/\partial\phi^m$ has maximal rank $N-d$. Therefore, Eq. (3.13) determines the Lagrange multipliers u^a and the process stops. Note that in a similar way Equation (3.12) determines the Lagrange multipliers λ^a. Taking Poisson brackets, it can easily be seen that all constraints are second class.

The equivalence of the embedded formulation to the intrinsic one can be easily established using the adapted coordinates that were defined in Eq. (3.3). In these coordinates, the constraints are $\psi_1^a = \Lambda_a$, $\psi_2^a = \phi^a$, $\psi_3^a = \Pi_a$ and $\psi_4^a = \partial_i\,\partial_i\phi^a + \lambda^a$. Note that the second and fourth constraint imply $\lambda^a = 0$. From (3.3a), we then have $\phi^\alpha = \varphi^\alpha$ and $\Pi_\alpha = \pi_\alpha$. Therefore, the constrained submanifold of $T^*\mathscr{Q}'_{\rm NSM}$ can be identified with $T^*\mathscr{Q}_{\rm NSM}$. Finally, using (3.1), it can be seen that the Hamiltonian (3.8), restricted to the constrained submanifold, is identical to the Hamiltonian (2.12).

IV. THE LIFTED FORMULATION

The G/H-valued NSM can be reformulated in terms of a G-valued field subject to a gauge invariance under the group H. Borrowing from the mathematical terminology, we call this G-valued field a "lifted" field and the corresponding formulation of the NSM a "lifted" formulation. In certain cases, e.g., when G/H is a Grassmann manifold, it is customary not to lift the fields all the way to G, but only to $G/\overline{H}$, where $\overline{H}$ is a normal subgroup of H. Global aspects of this procedure have been discussed in the Lagrangian formalism in Ref. 4. Here, we describe the corresponding constrained Hamiltonian form of the theory.

When $\overline{H}\subset H\subset G$ is a chain of closed subgroups and $\overline{H}$ is normal in H, $G/\overline{H}$ is a principal $K = H/\overline{H}$ bundle over G/H. There is a left action of G on $G/\overline{H}$ defined by $g(g'\overline{H}) = (gg')\overline{H}$ and a right action of $K = H/\overline{H}$ on $G/\overline{H}$ defined by $(g\overline{H})\,(h\overline{H}) = (gh)\overline{H}$. The bundle projection μ: $G/\overline{H}\to G/H$ consists of taking equivalence classes under this action of K: $\mu(g\overline{H}) = gH$. The vector fields generating the actions of G and K on $G/\overline{H}$ will be denoted $\overline{K}_a{}^{\bar\alpha}$ and $F_{\tilde a}{}^{\bar\alpha}$, respectively [as before, $a = 1,\dots,\dim G$ is an index in $\mathscr{L}(G)$, $\tilde a = 1,\dots,\dim K$ is an index in $\mathscr{L}(K)$ and $\bar\alpha = 1,\dots,\bar d$, with $\bar d = \dim G/\overline{H}$, is an index labeling local coordinates $\{\bar y^{\bar\alpha}\}$ on $G/\overline{H}$]. The vector fields $F_{\tilde a}$ are called the fundamental vector fields.

The tangent space to $G/\overline{H}$ at the origin $\bar o = e\overline{H}$ can be identified with a subspace $\overline{\mathscr{P}}\in\mathscr{L}(G)$ complementary to $\mathscr{L}(\overline{H})$; we assume that $\mathscr{P}\subset\overline{\mathscr{P}}$. Clearly, $\mathscr{P}$ defines a subspace of $T_{\bar o}(G/\overline{H})$, which is complementary to the vertical subspace spanned by the vectors $F_{\tilde a}$. The bundle $G/\overline{H}\to G/H$ carries a canonical G-invariant connection that is uniquely defined by the condition that the horizontal space at $\bar o$ is $\mathscr{P}$. It can be shown that the condition that G/H be reductive is sufficient to guarantee that this connection is well defined, in the sense that the horizontal space at $g\overline{H}$ does not depend on the particular group element g which is used to transport $\mathscr{P}$ from $\bar o$ to $g\overline{H}$. To this canonical connection is associated a connection form ω, which is a left-invariant $\mathscr{L}(K)$ valued-one on $G/\overline{H}$. In local coordinates $\omega = \omega^{\tilde a}{}_{\bar\alpha}\,d\bar y^{\bar\alpha}\otimes T_{\tilde a}$, where $\{T_{\tilde a}\}$ is a basis in $\mathscr{L}(K)$. It is a fundamental property of the connection form that $\omega(F_{\tilde a}) = T_{\tilde a}$; in components

$$\omega^{\tilde a}{}_{\bar\alpha}F_{\tilde b}{}^{\bar\alpha} = \delta^{\tilde a}{}_{\tilde b}. \tag{4.1}$$

There follows that the tensors

$$V^{\bar\alpha}{}_{\bar\beta} = F_{\tilde a}{}^{\bar\alpha}\omega^{\tilde a}{}_{\bar\beta}, \tag{4.2a}$$

$$H^{\bar\alpha}{}_{\bar\beta} = \delta^{\bar\alpha}{}_{\bar\beta} - V^{\bar\alpha}{}_{\bar\beta}, \tag{4.2b}$$

are the vertical and horizontal projectors. Given a map φ: $\mathbf{R}^n\to G/H$, we say that a map $\bar\varphi$: $\mathbf{R}^n\to G/\overline{H}$ is a lift of φ if $\mu(\bar\varphi(x)) = \varphi(x)$. If $\bar\varphi$ is a lift of φ, then also $\bar\varphi'$, defined by $\bar\varphi'(x) = (\bar\varphi(x))k(x)$ for some map k: $\mathbf{R}^n\to K$, is a lift of φ. Therefore, the lifted NSM has a nontrivial gauge group. We define the covariant derivative of $\bar\varphi$ by

$$\begin{aligned}D_\mu\bar\varphi^{\bar\alpha} &= H^{\bar\alpha}{}_{\bar\beta}\partial_\mu\bar\varphi^{\bar\beta}\\ &= \partial_\mu\bar\varphi^{\bar\alpha} - B^{\tilde a}_\mu F_{\tilde a}{}^{\bar\alpha}(\bar\varphi),\end{aligned} \tag{4.3}$$

where $B^{\tilde a}_{\mu} = \partial_\mu\bar\varphi^{\bar\beta}\omega^{\tilde a}{}_{\bar\beta}(\varphi)$ is a composite gauge potential.

Let $\bar h = \bar h_{\bar\alpha\bar\beta}\,d\bar y^{\bar\alpha}\otimes d\bar y^{\bar\beta}$ be a left-G- and right-K-invariant metric on $G/\overline{H}$; we assume that the corresponding inner product in $\overline{\mathscr{P}}$, restricted to $\mathscr{P}$, is the inner product corresponding to the metric h on G/H. The canonical configuration space of the NSM in the lifted formulation is $\overline{\mathscr{Q}}_{\rm NSM} = \Gamma(\mathbf{R}^n, G/\overline{H})$. The Lagrangian of the lifted NSM is

$$\mathscr{L} = -(1/2f^2)\bar{h}_{\bar{\alpha}\bar{\beta}}(\bar{\varphi})D^{\mu}\bar{\varphi}^{\bar{\alpha}}D_{\mu}\bar{\varphi}^{\bar{\beta}}. \tag{4.4}$$

Because of its gauge invariance this Lagrangian depends really only on φ and it can be seen that it coincides with the Lagrangian (2.8). Again, we defer comparison of this formulation with the intrinsic one until we have developed the Hamiltonian formalism.

The canonical momentum conjugate to $\bar{\varphi}^{\bar{\alpha}}$ is

$$\bar{\pi}_{\bar{\alpha}} = \frac{\delta \mathscr{L}}{\delta \partial_0 \bar{\varphi}^{\bar{\alpha}}} = \frac{1}{f^2}\bar{h}_{\bar{\alpha}\bar{\beta}}(\bar{\varphi})D_0\bar{\varphi}^{\bar{\beta}}. \tag{4.5}$$

Since $H^{\bar{\alpha}}{}_{\bar{\beta}}$ is degenerate, this relation is not invertible and, therefore, the lifted NSM is a singular dynamical system. The canonical Hamiltonian is

$$H_C = H_0 + \int d\mathbf{x}\, B_0^{\tilde{a}}\, \psi_{\tilde{a}}, \tag{4.6}$$

where

$$H_0 = \int d\mathbf{x}\left[\frac{f^2}{2}\bar{h}^{\bar{\alpha}\bar{\beta}}\bar{\pi}_{\bar{\alpha}}\bar{\pi}_{\bar{\beta}} + \frac{1}{2f^2}\bar{h}_{\bar{\alpha}\bar{\beta}}D_i\bar{\varphi}^{\bar{\alpha}}D_i\bar{\varphi}^{\bar{\beta}}\right], \tag{4.7}$$

and

$$\psi_{\tilde{a}} = F_{\tilde{a}}{}^{\bar{\beta}}(\bar{\varphi})\bar{\pi}_{\bar{\beta}}. \tag{4.8}$$

Since vertical and horizontal vectors are orthogonal in the metric $\bar{h}$, if we use the relation (4.5) between velocities and momenta, we find $\psi_{\tilde{a}} = 0$. These are the primary constraints of the theory. Their Poisson brackets are

$$\{\psi_{\tilde{a}}(\mathbf{x}),\psi_{\tilde{b}}(\mathbf{y})\} = f_{\tilde{a}\tilde{b}}{}^{\tilde{c}}\psi_{\tilde{c}}(\mathbf{x})\delta^{(n)}(\mathbf{x}-\mathbf{y}). \tag{4.9}$$

In order to understand their physical meaning, let $u = u^{\tilde{a}}T_{\tilde{a}}$: $\mathbf{R}^n \to \mathscr{L}(K)$ be an infinitesimal gauge transformation, and consider the function Q_u on $T^*\bar{\mathscr{Q}}_{\rm NSM}$ defined by

$$Q_u(\bar{\varphi},\bar{\pi}) = \int dx\, u^{\tilde{a}}\psi_{\tilde{a}}.$$

We have

$$\{Q_u,\bar{\varphi}^{\bar{\alpha}}\} = -u^{\tilde{a}}F_{\tilde{a}}{}^{\bar{\alpha}}, \tag{4.10a}$$

$$\{Q_u,\bar{\pi}_{\bar{\alpha}}\} = u^{\tilde{a}}\partial_{\bar{\alpha}}F_{\tilde{a}}{}^{\bar{\beta}}\bar{\pi}^{\bar{\beta}}. \tag{4.10b}$$

These are just the infinitesimal variations of $\bar{\varphi}^{\bar{\alpha}}$ and $\bar{\pi}_{\bar{\alpha}}$ under the gauge transformation u, so we conclude that the primary constraints are the generators of the gauge transformations of this theory. Notice that this is true also for time-dependent transformations. From (4.9), we get

$$\{Q_u,Q_v\} = -Q_{[u,v]}. \tag{4.11}$$

In order to see whether there are any secondary constraints, we define the primary Hamiltonian

$$H_P = H_0 + \int dx\, \lambda^{\tilde{a}}\psi_{\tilde{a}},$$

where the last term of H_C in Eq. (4.6) has been absorbed in the definition of the Lagrange multipliers $\lambda^{\tilde{a}}$. Using the right-K invariance of $\bar{h}$, one finds, after some algebra,

$$\{\psi_{\tilde{a}},H_0\} = 0; \tag{4.12}$$

this is just the statement that H_0 is gauge invariant. Therefore,

$$\{\psi_{\tilde{a}},H_P\} = \int dx\, \lambda^{\tilde{b}}\{\psi_{\tilde{a}},\psi_{\tilde{b}}\} \approx 0. \tag{4.13}$$

Thus there are no secondary constraints, and all primary constraints are first class.

The equivalence of the lifted formulation to the intrinsic one can be proven by choosing local coordinates in $G/\bar{H}$, which are adapted to the fibration $G/\bar{H} \to G/H$, in the sense that $\bar{y}^{\alpha} = y^{\alpha}$ when $\alpha = \bar{\alpha} = 1,...,d$ and $\{\bar{y}^{\tilde{\alpha}}\}$ with $\tilde{\alpha} = \bar{\alpha} - d = 1,...,\dim K$ are coordinates in the fibers. The fields $\bar{\varphi}^{\tilde{\alpha}}$ are the gauge degrees of freedom; this suggests that we take $\bar{\varphi}^{\tilde{\alpha}}$ as gauge conditions. Note that, in general, this is possible only locally, so we have an analog of the Gribov ambiguity[11]; we disregard this type of global problems here. In adapted coordinates, $F_{\tilde{a}}{}^{\alpha} = 0$, and $F_{\tilde{a}}{}^{\tilde{\alpha}}$ is nondegenerate; therefore, an equivalent set of primary constraints is $\bar{\pi}_{\tilde{\alpha}} \approx 0$. Since $\{\bar{\varphi}^{\tilde{\alpha}},\bar{\pi}_{\tilde{\beta}}\} = \delta^{\tilde{\alpha}}{}_{\tilde{\beta}}$, $\bar{\varphi}^{\tilde{\alpha}}$ are indeed good gauge conditions in the sense of Dirac's theory. The constraints $\bar{\varphi}^{\tilde{\alpha}} \approx 0$ and $\bar{\pi}_{\tilde{\alpha}} \approx 0$ define locally a submanifold of the phase space $T^*\bar{\mathscr{Q}}_{\rm NSM}$, which can be identified with the phase space of the intrinsic formulation, $T^*\mathscr{Q}_{\rm NSM}$. On this submanifold, $D_{\mu}\bar{\varphi}^{\tilde{\alpha}} = 0$ and $D_{\mu}\bar{\varphi}^{\alpha} = \partial_{\mu}\varphi^{\alpha}$. Since the restriction of the metric $\bar{h}$ to the horizontal spaces coincides with the metric h, one finds that on the constrained submanifold the Hamiltonian (4.6) is identical to the Hamiltonian (2.10). This proves that the two formulations of the NSM are equivalent.

V. THE GAUGED NONLINEAR SIGMA MODEL

We now consider the minimal coupling of a G/H-valued NSM to a G-Yang–Mills field. We call the dynamical system obtained in this way a gauged NSM (GNSM). The kinematical and topological aspects of this system have been discussed within the Lagrangian formalism in Ref. 4. In this paper, we disregard global problems and consider only the local dynamics. In order to avoid unnecessary complications, we begin by giving a minimal description of the GNSM, i.e., the scalar fields will be as in Sec. II. The canonical configuration space is $\mathscr{Q}_{\rm GNSM} = \mathscr{C} \times \Gamma(\mathbf{R}^n,\mathscr{L}(G)) \times \Gamma(\mathbf{R}^n,G/H)$, where $\mathscr{C}$ denotes the space of the "magnetic" potentials A_i^a (the space of connections in some principal G bundle over $\mathbf{R}^n$), and $\Gamma(\mathbf{R}^n,\mathscr{L}(G))$ is the space of the "electrostatic" potentials A_0^a. Then, the minimal coupling consists simply in replacing the derivatives $\partial_{\mu}\varphi^{\alpha}$ with the covariant derivatives $\nabla_{\mu}\varphi^{\alpha}$, which are defined by

$$\nabla_{\mu}\varphi^{\alpha} = \partial_{\mu}\varphi^{\alpha} + A_{\mu}^{a}K_{a}^{\alpha}(\varphi). \tag{5.1}$$

To see that this is really a covariant derivative, consider an infinitesimal gauge transformation $u = u^aT_a$: $\mathbf{R}^{n+1} \to \mathscr{L}(G)$; we have

$$\varphi^{\alpha} \to \varphi^{\alpha} - u^aK_a{}^{\alpha}(\varphi), \tag{5.2a}$$

$$A_{\mu}^{a} \to A_{\mu}^{a} + \nabla_{\mu}u^a, \tag{5.2b}$$

and, therefore,

$$\nabla_{\mu}\varphi^{\alpha} \to \nabla_{\mu}\varphi^{\alpha} - u^a\partial_{\beta}K_a{}^{\alpha}\nabla_{\mu}\varphi^{\beta}.. \tag{5.3}$$

The total Lagrangian is $\mathscr{L}(\varphi,A) = \mathscr{L}_{\rm YM}(A) + \mathscr{L}_{\rm GNSM}(\varphi,A)$, with

$$\mathscr{L}_{\rm YM}(A) = -(1/4e^2)F_{\mu\nu}{}^aF^{\mu\nu a}, \tag{5.4}$$

$$\mathscr{L}_{\rm GNSM}(\varphi,A) = -(1/2f^2)h_{\alpha\beta}(\varphi)\nabla^\mu\varphi^\alpha\nabla_\mu\varphi^\beta. \tag{5.5}$$

The canonical momenta conjugate to A_i^a and φ^α are

$$P_a^i = \frac{\delta\mathscr{L}}{\delta\partial_0 A_i^a} = -\frac{1}{e^2}E_i^a, \tag{5.6}$$

$$\pi_\alpha = \frac{\delta\mathscr{L}}{\delta\partial_0\varphi^\alpha} = \frac{1}{f^2}h_{\alpha\beta}\nabla_0\varphi^\beta, \tag{5.7}$$

where $E_i^a = F_{i0}^a = -\partial_0 A_i^a + \nabla_i A_0^a$. The canonical momenta conjugate to A_0^a vanish:

$$P_a^0 = \frac{\delta\mathscr{L}}{\delta\partial_0 A_0^a} = 0; \tag{5.8}$$

these are the primary constraints of the theory. (Note that, with our metric, $P_i^a = P_a^i$ and $P_0^a = -P_a^0$.) The canonical Hamiltonian

$$H_C = \int d\mathbf{x}\,[\pi_\alpha\partial_0\varphi^\alpha + P_a^i\partial_0 A_i^a - \mathscr{L}]$$

can be written in the form

$$H_C = H_0 - \int d\mathbf{x}\, A_0^a G_a, \tag{5.9}$$

where

$$H_0 = \int d\mathbf{x}\left[\frac{e^2}{2}P_a^iP_a^i + \frac{1}{4e^2}F_{ij}^aF_{ij}^a + \frac{f^2}{2}h^{\alpha\beta}\pi_\alpha\pi_\beta + \frac{1}{2f^2}h_{\alpha\beta}\nabla_i\varphi^\alpha\nabla_i\varphi^\beta\right] \tag{5.10}$$

and

$$G_a = \nabla_i P_a^i + K_a{}^\alpha\pi_\alpha. \tag{5.11}$$

The canonical Poisson brackets are (2.7) and

$$\{A_0^a(\mathbf{x}),A_0^b(\mathbf{y})\} = 0, \tag{5.12a}$$

$$\{A_0^a(\mathbf{x}),P_b^0(\mathbf{y})\} = \delta_b^a\delta^{(n)}(\mathbf{x}-\mathbf{y}), \tag{5.12b}$$

$$\{P_a^0(\mathbf{x}),P_b^0(\mathbf{y})\} = 0, \tag{5.12c}$$

$$\{A_i^a(\mathbf{x}),A_j^b(\mathbf{y})\} = 0, \tag{5.12d}$$

$$\{A_i^a(\mathbf{x}),P_b^j(\mathbf{y})\} = \delta_b^a\delta^j_i\delta^{(n)}(\mathbf{x}-\mathbf{y}), \tag{5.12e}$$

$$\{P_a^i(\mathbf{x}),P_b^j(\mathbf{y})\} = 0. \tag{5.12f}$$

With these, we can derive Hamilton's equations:

$$\frac{dA_i^a}{dt} = \{A_i^a,H_C\} = e^2P_i^a + \nabla_i A_0^a, \tag{5.13a}$$

$$\frac{dP_i^a}{dt} = \{P_i^a,H_C\} = \frac{1}{e^2}(\partial_k F_{ki}^a - f_{acd}A_j^cF_{ij}^d) - \frac{1}{f^2}h_{\alpha\beta}\nabla_i\varphi^\beta K_a^\alpha - f_{abc}A_0^bP_i^c, \tag{5.13b}$$

$$\frac{d\varphi^\alpha}{dt} = \{\varphi^\alpha,H_C\} = f^2h^{\alpha\beta}\pi_\beta - A_0^bK_b{}^\alpha, \tag{5.14a}$$

$$\begin{aligned}\frac{d\pi_\gamma}{dt} &= \{\pi_\gamma,H_C\}\\ &= -\frac{f^2}{2}\partial_\gamma h^{\alpha\beta}\pi_\alpha\pi_\beta\\ &\quad -\frac{1}{2f^2}\partial_\gamma h_{\alpha\beta}\nabla_i\varphi^\alpha\nabla_i\varphi^\beta + \frac{1}{f^2}\partial_i(h_{\gamma\beta}\nabla_i\varphi^\beta)\\ &\quad -\frac{1}{f^2}h_{\alpha\beta}\partial_\gamma K_a{}^\alpha A_i^a\nabla_i\varphi^\beta + eA_0^a\partial_\gamma K_a{}^\beta\pi_\beta.\end{aligned} \tag{5.14b}$$

Equations (5.13a) and (5.14a) simply reproduce the relations (5.6) and (5.7) between velocities and momenta. From Eqs. (5.13), we get

$$-\nabla_\mu F^{\mu ia} + \frac{e^2}{f^2}K_\beta^a\nabla_i\varphi^\beta = 0, \tag{5.15}$$

which is the ith component of the YM equations. Similarly, from Eqs. (5.14), we get, after contracting with $h^{\alpha\beta}$ and after some algebra,

$$\partial_\mu\nabla^\mu\varphi^\alpha + \Gamma^\alpha_{\beta\gamma}\nabla_\mu\varphi^\beta\nabla^\mu\varphi^\gamma + A_\mu^a\partial_\alpha K_a{}^\beta\nabla^\mu\varphi^\alpha = 0, \tag{5.16}$$

which is just the equation one gets by varying the Lagrangian (5.5) with respect to φ^α. It can be shown that this equation is equivalent to the statement that the current of the scalar fields

$$J_a^\mu = -\frac{\delta\mathscr{L}}{\delta A_\mu^a} = \frac{1}{f^2}K_{a\alpha}\nabla^\mu\varphi^\alpha \tag{5.17}$$

is covariantly conserved.[4] Therefore, (5.16) is a consequence of the YM equations.

The time component of the YM equations (Gauss' law), given by

$$0 = -\nabla_\mu F^{\mu 0a} + e^2J^{0a} = \nabla_i E_i^a - \frac{e^2}{f^2}K^a{}_\alpha\nabla_0\varphi^\alpha, \tag{5.18}$$

is not obtained in the Hamiltonian formalism as an equation of motion, but rather as a constraint. In fact, if we define as usual the primary Hamiltonian by

$$H_P = H_C + \int d\mathbf{x}\,\mu^aP_a^0,$$

then consistency of the primary constraints under time evolution requires that

$$0 \approx \frac{dP_a^0}{dt} = \{P_a^0,H_P\} = G_a. \tag{5.19}$$

Therefore, G_a are the secondary constraints. It is easily seen using Eqs. (5.11), (5.6), and (5.7) that the constraint (5.19) is equivalent to the Lagrangian equation (5.18). The algebra of the constraints is given by (5.12c) and by

$$\{G_a(\mathbf{x}),P_b^0(\mathbf{y})\} = 0, \tag{5.20}$$

$$\{G_a(\mathbf{x}),G_b(\mathbf{y})\} = f_{ab}{}^cG_c(\mathbf{x})\delta^{(n)}(\mathbf{x}-\mathbf{y}). \tag{5.21}$$

A rather lengthy calculation which makes repeated use of Eqs. (2.4) and (2.5) shows that

$$\{G_a,H_0\} = 0, \tag{5.22}$$

so, using (5.20) and (5.21),

$$\frac{dG_a}{dt} = \{G_a, H_P\} = \int d\mathbf{x}\, \mu^{\,b} f_{ab}{}^c G_c \approx 0. \tag{5.23}$$

This shows that there are no further constraints, and that all constraints are first class.

For every time-independent infinitesimal gauge transformation u: $\mathbf{R}^n \to \mathscr{L}(G)$, we define a functional G_u on $T^*(\mathscr{C} \times \Gamma(\mathbf{R}^n, G/H))$ by

$$G_u(A^i_a, P^a_i, \varphi^\alpha, \pi_\alpha) = \int d\mathbf{x}\, u^a(\mathbf{x}) G_a(\mathbf{x}).$$

From (5.21), we find

$$\{G_u, G_v\} = G_{[u,v]}. \tag{5.24}$$

The nonvanishing Poisson brackets of G_u with the canonical variables are

$$\{G_u, A^a_i(\mathbf{x})\} = \nabla_i u^a(\mathbf{x}), \tag{5.25a}$$

$$\{G_u, P^i_a(\mathbf{x})\} = -f_{abc} u^b(\mathbf{x}) P^i_c(\mathbf{x}), \tag{5.25b}$$

$$\{G_u, \varphi^\alpha(\mathbf{x})\} = -u^a(\mathbf{x}) K^\alpha_a(\varphi(\mathbf{x})), \tag{5.25c}$$

$$\{G_u, \pi_\alpha(\mathbf{x})\} = u^b(\mathbf{x}) \partial_\alpha K^\beta_b(\varphi(\mathbf{x})) \pi_\beta(\mathbf{x}). \tag{5.25d}$$

Therefore, G_u generates the infinitesimal gauge transformation with parameter u.

Having found all constraints, we define the total Hamiltonian

$$H_T = H_0 + \int d\mathbf{x} (\mu^a P^0_a + v^a G_a), \tag{5.26}$$

where μ^a and v^a are Lagrange multipliers, and we have reabsorbed the last term of the canonical Hamiltonian (5.9) in the definition of v^a. Since all constraints are first class, the conditions that they be preserved by the time evolution generated by H_T does not determine any of the Lagrange multipliers. In order to have a uniquely defined dynamics, it is necessary to fix a gauge condition for each constraint. Further analysis requires that we distinguish between the generators of $\mathscr{L}(G)$, which are in $\mathscr{L}(H)$, from those in the complementary space $\mathscr{P}$ [see (2.2) and following]. We choose the unitary gauge, in which the scalar field is constant. Without loss of generality, we can assume that this constant is the origin of G/H:

$$\varphi^\alpha(\mathbf{x}) = 0. \tag{5.27}$$

Looking at the form of Eq. (5.7), it is seen that the preservation of this constraint under the time evolution generated by H_C requires that

$$\phi_\alpha = \pi_\alpha - (1/f^2) A^a_0 K_{a\alpha} = 0. \tag{5.28}$$

This suggests that we should take $\phi_\alpha = 0$ as gauge conditions on the momenta supplementing the conditions $\varphi^\alpha = 0$ on the coordinates. As is well known from the Lagrangian treatment of the theory, the unitary gauge condition is not sufficient to provide a complete gauge fixing; the subgroup H remains unbroken. In fact, we have given so far only $2d$ gauge conditions; in order to have a complete gauge fixing, it is necessary to fix further 2 dim H constraints $\chi_{\hat{a}}$ and $\psi_{\hat{a}}$. We will not need to specify these functions at all; it suffices to observe that we could choose them to be functions of $A^{\hat{a}}_0$, $A^{\hat{a}}_i$, $P^0_{\hat{a}}$, and $P^i_{\hat{a}}$ only.

The Poisson brackets of the gauge conditions (5.27) and (5.28) with the constraints are

$$\{P^0_a(\mathbf{x}), \varphi^\alpha(\mathbf{y})\} = 0, \tag{5.29a}$$

$$\{P^0_a(\mathbf{x}), \phi_\alpha(\mathbf{y})\} = (1/f^2) K^a_\alpha(\varphi(\mathbf{x})) \delta^{(n)}(\mathbf{x}-\mathbf{y}), \tag{5.29b}$$

$$\{G_a(\mathbf{x}), \varphi^\alpha(\mathbf{y})\} = -K^\alpha_a(\varphi(\mathbf{x})) \delta^{(n)}(\mathbf{x}-\mathbf{y}), \tag{5.29c}$$

$$\{G_a(\mathbf{x}), \phi_\alpha(\mathbf{y})\} = (\partial_\alpha K^\beta_a \phi_\beta(\mathbf{x}) - (1/f^2) f_{ab}{}^c A^b_0 K_{c\beta}) \delta^{(n)}(\mathbf{x}-\mathbf{y}). \tag{5.29d}$$

Note that (5.29c) is weakly zero when $a = \hat{a}$, since $K^\alpha_{\hat{a}}(o) = 0$, and is weakly nondegenerate when $a = \bar{a}$. Thus, on the subspace defined by all constraints except $\chi_{\hat{a}}$ and $\psi_{\hat{a}}$, we have the following matrix of Poisson brackets of constraints [a common factor $\delta^{(n)}(\mathbf{x}-\mathbf{y})$ has been dropped for notational simplicity]:

	$P^0_{\hat{b}}$ ↓	$G_{\hat{b}}$ ↓	$P^0_{\bar{b}}$ ↓	$G_{\bar{b}}$ ↓	φ^β ↓	ϕ_β ↓	$\chi_{\hat{b}}$ ↓	$\psi_{\hat{b}}$ ↓
$P^0_{\hat{a}} \to 0$	0	0	0	0	0	0	$\times$	$\times$
$G_{\hat{a}} \to 0$	0	0	0	0	0	$-(1/f^2) f_{\hat{a}b}{}^c A^b_0 K_{c\beta}$	$\times$	$\times$
$P^0_{\bar{a}} \to 0$	0	0	0	0	0	$(1/f^2) K_{\bar{a}\beta}$	0	0
$G_{\bar{a}} \to 0$	0	0	0	0	$eK_{\bar{a}}{}^\beta$	$-(1/f^2) f_{\bar{a}b}{}^c A^b_0 K_{c\beta}$	$\times$	$\times$
$\varphi^\alpha \to 0$	0	0	$-eK_{\bar{b}}{}^\alpha$	0	δ^α_β	0	0	0
$\phi_\alpha \to 0$	$(1/f^2) f_{\hat{b}c}{}^d A^c_0 K_{d\alpha}$	$(1/f^2) K_{\bar{b}\alpha}$	$(1/f^2) f_{\bar{b}c}{}^d A^c_0 K_{d\alpha}$	δ^β_α	$(1/f^2) A^a_0 (\partial_\alpha K_{a\beta} - \partial_\beta K_{a\alpha})$	0	0	0
$\chi_{\hat{a}} \to \times$	$\times$	0	$\times$	0	0	$\times$	$\times$	
$\psi_{\hat{a}} \to \times$	$\times$	0	$\times$	0	0	$\times$	$\times$	

where $\times$ means that the Poisson bracket will be generally nonzero but its explicit form is not known until $\chi_{\hat{a}}$ and $\psi_{\hat{a}}$ are given. If these constraints are good gauge conditions for the subgroup H, then the determinant of the 4×4 matrices in the upper right and lower left corners are nonzero. In this case, also the determinant of the full matrix is nonzero. This shows that φ^α and ϕ_α are good gauge conditions. The Lagrange multipliers μ^a and v^a can be determined by requiring that the gauge conditions be preserved by the time evolution generated by H_T. We have

$$0=\frac{d\varphi^{\alpha}}{dt}=\{\varphi^{\alpha},H_T\}\approx f^2h^{\alpha\beta}\pi_\beta+\nu^{\bar{b}}K_{\bar{b}}{}^{\alpha}(o), \qquad (5.30a)$$

$$0=\frac{d\phi_\alpha}{dt}=\{\phi_\alpha,H_T\}\approx-\frac{1}{f^2}K_{\bar{b}\alpha}(o)\partial_iA^{\bar{b}}_i-\frac{1}{f^2}\mu^{\bar{a}}K_{\bar{a}\alpha}(o)$$
$$+\frac{1}{f^2}\nu^{\hat{a}}f_{\hat{a}\bar{b}}{}^{\bar{a}}A^{\bar{b}}_0K_{\bar{a}\alpha}(o), \qquad (5.30b)$$

$$0=\frac{d\chi_{\hat{a}}}{dt}=\{\chi_{\hat{a}},H_T\}\approx\{\chi_{\hat{a}},H_0\}+\int d\mathbf{x}\,\mu^{\bar{b}}\{\chi_{\hat{a}},P^0_{\bar{b}}\}$$
$$+\int d\mathbf{x}\,\nu^b\{\chi_{\hat{a}},G_b\}, \qquad (5.30c)$$

$$0=\frac{d\psi_{\hat{a}}}{dt}=\{\psi_{\hat{a}},H_T\}\approx\{\psi_{\hat{a}},H_0\}+\int d\mathbf{x}\,\mu^{\hat{b}}\{\psi_{\hat{a}},P^0_{\hat{b}}\}$$
$$+\int d\mathbf{x}\,\nu^b\{\psi_{\hat{a}},G_b\}. \qquad (5.30d)$$

From these equations, one can in principle determine the Lagrange multipliers. In particular, we have

$$\nu^{\bar{a}}=-A^{\bar{a}}_0, \qquad (5.31a)$$

$$\mu^{\bar{a}}=-\partial_iA^{\bar{a}}_i+f_{\bar{a}\hat{b}\bar{c}}\nu^{\hat{b}}A^{\bar{c}}_0. \qquad (5.31b)$$

The Lagrange multipliers $\mu^{\hat{a}}$ and $\nu^{\hat{a}}$ cannot be determined unless $\chi_{\hat{a}}$ and $\psi_{\hat{a}}$ are explicitly known. Using Eqs. (5.31), the total Hamiltonian can be rewritten in the form

$$H_T=H_0-\int d\mathbf{x}\,\partial_iA^{\bar{a}}_iP^0_{\bar{a}}-\int d\mathbf{x}\,A^{\bar{a}}_0G_{\bar{a}}$$
$$+\int d\mathbf{x}\,\mu^{\hat{a}}P^0_{\hat{a}}+\int d\mathbf{x}\,\nu^{\hat{a}}\widetilde{G}_{\hat{a}}, \qquad (5.32)$$

where

$$\widetilde{G}_{\hat{a}}=G_{\hat{a}}+f_{\hat{a}\bar{b}}{}^{\bar{c}}A^{\bar{b}}_0P^0_{\bar{c}} \qquad (5.33)$$

are first-class constraints satisfying the algebra

$$\{\widetilde{G}_{\hat{a}}(\mathbf{x}),\widetilde{G}_{\hat{b}}(\mathbf{y})\}=f_{\hat{a}\hat{b}}{}^{\hat{c}}\widetilde{G}_{\hat{c}}(\mathbf{x})\delta^{(n)}(\mathbf{x}-\mathbf{y}). \qquad (5.34)$$

At this point, it can easily be seen that the GNSM is equivalent to a massive YM theory (the constraint analysis of the massive YM theory is given in the Appendix). Consider the subspace $\mathscr{N}$ of the phase space, which is defined by the constraints (5.27) and (5.28). Clearly, the map $T^*\mathscr{C}\to\mathscr{N}$ defined by $(A^a_0,A^a_i,P^0_a,P^i_a)\mapsto(A^a_0,A^a_i,P^0_a,P^i_a,0,(1/f^2)\times K_{ba}A^b_0)$ is a diffeomorphism. Therefore, $\mathscr{N}$ can be identified with the phase space of a YM theory. Then, it can easily be seen that on $\mathscr{N}$

$$K_{\hat{a}}{}^{\alpha}\pi_\alpha\sim 0, \qquad (5.35a)$$

$$K_{\bar{a}}{}^{\alpha}\pi_\alpha\sim(1/f^2)A^{\bar{a}}_0, \qquad (5.35b)$$

[where $\sim$ means that only the constraints (5.27) and (5.28) are being used] and, therefore, the Gauss law constraints (5.19) are weakly equivalent to the constraints (A7). Furthermore,

$$\frac{f^2}{2}h^{\alpha\beta}\pi_\alpha\pi_\beta+\frac{1}{2f^2}h_{\alpha\beta}(\varphi)\nabla_i\varphi^{\alpha}\quad\nabla_i\varphi^{\beta}$$
$$\sim\frac{1}{2f^2}(A^{\bar{a}}_0A^{\bar{a}}_0+A^{\bar{a}}_iA^{\bar{a}}_i). \qquad (5.36)$$

So, if we identify $M=f^{-1}$, the restriction of the total Hamiltonian (5.32) to $\mathscr{N}$ is identical to the total Hamiltonian (A15). In conclusion, the restriction of the GNSM to the submanifold $\mathscr{N}$ has the same constraints and the same Hamiltonian as a massive YM theory, and, therefore, the two theories are equivalent.

VI. OTHER FORMULATIONS OF THE GAUGED SIGMA MODEL

In the previous section, the scalar sector of the GNSM was presented in the intrinsic formulation. This has the advantage of minimizing the number of the constraints, since all constraints come from the gauging of the group G and, hence, are equal in number to the constraints of a pure YM theory. On the other hand, there may be reasons to prefer other formulations for the scalar sector. In the resulting formulations of the GNSM, the constraints that appeared in Secs. III and IV are present in addition to those discussed in Sec. V. There is not much interference between these sets of constraints, so we give only a brief discussion of these formulations. The interested reader can easily work out all the details following the pattern of the previous sections.

The GNSM in the embedded formulation is obtained by replacing the fields φ^{α} of Sec. V with the fields ϕ^m and λ^a of Sec. III. The configuration space is $\mathscr{Q}'_{\rm GNSM}=\mathscr{C}\times\Gamma(\mathbf{R}^n,\mathscr{L}(G))\times\Gamma(\mathbf{R}^n,\mathbf{R}^N)\times\Gamma(\mathbf{R}^n,\mathbf{R}^{N-d})$. Instead of (5.5), we have

$$\mathscr{L}_{\rm GNSM}(A,\phi)=-\frac{1}{2f^2}\nabla^{\mu}\phi^m\nabla_\mu\phi^m$$
$$+\sum_{a=1}^{N-d}\lambda^aZ^a(\phi), \qquad (6.1)$$

where

$$\nabla_\mu\phi^m=\partial_\mu\phi^m+A^a_\mu(T_a)^m{}_n\phi^n. \qquad (6.2)$$

When this covariant derivative acts on fields satisfying the constraints $Z^a(\phi)=0$, it can be identified with the covariant derivative (5.1). More precisely, using (3.2), we have

$$\nabla_\mu(j\circ\varphi)^m=\frac{\partial j^m}{\partial y^{\alpha}}\nabla_\mu\varphi^{\alpha}. \qquad (6.3)$$

The canonical momenta are

$$P^0_a=\frac{\delta\mathscr{L}}{\delta\partial_0A^a_0}=0, \qquad (6.4)$$

$$P^i_a=\frac{\delta\mathscr{L}}{\delta\partial_0A^a_i}=-\frac{1}{e^2}E^a_i, \qquad (6.5)$$

$$\Pi_m=\frac{\delta\mathscr{L}}{\delta\partial_0\phi^m}=\frac{1}{f^2}\nabla_0\phi^m, \qquad (6.6)$$

$$\Lambda_a=\frac{\delta\mathscr{L}}{\delta\partial_0\lambda^a}=0. \qquad (6.7)$$

The canonical Hamiltonian is

$$H_C = H_0 + \int d\mathbf{x}\, A_0^a G_a - \int d\mathbf{x}\, \lambda^a Z^a, \tag{6.8}$$

where

$$H_0 = \int d\mathbf{x}\left[\frac{e^2}{2} P^i_a P^i_a + \frac{1}{e^2} F^a_{ij} F^a_{ij} + \frac{f^2}{2}\Pi_m\Pi_m + \frac{1}{2f^2}\nabla_i\phi^m\nabla_i\phi^m\right] \tag{6.9}$$

and

$$G_a = \nabla_i P^i_a + (T_a)^m{}_n\phi^n\Pi_m. \tag{6.10}$$

The primary constraints of this theory are (6.4) and (6.7). Upon defining as usual the primary Hamiltonian H_P, we get, from (6.4), the secondary constraints

$$G_a = 0, \tag{6.11}$$

while from (6.7), we get the sequence of constraints

$$0 \approx \psi^a_2 = Z^a(\phi), \tag{6.12a}$$

$$0 \approx \psi^a_3 = f^2\Pi_m \frac{\partial Z^a}{\partial\phi^m}, \tag{6.12b}$$

$$0 \approx \psi^a_4 = f^2 M^{ab}\lambda^b + \left(f^4\frac{\partial^2 Z^a}{\partial\phi^m\,\partial\phi^n}\Pi_m\Pi_n + \frac{\partial Z^a}{\partial\phi^n}\nabla_i\nabla_i\phi^n\right), \tag{6.12c}$$

where M^{ab} is as in Sec. III. It is easily seen that the constraints (6.7) and (6.12) are all second class. The constraint (6.4) is obviously first class. As for (6.11), consider the Poisson brackets with the constraints (6.12): Using Eq. (3.4), we get

$$\{G_a,\psi^b_2\} = -(T_a)^m{}_n\phi^n\frac{\partial}{\partial\phi^m}Z^b = 0. \tag{6.13}$$

Using the Jacobi identity and (6.13), we find

$$\{G_a,\psi^b_3\} = \{G_a,\{\psi^b_2,H_P\}\} = -\{\psi^b_2,\{H_P,G_a\}\} - \{H_P,\{G_a,\psi^b_2\}\} = 0. \tag{6.14}$$

Similarly, using the Jacobi identity and (6.14),

$$\{G_a,\psi^b_4\} = \{G_a,\{\psi^b_3,H_P\}\} = -\{\psi^b_3,\{H_P,G_a\}\} - \{H_P,\{G_a,\psi^b_3\}\} = 0. \tag{6.15}$$

Therefore, also G_a are first class. Using the same methods as in the end of Sec. III, it is then possible to prove that the embedded formulation of the GNSM is equivalent to the ones presented in Sec. V.

The GNSM in lifted formulation is obtained by replacing the fields φ^α of Sec. V with the fields $\overline{\varphi}^{\overline{\alpha}}$ as in Sec. IV. The canonical configuration space is $\overline{\mathscr{Q}}_{\rm GNSM} = \mathscr{C}\times\Gamma(\mathbf{R}^n,\mathscr{L}(G))\times\Gamma(\mathbf{R}^n,G/\overline{H})$. Instead of (5.5), we have

$$\mathscr{L}_{\rm GNSM}(A,\overline{\varphi}) = -(1/2f^2)\bar{h}_{\overline{\alpha}\overline{\beta}}\mathscr{D}_\mu\overline{\varphi}^{\overline{\alpha}}\mathscr{D}^\mu\overline{\varphi}^{\overline{\beta}}, \tag{6.16}$$

where $\mathscr{D}_\mu\overline{\varphi}^{\overline{\alpha}}$ is the "doubly covariant" derivative, defined by

$$\mathscr{D}_\mu\overline{\varphi}^{\overline{\alpha}} = D_\mu\overline{\varphi}^{\overline{\alpha}} + A^a_\mu\overline{K}_a{}^{\overline{\alpha}}(\overline{\varphi}). \tag{6.17}$$

The momenta conjugate to A^a_i and $\overline{\varphi}^{\overline{\alpha}}$ are

$$P^i_a = \frac{\delta\mathscr{L}}{\delta\partial_0 A^a_i} = -\frac{1}{e^2}E^a_i, \tag{6.18}$$

$$\overline{\pi}_{\overline{\alpha}} = \frac{\delta\mathscr{L}}{\delta\partial_0\overline{\varphi}^{\overline{\alpha}}} = \frac{1}{f^2}\bar{h}_{\overline{\alpha}\overline{\beta}}\mathscr{D}_0\overline{\varphi}^{\overline{\beta}}. \tag{6.19}$$

The primary constraints are

$$P^0_a = \frac{\delta\mathscr{L}}{\delta\partial_0 A^a_0} \approx 0, \tag{6.20}$$

$$\psi_{\bar{a}} = F_{\bar{a}}{}^{\overline{\alpha}}\overline{\pi}_{\overline{\alpha}} \approx 0, \tag{6.21}$$

and the canonical Hamiltonian is

$$H_C = H_0 + \int d\mathbf{x}\, B^{\bar{a}}_0\psi_{\bar{a}} - \int d\mathbf{x}\, A^a_0 G_a, \tag{6.22}$$

where $B^{\bar{a}}_\mu$ is as in (4.3), and

$$H_0 = \int d\mathbf{x}\left[\frac{e^2}{2}P^i_a P^i_a + \frac{1}{4e^2}F^a_{ij}F^a_{ij} + \frac{f^2}{2}\bar{h}^{\overline{\alpha}\overline{\beta}}\overline{\pi}_{\overline{\alpha}}\overline{\pi}_{\overline{\beta}} + \frac{1}{2f^2}\bar{h}_{\overline{\alpha}\overline{\beta}}\mathscr{D}_i\overline{\varphi}^{\overline{\alpha}}\mathscr{D}_i\overline{\varphi}^{\overline{\beta}}\right], \tag{6.23}$$

and

$$G_a = \nabla_i P^i_a + e\overline{K}_a{}^{\overline{\alpha}}\overline{\pi}_{\overline{\alpha}}. \tag{6.24}$$

Following, once again, the standard procedure, one finds that (6.20) gives rise to the secondary constraints $G_a = 0$, whereas (6.21) does not give rise to any secondary constraints.

As in Secs. IV and V, it can be checked that the constraints generate the gauge transformations for the groups G and K. The nonvanishing Poisson brackets of the constraints are as in (5.21) and (4.9) and, therefore, all constraints are first class.

The equivalence of this formulation to the intrinsic one can be shown in the same way as in Sec. IV.

VII. CONCLUDING REMARKS

In this paper, we have considered the local dynamics of the NSM without paying much attention to global properties. We would like to conclude with some remarks on this point.

First of all, we observe that even in the intrinsic formulation the fields $\varphi^\alpha(\mathbf{x})$ are only defined when the image of the map φ lies within the domain of the chosen coordinate system on N. In general, several charts will be needed to coordinatize N (e.g., at least two if N is a sphere), and questions of compatibility could, in principle, arise. However, for the classical case considered in this paper, this poses no real problem. Indeed, it is a property of the tensorial formalism that we have been employing, that all the formulas we have written are the local coordinate expression of globally valid ones.

At the quantum level, however, the problem is more subtle. As is well known, even for a one-dimensional quantum mechanical system (corresponding to the case when M

is a point and $N = \mathbf{R}$), coordinate transformations lead to unitarily inequivalent quantum theories. Therefore, it is doubtful that a naive quantization procedure based on replacing Poisson brackets (or Dirac brackets) of the fields φ^α by commutators could reflect the global properties of the classical theory correctly. An attempt to overcome these difficulties has recently been proposed.[13]

By the same token, even if we were successful in quantizing the different formulations of the NSM, which we have discussed, they should not be *a priori* expected to yield unitarily equivalent theories, and the NSM in embedded formulation should not *a priori* be expected to be equivalent to a limiting case of a Higgs model (for a discussion of this point, see Ref. 14).

Finally, let us note that in our discussion the topology, and even the dimension, of the space M and the internal manifold N were largely immaterial. For the sake of generality, we have chosen to work with a Lagrangian that does not depend on these choices. However, the NSM exhibits rich topological structures, which depend crucially on the dimension. These properties are reflected by the presence of "topological terms" in the Lagrangian, which, in turn, give rise, in the Hamiltonian treatment of the theory, to modifications of the canonical symplectic structure. A Hamiltonian analysis of these topological properties has to be done on a case by case basis (see, for example, Ref. 15).

APPENDIX: MASSIVE YANG-MILLS THEORY

We consider a YM theory with gauge group G such that the components of the gauge potentials in the subspace $\mathscr{P}$, defined as in (2.2), have mass M. The Lagrangian is

$$\mathscr{L}_{\rm MYM}(A) = -(1/4e^2)F^a_{\mu\nu}F^{\mu\nu a} - \tfrac{1}{2}M^2 A^{\bar a}_\mu A^{\mu\bar a}. \tag{A1}$$

The case when all components are massive can be recovered by letting $H = \{e\}$ and $\mathscr{P} = \mathscr{L}(G)$ (Ref. 12). The canonical momenta are

$$P^0_a = \frac{\delta\mathscr{L}}{\delta\partial_0 A^a_0} = 0, \tag{A2}$$

$$P^i_a = \frac{\delta\mathscr{L}}{\delta\partial_0 A^a_i} = -\frac{1}{f^2}E^a_i. \tag{A3}$$

Equation (A2) gives the primary constraints. The canonical Hamiltonian is

$$H_C = H_0 - \int d\mathbf{x}\, A^a_0 G_a, \tag{A4}$$

where

$$H_0 = \int d\mathbf{x}\left[\frac{e^2}{2}P^i_a P^i_a + \frac{1}{4e^2}F^a_{ij}F^a_{ij} + \frac{1}{2}M^2(A^{\bar a}_i A^{\bar a}_i + A^{\bar a}_0 A^{\bar a}_0)\right], \tag{A5}$$

and

$$G_{\hat a} = \nabla_i P^i_{\hat a}, \tag{A6a}$$

$$G_{\bar a} = \nabla_i P^i_{\bar a} + M^2 A^{\bar a}_0. \tag{A6b}$$

From the consistency of the primary constraints (A2), we get the secondary constraints

$$G_a = 0. \tag{A7}$$

The nonvanishing Poisson brackets of the constraints are

$$\{P^0_{\bar a}(\mathbf{x}), G_{\bar b}(\mathbf{y})\} = -M^2\delta_{\bar a\bar b}\delta^{(n)}(\mathbf{x}-\mathbf{y}), \tag{A8a}$$

$$\{G_{\hat a}(\mathbf{x}), G_{\hat b}(\mathbf{y})\} = f_{\hat a\hat b}{}^{\hat c}G_{\hat c}(\mathbf{x})\delta^{(n)}(\mathbf{x}-\mathbf{y}), \tag{A8b}$$

$$\{G_{\hat a}(\mathbf{x}), G_{\bar b}(\mathbf{y})\} = f_{\hat a\bar b}{}^{\bar c}(G_{\bar c}(\mathbf{x}) - M^2 A^{\bar c}_0(\mathbf{x}))\delta^{(n)}(\mathbf{x}-\mathbf{y}), \tag{A8c}$$

$$\{G_{\bar a}(\mathbf{x}), G_{\bar b}(\mathbf{y})\} = [f_{\bar a\bar b}{}^{\hat c}G_{\hat c}(\mathbf{x}) + f_{\bar a\bar b}{}^{\bar c}(G_{\bar c}(\mathbf{x}) - M^2 A^{\bar c}_0(\mathbf{x}))]\delta^{(n)}(\mathbf{x}-\mathbf{y}). \tag{A8d}$$

We define the total Hamiltonian

$$H_T = H_0 + \int d\mathbf{x}\,\mu^a P^0_a + \int d\mathbf{x}\,\nu^a G_a, \tag{A9}$$

where we have reabsorbed $-A^a_0$ in the Lagrange multiplier ν^a. Consistency of the constraints requires that

$$0 = \frac{dP^0_{\hat a}}{dt} = \{P^0_{\hat a}, H_T\} \approx 0, \tag{A10a}$$

$$0 = \frac{dP^0_{\bar a}}{dt} = \{P^0_{\bar a}, H_T\} \approx -M^2 A^{\bar a}_0 - M^2\nu^{\bar a}, \tag{A10b}$$

$$0 = \frac{dG_{\hat a}}{dt} = \{G_{\hat a}, H_T\} \approx -M^2\nu^{\bar b}f_{\hat a\bar b}{}^{\bar c}A^{\bar c}_0, \tag{A10c}$$

$$0 = \frac{dG_{\bar a}}{dt} = \{G_{\bar a}, H_T\} \approx M^2\partial_i A^{\bar a}_i + M^2\mu^{\bar a} - M^2\nu^b f_{\bar a b\bar c}A^{\bar c}_0. \tag{A10d}$$

This shows that there are no further constraints. From (A8), we see that all the constraints are second class except for $P^0_{\hat a}$. Since the first-class constraints always occur in pairs and since the dimension of H may well be odd, one expects that there are linear combinations of the second-class constraints which are first class. In fact, consider the functions

$$\widetilde{G}_{\hat a} = G_{\hat a} + f_{\hat a\bar b}{}^{\bar c}A^{\bar b}_0 P^0_{\bar c}. \tag{A11}$$

From the Poisson brackets,

$$\{\widetilde{G}_{\hat a}(\mathbf{x}), P^0_{\hat b}(\mathbf{y})\} = 0, \tag{A12a}$$

$$\{\widetilde{G}_{\hat a}(\mathbf{x}), P^0_{\bar b}(\mathbf{y})\} = -f_{\hat a\bar b}{}^{\bar c}P^0_{\bar c}(\mathbf{x})\delta^{(n)}(\mathbf{x}-\mathbf{y}), \tag{A12b}$$

$$\{\widetilde{G}_{\hat a}(\mathbf{x}), \widetilde{G}_{\hat b}(\mathbf{y})\} = f_{\hat a\hat b}{}^{\hat c}\widetilde{G}_{\hat c}(\mathbf{x})\delta^{(n)}(\mathbf{x}-\mathbf{y}), \tag{A12c}$$

$$\{\widetilde{G}_{\hat a}(\mathbf{x}), \widetilde{G}_{\bar b}(\mathbf{y})\} = f_{\hat a\bar b}{}^{\bar c}G_{\bar c}(\mathbf{x})\delta^{(n)}(\mathbf{x}-\mathbf{y}), \tag{A12d}$$

one sees that the constraints (A12) are first class and pair up with the constraints $P^0_{\hat a}$. These constraints generate the gauge transformations for the unbroken group H. From (A10), we get relations that determine the Lagrange multipliers relative to the second-class constraints:

$$\nu^{\bar a} = -A^{\bar a}_0, \tag{A13a}$$

$$\mu^{\bar a} = -\partial_i A^{\bar a}_i + f_{\bar a\hat b\bar c}\nu^{\hat b}A^{\bar c}_0. \tag{A13b}$$

The total Hamiltonian can now be rewritten:

$$H_T = H_0 - \int d\mathbf{x}\, \partial_i A_i^{\bar{a}} P_a^0 - \int d\mathbf{x}\, A_0^{\bar{a}} G_{\bar{a}}$$

$$+ \int d\mathbf{x}\, \mu^{\hat{a}} P_{\hat{a}}^0 + \int d\mathbf{x}\, \nu^{\hat{a}} \widetilde{G}_{\hat{a}}. \tag{A14}$$

The Lagrange multipliers relative to the first-class constraints have remained undetermined. In order to fix completely the form of the Hamiltonian, it is necessary to choose 2 dim H gauge conditions $\chi_{\hat{a}}$ and $\psi_{\hat{a}}$.

[1] B. Lee, *Chiral Dynamics* (Gordon and Breach, New York, 1972).
[2] A. Polyakov, *Gauge fields and Strings* (Harwood Academic, Chur, 1987).
[3] S. J. Gates, M. T. Grisaru, M. Roček, and W. Siegel, *Superspace: Or One Thousand and One Lessons on Supersymmetry* (Benjamin Cummings, Reading, MA, 1983); M. Green, J. H. Schwartz, and E. Witten, *Superstring Theory* (Cambridge U. P., Cambridge, MA, 1987).
[4] R. Percacci, *Geometry of Nonlinear Field Theories* (World Scientific, Singapore, 1986).
[5] P. A. M. Dirac, "Lectures on quantum mechanics," Belfer Graduate School of Science Monograph series, Yeshiva University (1964); A. Hanson, T. Regge, and C. Teitelboim, *Constrained Hamiltonian Systems* (Accademia Nazionale dei Lincei, Rome, 1976).
[6] M. Veltman, Acta Phys. Pol. B **8**, 475 (1977).
[7] E. D'Hoker and E. Farhi, Nucl. Phys. B **248**, 59, 77 (1984); L. D. Faddeev and S. L. Shatashvili, Phys. Lett. B **167**, 225 (1986); R. Percacci and R. Rajaraman, *ibid.* **201**, 256 (1988); R. Percacci and R. Rajaraman, Int. J. Mod. Phys. A **4**, 4177 (1989).
[8] D. Popović, Phys. Rev. D **34**, 1764 (1986); J. Dell, J. L. de Lyra, and L. Smolin, Phys. Rev. D **34**, 3012 (1986).
[9] H. Sugawara, Phys. Rev. **170**, 1659 (1968).
[10] J. D. Moore and R. Schlafly, Math. Z. **173**, 119 (1980).
[11] I. Singer, Commun. Math. Phys. **60**, 7 (1978).
[12] P. Senjanović, Ann. Phys. **100**, 227 (1976).
[13] C. Isham, in *Relativity, Groups and Topology II*, edited by B. de Witt and R. Stora (Elsevier, New York, 1984).
[14] L. H. Chan, Phys. Rev. D **36**, 3755 (1987).
[15] E. Witten, Commun. Math. Phys. **92**, 455 (1984); N. K. Pak and R. Percacci, Phys. Rev. D **36**, 2420 (1987).

Singular indecomposable representations of sl(2,ℂ) and relativistic wave equations

A. J. Bracken and A. Cant[a)]
Department of Mathematics, The University of Queensland, Brisbane, Australia

(Received 21 February 1989; accepted for publication 28 June 1989)

A detailed summary is given of the structure of singular indecomposable representations of sl(2,ℂ), as developed by Gel'fand and Ponomarev [Usp. Mat. Nauk **23**, 3 (1968); translated in Russ. Math. Surveys **23**, 1 (1968)]. A variety of four-vector operators Γ_μ is constructed, acting within direct sums of such representations, including some with nonsingular Γ_0. Associated wave equations of Gel'fand–Yaglom type are considered that admit timelike solutions and lead to mass-spin spectra of the Majorana type. A subclass of these equations is characterized in an invariant way by obtaining basis-independent expressions for the commutator and anticommutator of Γ_μ and Γ_ν. A brief discussion is given of possible applications to physics of these equations and of others in which nilpotent scalar operators appear.

I. INTRODUCTION

Many authors have studied first-order linear relativistically invariant wave equations of the type

$$(\Gamma_\mu\,\partial^\mu + i\kappa)\psi(x) = 0, \tag{1.1}$$

in which the wave function ψ takes its values in a vector space V carrying a representation π of the Lorentz group SL(2,ℂ), the Γ_μ (for $\mu = 0,1,2,3$) and κ are linear operators on V, and $\partial^\mu = \partial/\partial x_\mu$. The first systematic treatment was that of Gel'fand and Yaglom,[1] who gave detailed formulas for the structure of possible Γ_μ in the case where π is a direct sum of irreducible representations of SL(2,ℂ). These representations may or may not be infinite dimensional.

The results of Ref. 1 were obtained in a particular basis for V, but various authors have later emphasized the importance of invariant properties of wave equations. These are the properties that do not depend on the choice of basis in V or on the corresponding explicit form of the Γ_μ and κ. In particular, starting with the early work on wave equations (see, for example, Lubanski[2] and Harish-Chandra[3]), there has been great interest in what we shall refer to as their *algebraic structure*, by which we mean especially the algebras generated by the vector operator Γ_μ. This involves, in particular, a description of the commutator $[\Gamma_\mu,\Gamma_\nu]$ and the anticommutator $\{\Gamma_\mu,\Gamma_\nu\}$. Representations of such algebras lead to entire families of equations, giving us a systematic way of classifying some of the vast number of relativistic wave equations. Lubanski concentrated on the case where the commutator $[\Gamma_\mu,\Gamma_\nu]$ is a nonzero multiple of the generator $J_{\mu\nu}$ of SL(2,ℂ), so that the complex Lie algebra generated by the Γ_μ is just so(5,ℂ). An analysis of the possible Lie algebras generated by the Γ_μ was later carried out by Cant and Hurst,[4] who showed that arbitrarily large simple Lie algebras can be obtained. This contradicted earlier claims that had been made (see, for example, Refs. 5 and 6). It was also shown in Ref. 4 how a knowledge of the Lie algebra can help in deriving the mass and spin spectra associated with Eq. (1.1). Bracken[7] also used algebraic properties to characterize a class of wave equations, and to determine the associated mass and spin spectra, in a study of the family with κ a multiple of the identity operator on V, and

$$\pi = [\tfrac{1}{2},l_1] \oplus [\tfrac{1}{2}, -l_1], \quad l_1 \in \mathbb{C}. \tag{1.2}$$

[We use the standard notation[1] for the irreducible representations of SL(2,ℂ).] In fact, it has been shown[8] that an infinite-dimensional Lie algebra is generated in this case, except when $l_1 = 0$, where we recover one of the Majorana equations, or $2(l_1 - 1) \in \mathbb{N}$, where π is finite dimensional and the Lie algebra generated is $\mathrm{sp}(2(l_1)^2 - \tfrac{1}{2})$. A further study of algebraic structure, concentrating on the role of real Lie algebras in wave equations, was carried out by Cant.[9]

Such algebraic properties will, we believe, be especially useful when one considers variations on the classical theme of wave equations based on direct sums of irreducible representations. For example, wave equations of the form (1.1), with the added property that V carries a representation of the larger group $\overline{\mathrm{SL}}(4,R)$, have been studied and applied to the problem of describing the gravitational interactions of hadrons.[10] Such equations are infinite dimensional and SL(2,ℂ) invariant, although not fully SL(4,R) invariant. The results of Ref. 9 were particularly useful in this work. Infinite-dimensional equations associated with representations of SO(4,2) have also been widely discussed in the literature.[11]

In this paper we shall be concerned with a different direction of generalization: we consider the case where π is a direct sum of reducible but *indecomposable* representations of the algebra[12] sl(2,ℂ). All such representations are infinite dimensional, and they can be very complicated objects. The first examples presented were the "expansors" described by Dirac,[13] who more recently emphasized the potential importance of indecomposable representations for physics.[14] A class of indecomposable representations was studied by Gel'fand and Ponomarev,[15] and divided into two subclasses, called singular and nonsingular. Bender and Griffiths,[16] in a study of the transformation properties of massless fields, examined the composition series for the tensor product of the

[a)] Present address: Department of Industry, Technology and Commerce, GPO Box 9839, Canberra, Australia.

 0022-2488/89/122963-10$02.50

four-vector representation of SL(2,C) with an infinite-dimensional irreducible representation, and found that it can contain indecomposable representations. Hlavaty and Niederle[17] applied the results of Ref. 15 in constructing some examples of wave equations based on indecomposable representations. Their work describes the general structure of Γ_μ associated with a representation π, which is a direct sum of nonsingular indecomposable representations, but their results are slight in the more complicated, and possibly more interesting case, when singular indecomposable representations are involved: there they gave only one example of a wave equation (1.1), associated with a direct sum of particularly simple singular indecomposable representations that are in fact operator irreducible. Operators Γ_μ associated with such representations, sometimes called "integer-point" representations in the literature,[18] had been constructed earlier by Ruhl.[19] Hlavaty *et al.* showed for their example that no timelike solutions exist.

A different approach to indecomposable representations of sl(2,C) and associated four-vector operators has been developed by Gruber and his associates.[20] To our knowledge, the relationship of the representations constructed there to those of Gel'fand and Ponomarev[15] has not been fully elucidated.

Our object in the present work is twofold. First, to summarize the results of Gel'fand and Ponomarev on singular indecomposable representations of sl(2,C), in a form more readily accessible to physicists, and second, to give some examples of wave equations based on such representations, especially ones that do admit timelike solutions, unlike the example given in Ref. 17.

We shall show further that a subclass of these equations can be characterized in an invariant way, at least partly, by virtue of the simple form taken by the commutator and anticommutator of Γ_μ and Γ_ν. This subclass is a direct generalization of that considered in Ref. 7, which includes one of Majorana's equations,[21] and indeed the subclass of wave equations we discuss does lead to mass-spin spectra of the Majorana type.

II. SINGULAR INDECOMPOSABLE REPRESENTATIONS OF sl(2,C)

Let $\mathscr{g}\cong \mathrm{sl}(2,\mathbb{C})$ denote the real Lie algebra of the homogeneous Lorentz group, and $\mathscr{k}\cong \mathrm{su}(2)$ its maximal compact subalgebra. We take the standard basis $\{h_1,h_2,h_3,f_1,f_2,f_3\}$ for (the complexification of) $\mathscr{g}$; $\{h_1,h_2,h_3\}$ for $\mathscr{k}$. The defining Lie product relations are

$$[h_p,h_q] = -[f_p,f_q] = i\epsilon_{pqr}h_r,$$
$$[h_p,f_q] = i\epsilon_{pqr}f_r, \tag{2.1}$$

where p,q,r run over 1,2,3, repeated subscripts are summed over those values, and ϵ_{pqr} is the usual alternating symbol. We work with h_3,f_3, $h_\pm = h_1 \pm ih_2$, and $f_\pm = f_1 \pm if_2$ in what follows. A representation τ of $\mathscr{g}$ is said to be $\mathscr{k}$ finite or a *Harish-Chandra representation* if in the direct sum decomposition of the restriction of τ to $\mathscr{k}$, equivalent irreducible representations of $\mathscr{k}$ occur with finite multiplicities only. Thus

$$\tau|_{\mathscr{k}} = \bigoplus_{l\in\mathscr{S}} \tau_l \tag{2.2}$$

where $\mathscr{S}$ is a subset of $\mathbb{N}/2\equiv\{0,\frac{1}{2},1,\frac{3}{2},...\}$, and each τ_l is the direct sum of a finite number of copies of the $(2l+1)$-dimensional irreducible representation φ_l of $\mathscr{k}$. Such a τ is called *indecomposable* if it cannot be decomposed into a direct sum of representations of $\mathscr{g}$. The indecomposable Harish-Chandra representations of $\mathscr{g}$ have been classified by Gel'fand and Ponomarev,[15] and we now summarize their results.

Let τ be such a representation, V the associated $\mathscr{g}$ module, and $H_3 = \tau(h_3)$, $F_3 = \tau(f_3)$, etc. The Casimir operators are given, as usual,[1] by

$$\Delta_1 = \tfrac{1}{2}(H_-F_+ + F_-H_+) + H_3F_3,$$
$$\Delta_2 = H_-H_+ - F_-F_+ + (H_3)^2 - (F_3)^2 + 2H_3, \tag{2.3}$$

and commute with H_3, $H_\pm$, F_3, and $F_\pm$ on V. If τ is in fact *irreducible*, then Δ_1 and Δ_2 are multiples of the identity operator on V. This is a necessary but not sufficient condition for subspace irreducibility. More generally, it is only true that Δ_1 and Δ_2 each have on V exactly one eigenvalue λ_1, λ_2, of the form

$$\lambda_1 = -il_0l_1, \quad \lambda_2 = l_0^2 + l_1^2 - 1, \tag{2.4}$$

with $l_0\in\mathbb{N}/2$ and $l_1\in\mathbb{C}$. If $l_1 - l_0$ is a nonzero integer, then τ is called singular; otherwise τ is nonsingular. The structure of nonsingular indecomposable representations of $\mathscr{g}$, as determined in Ref. 15, has been summarized by Hlavaty and Niederle.[17] In the present paper, we shall concentrate on the more complicated case of singular indecomposable representations.

Given such a representation τ then, with λ_1, λ_2, as in (2.4), we choose l_0 and l_1 without loss of generality, such that $0\leqslant l_0<|l_1|$. Then[15] the set $\mathscr{S}$ in (2.1) is given by $\mathscr{S} = \{l_0,l_0+1,l_0+2,...\}$. Corresponding to (2.2) we can write V as an algebraic direct sum

$$V\cong \bigoplus_{l\in\mathscr{S}} V_l, \tag{2.5}$$

where each V_l is an eigenspace of the Casimir operator $(H_3)^2 + H_-H_+ + H_3$ of $\mathscr{k}$, with eigenvalue $l(l+1)$. Each V_l can in turn be written as a direct sum

$$V_l\cong \bigoplus_{m=-l}^{l} V_{lm} \tag{2.6}$$

of eigenspaces V_{lm} of H_3 with eigenvalue $m\in\{l,l-1,..., -l\}$. The subspaces V_{lm} for $l = l_0,l_0+1,...,|l_1|-1$ all have dimension n_0, while the V_{lm} for $l = |l_1|, |l_1|+1,...$ all have dimension n_1, for some pair (n_0,n_1) of non-negative integers, not both of which are zero. If $n_1 = 0$ we must have $n_0 = 1$, in which case τ is the finite-dimensional irreducible representation labeled[1] $[l_0,l_1]$. If $n_0 = 0$ and $n_1 = 1$, τ is the infinite-dimensional irreducible representation $[|l_1|,\mathrm{sgn}(l_1)l_0]$, sometimes[18,19] called the "tail" of $[l_0,l_1]$. In all remaining cases, with $n_0 = 0, n_1>1$ or $n_0>0$, $n_1>0$, τ is subspace reducible, i.e., V contains a proper subspace invariant under the action of H_3, F_3, etc. Such a τ can loosely be thought of as n_0 copies of $[l_0,l_1]$ and n_1 copies of its tail "glued indecomposably" together.[16] However, τ is *not* in general determined to within equivalence by giving l_0, l_1, n_0 and n_1 alone. It is necessary to specify the action of H_3, F_3, etc., in a suitably chosen basis for V. In Ref. 15, it is shown that a basis of vectors $\xi_{lm\alpha}$ can be found, with $l = l_0, l_0+1,...;\quad m = l, l-1,..., -l;\quad \alpha = 1,2,...,n_0$ for $l_0 \leqslant l < |l_1|$; and $\alpha = 1,2,...,n_1$ for $l \geqslant |l_1|$, such that (adapting the notation of Ref. 17)

$$
\begin{aligned}
H_3\xi_{lm\alpha} &= m\xi_{lm\alpha},\\
H_\pm\,\xi_{lm\alpha} &= [(l\pm m+1)(l\mp m)]^{1/2}\xi_{lm\pm 1\alpha},\\
F_3\xi_{lm\alpha} &= [l^2-m^2]^{1/2}(M_l^\tau)_{\alpha\beta}\xi_{l-1m\beta} - m(Z_l^\tau)_{\alpha\beta}\xi_{lm\beta}\\
&\quad - [(l+1)^2-m^2]^{1/2}(P_l^\tau)_{\alpha\beta}\xi_{l+1m\beta},\\
F_\pm\,\xi_{lm\alpha} &= \pm[(l\mp m)(l\mp m-1)]^{1/2}\\
&\quad\times(M_l^\tau)_{\alpha\beta}\xi_{l-1m\pm1\beta}\\
&\quad - [(l\mp m)(l\pm m+1)]^{1/2}(Z_l^\tau)_{\alpha\beta}\xi_{lm\pm1\beta}\\
&\quad \pm[(l\pm m+1)(l\pm m+2)]^{1/2}\\
&\quad\times(P_l^\tau)_{\alpha\beta}\xi_{l+1m\pm1\beta}.
\end{aligned}
\tag{2.7}
$$

Here repeated subscripts β are to be summed over the values 1 to n_0 or n_1, as appropriate. Note that certain vectors on the right-hand sides in (2.7) are undefined, e.g., $\xi_{l-1m\beta}$ when $m = l$, but these can be ignored because they always appear with vanishing coefficients. The matrices M_l^τ, Z_l^τ, and P_l^τ, whose elements are $(M_l^\tau)_{\alpha\beta}$, $(Z_l^\tau)_{\alpha\beta}$, and $(P_l^\tau)_{\alpha\beta}$, respectively, have the appropriate dimensions. Thus P_l^τ is $n_0\times n_0$ for $l_0\leqslant l<|l_1|-1$; $n_1\times n_0$ for $l=|l_1|-1$; and $n_1\times n_1$ for $l\geqslant|l_1|$. Similarly, M_l^τ is $n_0\times n_0$ for $l_0\leqslant l<|l_1|$; $n_0\times n_1$ for $l=|l_1|$; and $n_1\times n_1$ for $l>|l_1|$, while Z_l^τ is $n_0\times n_0$ for $l_0\leqslant l<|l_1|$ and $n_1\times n_1$ for $l\geqslant|l_1|$. These matrices have the following form:

$$
P_l^\tau = \begin{cases} I_0, & l_0\leqslant l<|l_1|-1,\\ d_+, & l=|l_1|-1,\\ I_1, & l\geqslant|l_1|,\end{cases}
\tag{2.8a}
$$

$$
M_l^\tau = \begin{cases} [(l^2-l_0^2)(l_1^2-l^2)/(4l^2-1)l^2][I_0+(l_1^2/(l_1^2-l^2))a_0], & l_0\leqslant l<|l_1|,\\ d_-, & l=|l_1|,\\ [(l^2-l_0^2)(l_1^2-l^2)/(4l^2-1)l^2][I_1+(l_1^2/(l_1^2-l^2))a_1+(l_0^2/(l_0^2-l^2))\delta], & l>|l_1|,\end{cases}
\tag{2.8b}
$$

$$
Z_l^\tau = \begin{cases} [il_0l_1/l(l+1)]\sqrt{I_0+a_0}, & l_0\leqslant l<|l_1|,\\ [il_0l_1/l(l+1)]\sqrt{I_1+a_1+\delta}, & l\geqslant|l_1|,\end{cases}
\tag{2.8c}
$$

where I_0 and I_1 are the unit matrices of dimension $n_0\times n_0$ and $n_1\times n_1$, while d_+ is $n_1\times n_0$, d_- is $n_0\times n_1$, and δ is $n_1\times n_1$, such that

$$
\begin{aligned}
&d_-\delta=\delta d_+=0\\
&\delta \text{ and } d_+d_- \quad (\text{and hence } d_-d_+) \text{ are nilpotent.}
\end{aligned}
\tag{2.9}
$$

(The matrices Z_l^τ and M_l^τ, in cases with $l_0=0$, should be interpreted as vanishing when $l=0$, as should M_l^τ when $l=l_0=\frac{1}{2}$.) In addition, we have set

$$
\begin{aligned}
a_0 &= [(4l_1^2-1)/(l_1^2-l_0^2)]d_-d_+,\\
a_1 &= [(4l_1^2-1)/(l_1^2-l_0^2)]d_+d_-.
\end{aligned}
\tag{2.10}
$$

Each matrix square root in (2.8c) is of the form $\sqrt{I+A}$, with I a unit matrix and A nilpotent (say $A^{K+1}=0, A^K\neq 0$ in a particular case), and is to be interpreted through the binomial expansion as[22]

$$
\sqrt{I+A} = I+\tfrac{1}{2}A+\cdots + [(-1)^{K+1}(2K)!/(2K-1)(K!)^2 2^{2K}]A^K.
\tag{2.11}
$$

From the formulas (2.6)–(2.9) it can be deduced that the Casimir operators (2.3) leave each V_{lm} in (2.6) invariant, and act on V_{lm} as

$$
\begin{aligned}
\Delta_{1l} &= \begin{cases} -il_0l_1\sqrt{I_0+a_0}, & l_0\leqslant l<|l_1|,\\ -il_0l_1\sqrt{I_1+a_1+\delta}, & l\geqslant|l_1|,\end{cases}\\
\Delta_{2l} &= \begin{cases} (l_0^2+l_1^2-1)I_0+l_1^2a_0, & l_0\leqslant l<|l_1|,\\ (l_0^2+l_1^2-1)I_1+l_0^2\delta+l_1^2a_1, & l\geqslant|l_1|,\end{cases}
\end{aligned}
\tag{2.12}
$$

and it can be seen that they do indeed have one eigenvalue each, of the form (2.4), because of (2.11) and the nilpotency of a_0, a_1, and δ.

To complete the description of a singular indecomposable representation τ, it remains to complete the description of the matrices d_+, d_-, and δ subject to (2.9). Gel'fand and Ponomarev[15] have shown that inequivalent indecomposable sets of such matrices are in one-to-one correspondence with certain diagrams, which are therefore also in one-to-one correspondence with inequivalent indecomposable representations τ having the same values of l_0, l_1, n_0, and n_1. In other words, each such diagram may be regarded as providing the remaining labels necessary to characterize a corresponding τ, up to equivalence.

There are two categories of diagrams, of so-called "open" and "closed" types, and, correspondingly, there are singular indecomposable representations of type I and type II.

Definition 2.1: An *open diagram* is a finite set M of points in the lattice $\mathbb{Z}^2$, arranged as an unbroken staircase descending from left to right. Thus M contains one point [which can be taken without loss of generality to be the origin (0,0) in $\mathbb{Z}^2$], starting from which we can generate all of M by going successively either right or down to the nearest neighboring lattice point. Each point is colored black or white, with the restriction that nearest neighbors in M are both black if they are vertically adjacent, and are opposite in color if they are horizontally adjacent. (This implies that the length of each horizontal part of the staircase must be an even integer, unless that part includes the first or last point, when its length may be even or odd.) ■

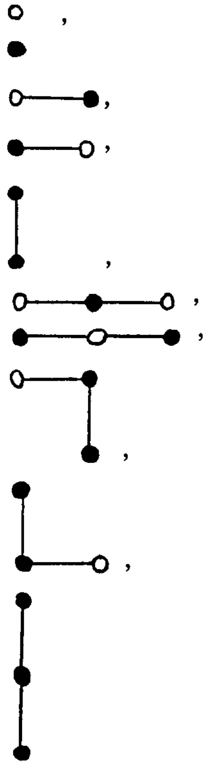
For the corresponding indecomposable representation τ of $\mathscr{G}$, n_0 equals the number of white points and n_1 the number of black points in the diagram.

The simplest diagrams are as follows:

$n_0 + n_1 = 1$: (2.13a) (2.13b)

$n_0 + n_1 = 2$: (2.13c) (2.13d) (2.13e)

$n_0 + n_1 = 3$: (2.13f) (2.13g) (2.13h) (2.13i) (2.13j)



To obtain from a given diagram the corresponding matrices $d_\pm, \delta$ of (2.7)–(2.9), we first associate with each of the points $(i,j) \in M$, a basis vector $e(i,j)$ in an (n_0+n_1)-dimensional complex vector space P. The basis vectors $e(i,j)$ corresponding to white points (i,j) span an n_0-dimensional subspace P_0 of P, while those corresponding to black points span an n_1-dimensional subspace P_1 of P; evidently

$$P = P_0 \oplus P_1. \tag{2.14}$$

Next we define linear operators a and b on P by

$$ae(i,j) = \begin{cases} e(i+1,j), & (i+1,j) \in M, \\ 0, & \text{otherwise}, \end{cases} \tag{2.15}$$

$$be(i,j) = \begin{cases} e(i,j+1), & (i,j+1) \in M, \\ 0, & \text{otherwise}. \end{cases} \tag{2.16}$$

Then a and b are nilpotent, with

$$ab = ba = 0,$$
$$aP_0 \subseteq P_1, \quad aP_1 \subseteq P_0$$
$$bP_0 = \{0\}, \quad bP_1 \subseteq P_1. \tag{2.17}$$

If we now identify P with $\mathbb{C}^{n_0+n_1}$, choosing the $e(i,j)$ in such a way that vectors in P_0 have their bottom n_1 components zero, and vectors in P_1 have their top n_0 components zero, we obtain a matrix realization of a and b with the form

$$a = \begin{bmatrix} 0 & d_- \\ d_+ & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 0 & 0 \\ 0 & \delta \end{bmatrix}, \tag{2.18}$$

where d_+, d_-, and δ have dimension $n_1 \times n_0$, $n_0 \times n_1$, and $n_1 \times n_1$, respectively, and satisfy (2.9).

For example, for the diagram (2.13h) we can take the points to be at (0,0), (1,0), and (1, − 1) in $\mathbb{Z}_2$, and set $e(0,0) = (1,0,0)^T$, $e(1,0) = (0,1,0)^T$, and $e(1,-1) = (0,0,1)^T$. Then (2.15) and (2.16) imply

$$a = \begin{bmatrix} 0 & 0 & 0 \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix}, \tag{2.19}$$

so that

$$d_+ = [1 \quad 0]^T, \quad d_- = [0 \quad 0], \quad \delta = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}. \tag{2.20}$$

In this way the matrices $d_\pm, \delta$ are determined by a given open diagram and, together with l_0 and l_1, complete the specification of a representation τ of type I. Note that for given l_0, l_1, the diagrams (2.13a) and (2.13b) correspond to the irreducible representations $[l_0, l_1]$ and $[|l_1|, \operatorname{sgn}(l_1) l_0]$.

Our treatment of an open diagram differs slightly from that of Ref. 15, in that we color the points to show explicitly the grading $P_0 \oplus P_1$. Apart from making the structure clearer, this is in fact necessary to distinguish between those inequivalent representations which would otherwise have the same "straight-row" diagram, consisting of n points in a horizontal line.

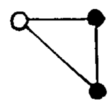
For example, (2.13c) and (2.13d) lead to

$$\begin{aligned} &d_+ = 1, \quad d_- = \delta = 0, \\ &d_- = 1, \quad d_+ = \delta = 0, \end{aligned} \tag{2.21}$$

respectively. For given l_0 and l_1, the corresponding inequivalent representations of $\mathscr{g}$ in this case are the well-known "operator-irreducible" indecomposable representations. For these the Casimir operators Δ_1 and Δ_2 of (2.3) are multiples of the identity by λ_1 and λ_2, as in (2.4), as follows from (2.21), (2.10), and (2.12), but the representations are nevertheless subspace reducible. In Ref. 23 they are denoted, respectively, by $\{l_0 \to l_1\}$ and $\{l_0 \leftarrow l_1\}$, this notation being intended to indicate that in the first case the subspace $V_{|l_1|} \oplus V_{|l_1|+1} \oplus \cdots$ of V is invariant, while in the second case $V_{l_0} \oplus V_{l_0+1} \oplus \cdots \oplus V_{|l_1|-1}$ is invariant. In Ref. 17 they are denoted by $(l_0, l_1, +)$ and $(l_0, l_1, -)$.

Definition 2.2: A *closed diagram* is obtained from any open diagram M that begins with a white point and ends with at least two successive black points. A line is drawn connecting the first and last points of M and the diagram is supplemented by a pair of labels (q,μ), $q \in \mathbb{N}$, $\mu \in \mathbb{C} \setminus 0$. ■

The simplest example has three points:



(q,μ). (2.22)

When $q = 1$, the procedure for constructing a and b, and subsequently $d_\pm$ and δ, is just as before, except that the definition (2.15) of a is supplemented by requiring

$$ae(k,l) = \mu e(0,0), \tag{2.23}$$

where (k,l) is the final and (0,0) the initial point of M. Again this leads to matrices a and b with the general form (2.18), from which d_+, d_-, and δ can be determined in order to complete the description of a singular indecomposable representation of $\mathscr{g}$ of type II. This representation is labeled (up to equivalence) by l_0, l_1, and the closed diagram [including the pair $(1,\mu)$].

For example, diagram (2.22) with $q = 1$ leads to

$$a = \begin{bmatrix} 0 & 0 & \mu \\ 1 & 0 & 0 \\ 0 & 0 & 0 \end{bmatrix}, \quad b = \begin{bmatrix} 0 & 0 & 0 \\ 0 & 0 & 1 \\ 0 & 0 & 0 \end{bmatrix} \tag{2.24}$$

and hence to

$$d_+ = [1 \quad 0]^T, \quad d_- = [0 \quad \mu], \quad \delta = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix}. \quad (2.25)$$

[Note that these matrices reduce to those in (2.20), corresponding to the open diagram (2.13h), if μ is set equal to zero. In fact, it is obviously true, in general, that the closed diagram with $q = 1$ and μ set equal to zero, reduces in this sense to the corresponding open diagram.]

More generally, for $q > 1$, a representation of type II is obtained by associating with each point (ij) of a closed diagram, a q-dimensional subspace $v(i,j)$ rather than a single vector $e(i,j)$ (as in the case $q = 1$). The definitions of a and b are generalized accordingly. Thus

$$av(i,j) \to v(i+1,j) \quad (2.26)$$

is an isomorphism if $(i+1,j) \in M$ (with a $q \times q$ matrix, which can be taken to the identity I_q); otherwise $av(i,j) = 0$, except that if $(0,0)$ is the first and (k,l) the last point of M, then a maps $v(k,l)$ into $v(0,0)$ with a $q \times q$ matrix μ_q, which can be taken to be a single Jordan block with eigenvalue μ. The mapping

$$bv(i,j) \to v(i,j+1) \quad (2.27)$$

is an isomorphism (with matrix I_q) if $(i,j+1) \in M$, and $bv(i,j) = 0$ otherwise. In this case the subspaces P_0 and P_1 of P are of dimension $n_0 = qm_0$, $n_1 = qm_1$, respectively, where m_0 and m_1 are the numbers of white and black points in the diagrams. Associating P with $\mathbb{C}^{n_0+n_1}$ as before, we read off the matrices $d_\pm$ and δ from the matrices of a and b in the same way. The corresponding indecomposable representation of $\mathscr{g}$ of type II is labeled by l_0, l_1, and the closed diagram [including (q,μ)].

For example, taking the diagram (2.22) with $q = 2$, the matrices can be obtained from those for the case $q = 1$, as in (2.24), (2.25), by replacing each zero by a 2×2 block of zeros, each 1 by the 2×2 unit matrix, and each μ by the 2×2 matrix

$$\mu_2 = \begin{bmatrix} \mu & 1 \\ 0 & \mu \end{bmatrix}. \quad (2.28)$$

III. VECTOR OPERATORS: GENERALITIES

Suppose that $\psi(x)$ in (1.1) is, for each x, an element of the $\mathscr{g}$ module V_π of a Harish-Chandra representation π of $\mathscr{g}$,

$$\pi \cong \bigoplus_r \tau_r,$$
$$V_\pi \cong \bigoplus_r V^{\tau_r}, \quad (3.1)$$

where each τ_r is (singular) indecomposable. Suppose further that Γ_μ and κ in (1.1) are linear mappings (operators) from V_π into itself. Let $H_3 = \pi(h_3)$, $F_3 = \pi(f_3)$, etc. Generalizing well-known results,[1,24] we know that a sufficient condition for (1.1) to be locally[12] invariant under homogeneous (and, indeed, inhomogeneous) Lorentz transformations is that κ commutes with H_3, F_3, etc., on V_π and

$$[\Gamma_0, H_\pm] = [\Gamma_0, H_3] = 0, \quad (3.2a)$$
$$\Gamma_0 = [[F_3,\Gamma_0],F_3], \quad (3.2b)$$
$$\Gamma_3 = i[F_3,\Gamma_0], \quad \Gamma_1 \pm i\Gamma_2 = i[F_\pm,\Gamma_0], \quad (3.2c)$$

on V_π. Then we say that κ is a scalar operator and Γ_μ a four-vector operator on V_π. In searching for possible locally invariant equations, we therefore seek representations π for which κ and Γ_μ can be found with these properties. If we restrict attention to equations for which κ is invertible on V_π, then[24] there is no significant loss of generality in supposing κ to be a nonzero numerical multiple of the identity operator on V_π. Then the problem reduces to finding representations π for which a four-vector operator Γ_μ can be found. It is sufficient to search for Γ_0 satisfying (3.2a) and (3.2b), as Γ_1, Γ_2, and Γ_3 can then be *defined* by (3.2c). We shall concentrate on this problem here, but make some remarks about wave equations with noninvertible κ at the end of the paper.

Let τ and τ' denote any two of the indecomposable representations (or possibly one and the same representation) contained in π, and let P^τ, $P^{\tau'}$ be the operators projecting V_π onto the corresponding subspaces V^τ, $V^{\tau'}$ in (3.1). Define

$$\Gamma_\mu^{\tau'\tau} = P^{\tau'}\Gamma_\mu P^\tau \quad (3.3)$$

and note that, since P^τ and $P^{\tau'}$ commute on V_π with H_3, F_3, etc., $\Gamma_\mu^{\tau'\tau}$ is a four-vector operator whenever this is true of Γ_μ. We concentrate on the determination of $\Gamma_\mu^{\tau'\tau}$, in effect restricting attention to the case when $\pi = \tau \oplus \tau'$ (or $\pi = \tau$, if $\tau' = \tau$). The Γ_μ in a more general case can evidently be built up from such $\Gamma_\mu^{\tau'\tau}$.

Decomposing V^τ and $V^{\tau'}$ as in (2.5) and (2.6), and introducing bases, as in Sec. II, we see from (3.2a) and (3.3) that $\Gamma_0^{\tau'\tau}$ carries V_{lm}^τ into $V_{lm}^{\tau'}$. Let $X_l^{\tau'\tau}$ denote the corresponding matrix; it is independent of m, again because of (3.2a). Expressions for the matrices $\Gamma_{p;lm}^{\tau'\tau}$, defined by the action of the operators $\Gamma_p^{\tau'\tau}$, $p = 1,2,3$, on V_{lm}^τ, can then be written down. For example, we find, using (2.7) and (3.2c), that

$$\begin{aligned} i\Gamma_{3;lm}^{\tau'\tau} &= [l^2 - m^2]^{1/2}(X_{l-1}^{\tau'\tau}M_l^\tau - M_l^{\tau'}X_l^{\tau'\tau}) \\ &\quad - m(X_l^{\tau'\tau}Z_l^\tau - Z_l^{\tau'}X_l^{\tau'\tau}) \\ &\quad - [(l+1)^2 - m^2]^{1/2}(X_{l+1}^{\tau'\tau}P_l^\tau - P_l^{\tau'}X_l^{\tau'\tau}). \end{aligned} \quad (3.4)$$

The heart of the problem then is to determine the $X_l^{\tau'\tau}$ from the remaining condition (3.2b), for the appropriate range of l values in the representation τ. This condition leads to the following system of coupled matrix equations, essentially the same as Eqs. (3.6) of Ref. 17, in which Z_l^τ, M_l^τ, P_l^τ, and the corresponding primed variables are to be regarded as given, and the $X_l^{\tau'\tau}$ are unknowns:

$$2P_{l+1}^{\tau'}X_{l+1}^{\tau'\tau}P_l^\tau - P_{l+1}^{\tau'}P_l^{\tau'}X_l^{\tau'\tau} - X_{l+2}^{\tau'\tau}P_{l+1}^\tau P_l^\tau = 0, \quad (3.5a)$$

$$2M_{l-1}^{\tau'}X_{l-1}^{\tau'\tau}M_l^\tau - M_{l-1}^{\tau'}M_l^{\tau'}X_l^{\tau'\tau} - X_{l-2}^{\tau'\tau}M_{l-1}^\tau M_l^\tau = 0, \quad (3.5b)$$

$$\begin{aligned} X_{l+1}^{\tau'\tau}[P_l^\tau Z_l^\tau + Z_{l+1}^\tau P_l^\tau] &+ [P_l^{\tau'}Z_l^{\tau'} + Z_{l+1}^{\tau'}P_l^{\tau'}]X_l^{\tau'\tau} \\ &- 2P_l^{\tau'}X_l^{\tau'\tau}Z_l^\tau - 2Z_{l+1}^{\tau'}X_{l+1}^{\tau'\tau}P_l^\tau = 0, \end{aligned} \quad (3.5c)$$

$$\begin{aligned} X_{l-1}^{\tau'\tau}&[Z_{l-1}^\tau M_l^\tau + M_l^\tau Z_l^\tau] \\ &+ [Z_{l-1}^{\tau'}M_l^{\tau'} + M_l^{\tau'}Z_l^{\tau'}]X_l^{\tau'\tau} \end{aligned}$$

$$-2Z^{\tau'}_{l-1}X^{\tau'\tau}_{l-1}M^{\tau}_{l}-2M^{\tau'}_{l}X^{\tau'\tau}_{l}Z^{\tau}_{l}=0, \quad (3.5d)$$

$$\begin{aligned} X^{\tau'\tau}_{l}&[P^{\tau}_{l-1}M^{\tau}_{l}+M^{\tau}_{l+1}P^{\tau}_{l}+Z^{\tau}_{l}Z^{\tau}_{l}]\\ &+[Z^{\tau'}_{l}Z^{\tau'}_{l}+P^{\tau'}_{l-1}M^{\tau'}_{l}+M^{\tau'}_{l+1}P^{\tau'}_{l}]X^{\tau'\tau}_{l}\\ &-2P^{\tau'}_{l-1}X^{\tau'\tau}_{l-1}M^{\tau}_{l}-2M^{\tau'}_{l+1}X^{\tau'\tau}_{l+1}P^{\tau}_{l}\\ &-2Z^{\tau'}_{l}X^{\tau'\tau}_{l}Z^{\tau}_{l}=0, \end{aligned} \quad (3.5e)$$

$$\begin{aligned} (l+1)^2&[X^{\tau'\tau}_{l}M^{\tau}_{l+1}P^{\tau}_{l}\\ &-2M^{\tau'}_{l+1}X^{\tau'\tau}_{l+1}P^{\tau}_{l}+M^{\tau'}_{l+1}P^{\tau'}_{l}X^{\tau'\tau}_{l}]\\ &+l^2[X^{\tau'\tau}_{l}P^{\tau}_{l-1}M^{\tau}_{l}-2P^{\tau'}_{l-1}X^{\tau'\tau}_{l-1}M^{\tau}_{l}\\ &+P^{\tau'}_{l-1}M^{\tau'}_{l}X^{\tau'\tau}_{l}]=X^{\tau'\tau}_{l}. \end{aligned} \quad (3.5f)$$

These equations must hold for each allowed l value in the representation τ, and $X^{\tau'\tau}_{l}$ must be set equal to zero unless both V^{τ}_{l} and $V^{\tau'}_{l}$ are non-null.

If τ and τ' are irreducible [so that each matrix in (3.5) is a single number], it is well known[1] that a necessary and sufficient condition for the existence of a nontrivial solution is that the labels (l_0,l_1) and (l'_0,l'_1) of τ and τ' (which serve to characterize the representations completely in such a case) satisfy the "interlocking" condition that one of the pairs (l'_0,l'_1), $(-l'_0,-l'_1)$ is equal to one of the pairs (l_0,l_1+1), (l_0,l_1-1), (l_0+1,l_1), or (l_0-1,l_1). Moreover, the structure of the solutions is known for all such cases.[1] Hlavaty and Niederle[17] have in fact extended these results to the case where τ and τ' are nonsingular indecomposable representations.

In the singular case, we expect that it is still true that the labels (l_0,l_1), (l'_0,l'_1) of τ and τ' (which now are not sufficient to characterize τ and τ' completely) must satisfy the interlocking condition if a nontrivial solution is to exist, although we have no general proof. On the other hand, it would be surprising if this condition is sufficient for existence, with no restriction on the diagrams for τ and τ', but this too remains an open question.

IV. VECTOR OPERATORS: EXAMPLES

For a given pair (τ,τ') of singular indecomposable representations,[25] it is not, in general, easy to determine if Eqs. (3.5) admit a nontrivial solution. The only solutions previously presented (apart of course from those corresponding to irreducible τ and τ') seem to have been those corresponding to the case of interlocking τ and τ' of the operator-irreducible type, with diagrams of the form (2.13c) or (2.13d). Then $n_0=n_1=1$, so that all the matrices in (3.5) are simply numbers, and the problem of finding solutions is not substantially more difficult than in the irreducible case.

In Ref. 17, a nontrivial solution is found in a case of this type, where τ has labels $(\frac{1}{2},l_1)$ and τ' has labels $(\frac{1}{2},-l_1)$, with $l_1\in\{\frac{3}{2},\frac{5}{2},...\}$, and the diagrams for τ and τ' are (2.13c) and (2.13d), respectively. (These are the operator-irreducible representations $[\frac{1}{2}\rightarrow l_1]$ and $[\frac{1}{2}\leftarrow -l_1]$ described earlier.) It was shown, however, that this leads to an operator Γ_0 with no nonzero eigenvalues, so that (1.1) has no timelike solutions in this case.[17] In fact, four-vector operators in the case of interlocked operator-irreducible τ and τ' were described earlier by Ruhl.[19]

In seeking to find, for representations with arbitrarily complicated diagrams, examples of four-vector operators that lead in at least some cases to wave equations with timelike solutions, we shall make the following simplifying assumptions: (1) τ and τ' have labels $(\frac{1}{2},l_1)$ and $(\frac{1}{2},-l_1)$, where $l_1\in\{\frac{3}{2},\frac{5}{2},...\}$ (note that the interlocking condition is then satisfied); and (2) τ and τ' have the *same* diagram.

Note that condition (1) but not condition (2) is satisfied by the example of Ref. 17. The matrices $d_{\pm}$, δ (and hence a_0, a_1) can now be taken to be the same for τ' as for τ, and we obtain

$$Z^{\tau'}_{l}=-Z^{\tau}_{l};\quad P^{\tau'}_{l}=P^{\tau}_{l};\quad M^{\tau'}_{l}=M^{\tau}_{l}, \quad (4.1)$$

and, for the matrices of the Casimir operators (2.12),

$$\Delta^{\tau}_{1l}=-\Delta^{\tau'}_{1l},$$
$$\Delta^{\tau}_{2l}=\Delta^{\tau'}_{2l}. \quad (4.2)$$

Under these conditions, $X^{\tau'\tau}_{l}$ $(\frac{1}{2}\leqslant l<l_1)$ and a_0 are $n_0\times n_0$ matrices, while $X^{\tau'\tau}_{l}$ $(l\geqslant l_1)$, a_1, and δ are $n_1\times n_1$ matrices, where n_0 and n_1 are determined by the number of white and black points in the (common) diagram for τ and τ', as described in Sec. II. In order to simplify ordering problems in (3.5), we limit ourselves further by seeking only solutions such that (3) $X^{\tau'\tau}_{l}$ commutes with a_0 for $\frac{1}{2}\leqslant l<l_1$, and with a_1 and δ for $l\geqslant l_1$.

It follows from (2.12) and (4.2) that condition (3) is equivalent to requiring that $\Gamma^{\tau'\tau}_{\mu}$ satisfies

$$[\Gamma^{\tau'\tau}_{\mu},\Delta_2]=0=\{\Gamma^{\tau'\tau}_{\mu},\Delta_1\} \quad (4.3)$$

on $V^{\tau}\oplus V^{\tau'}$. Equations (4.3) are known[7] to hold for all solutions in the case that τ and τ' are irreducible [given conditions (1) and (2)], when Δ_2 is a multiple of the identity operator (on $V^{\tau}\oplus V^{\tau'}$) and Δ_1 is a (generalized) Dirac γ_5 matrix, but it is not clear if the imposition of condition (3) places a nontrivial restriction on the solution of (3.5) in the present situation.

Having imposed conditions (1)–(3), we now consider (3.5a) and find using (2.8) that

$$X^{\tau'\tau}_{l+2}=2X^{\tau'\tau}_{l+1}-X^{\tau'\tau}_{l} \quad (4.4)$$

for $\frac{1}{2}\leqslant l<l_1-2$, and for $l\geqslant l_1$, so that

$$X^{\tau'\tau}_{l}=\begin{cases} lA+C, & \frac{1}{2}\leqslant l<l_1,\\ lB+D, & l\geqslant l_1, \end{cases} \quad (4.5)$$

where A, B, C, and D are matrices independent of l. According to condition (3), A and C commute with a_0, and B and D commute with a_1 and δ. Use of (4.1) and (2.8) in (3.5c), with $l_0\leqslant l<l_1-1$, then gives

$$\begin{aligned}&[(2l-1)A+2C]P^{\tau}_{l}Z^{\tau}_{l}\\ &\quad=[(2l+3)A+2C]Z^{\tau}_{l+1}P^{\tau}_{l}.\end{aligned} \quad (4.6)$$

But Eqs. (2.8) imply that $lP^{\tau}_{l}Z^{\tau}_{l}=(l+2)Z^{\tau}_{l+1}P^{\tau}_{l}$, so that

$$\begin{aligned}&(l+2)[(2l-1)A+2C]P^{\tau}_{l}Z^{\tau}_{l}\\ &\quad=l[(2l+3)A+2C]P^{\tau}_{l}Z^{\tau}_{l},\end{aligned} \quad (4.7)$$

and since $P^{\tau}_{l}Z^{\tau}_{l}$ is nonsingular, we obtain $C=\frac{1}{2}A$. In a similar way, using (3.5c) with $l\geqslant l_1$, we obtain $D=\frac{1}{2}B$. Thus

$$X_l^{\tau'\tau} = \begin{cases} (l+\frac{1}{2})A, & \frac{1}{2}\leqslant l<l_1, \\ (l+\frac{1}{2})B, & l\geqslant l_1. \end{cases} \tag{4.8}$$

Then (3.5a) with $l = l_1 - 2$ or $l = l_1 - 1$ implies

$$d_+A = Bd_+. \tag{4.9}$$

Equation (3.5b) yields no new condition for $\frac{5}{2}\leqslant l < l_1$ or $l > l_1 + 2$, but for $l = l_1 + 1$ or $l = l_1$ we obtain

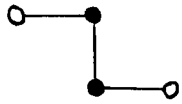
$$Ad_- = d_-B. \tag{4.10}$$

The remaining equations (3.5) require just one more condition, that

$$B\delta = \delta B = 0. \tag{4.11}$$

Since (4.9) and (4.10) imply, with (2.10), that A commutes with a_0 and B commutes with a_1, the problem reduces to the following: for a given diagram and hence for given $d_\pm$, δ, find matrices A and B such that (4.9)–(4.11) are satisfied.

For small values of n_0 and n_1, we can now easily construct all solutions subject to the conditions (1)–(3). For example, the diagram



(4.12)

leads to

$$d_+ = \begin{bmatrix} 1 & 0 \\ 0 & 0 \end{bmatrix}, \quad d_- = \begin{bmatrix} 0 & 0 \\ 0 & 1 \end{bmatrix}, \quad \delta = \begin{bmatrix} 0 & 1 \\ 0 & 0 \end{bmatrix} \tag{4.13}$$

[and hence to $a_0 = a_1 = \left[\begin{smallmatrix} 0 & 0 \\ 0 & 0 \end{smallmatrix}\right]$], and it follows from (4.9)–(4.11) that

$$A = \begin{bmatrix} 0 & 0 \\ \alpha & 0 \end{bmatrix}, \quad B = \begin{bmatrix} 0 & \beta \\ 0 & 0 \end{bmatrix}, \tag{4.14}$$

where α and β are arbitrary constants. Similarly, for diagram (2.22) with $q = 2$ and $\mu \neq 0$, we find that A equals the 2×2 zero matrix and B is a 4×4 matrix with an arbitrary 2×2 block in the upper right-hand corner, and zeros elsewhere.

Of more interest is the observation that a class of solutions (A,B) can now be determined as follows, whatever the common diagram of τ and τ'. Since (2.9) and (2.10) imply that

$$d_+a_0 = a_1d_+, \quad a_0d_- = d_-a_1,$$
$$\delta a_1 = a_1\delta = 0, \tag{4.15}$$

we can satisfy (4.9) and (4.10) by taking

$$A = \alpha_0 I_0 + \alpha_1 a_0 + \alpha_2 a_0^2 + \cdots + \alpha_N a_0^N,$$
$$B = \alpha_0 I_1 + \alpha_1 a_1 + \alpha_2 a_1^2 + \cdots + \alpha_N a_1^N + c\delta^M, \tag{4.16}$$

where c and the α_i, $i = 0,1,2,\ldots,N$ are arbitrary complex constants, N is the largest non-negative integer such that at least one of a_0^N, a_1^N is nonzero, and M is the largest non-negative integer such that δ^M is nonzero.

If $\delta = 0$, then (4.11) is satisfied trivially, and (4.16) defines a class of solutions (4.8) parametrized by $\alpha_0,\alpha_1,\ldots,\alpha_N$. The vanishing of δ is easily seen to require that the diagram of τ and τ' be a straight row: we discuss this case further in the next section .

If $\delta \neq 0$, we must set $\alpha_0 = 0$ in (4.16) in order to satisfy (4.11). We then still obtain a class of solutions (4.8), parametrized by $\alpha_1,\alpha_2,\ldots,\alpha_N$ and c, but note immediately that $X_l^{\tau'\tau}$ is then nilpotent for each l, and that the same will be true of $\Gamma_0^{\tau'\tau}$. The corresponding wave equation (1.1) with $\Gamma_\mu = \Gamma_\mu^{\tau'\tau}$ [or with $\Gamma_\mu = \Gamma_\mu^{\tau'\tau} + \Gamma_\mu^{\tau\tau'}$ as in (4.17) below] will not admit timelike solutions in such a case. The question arises as to whether or not (4.8) and (4.16) give in the way described, all solutions of (3.5) under conditions (1)–(3). That the answer is no, at least when $\delta \neq 0$, is shown by the counterexample (4.12)–(4.14), with $\alpha \neq 0$.

We attempted to find other solutions to (3.5), satisfying conditions (1) and (2), but not (3), with the help of the symbolic manipulation computer package[26] MUMATH, but were unsuccessful. Furthermore, for all diagrams leading to $\delta \neq 0$, we found only nilpotent solutions $X_l^{\tau'\tau}$. We conjecture that this is a general rule, at least when conditions (1) and (2) hold.

We also found only nilpotent solutions in cases like that considered in Ref. 17, where condition (2) does not hold. This was in fact our motivation for imposing that condition.

Note that if the roles of τ and τ' are interchanged in (3.5), then $X_l^{\tau\tau'}$ satisfies the same equations as $X_l^{\tau'\tau}$, as a consequence of (4.1). Therefore we have also found solutions $X_l^{\tau\tau'}$ of the same general form (4.8), (4.16), and we can consider

$$\Gamma_\mu = \Gamma_\mu^{\tau'\tau} + \Gamma_\mu^{\tau\tau'} \tag{4.17}$$

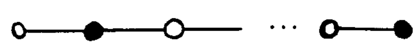
in (1.1). This will, in general, be necessary if equations possessing timelike solutions are to be obtained, as is familiar from the case of the Dirac matrices γ_μ, which couple the irreducible representations $\tau = [\frac{1}{2},\frac{3}{2}]$, $\tau' = [\frac{1}{2}, -\frac{3}{2}]$; here

$$\Gamma_\mu^{\tau'\tau} = \alpha\gamma_\mu(1+\gamma_5), \quad \Gamma_\mu^{\tau\tau'} = \beta\gamma_\mu(1-\gamma_5),$$
$$\gamma_\mu = \Gamma_\mu^{\tau'\tau} + \Gamma_\mu^{\tau\tau'} \tag{4.18}$$

(choosing $\alpha = \beta = \frac{1}{2}$).

V. A CLASS OF WAVE EQUATIONS

We consider the case when τ and τ' have the straight row diagram



(5.1)

with $2k$ points and, as in Sec. IV, condition (1) holds. Following the prescription outlined in Sec. II we obtain

$$d_+ = I_k, \quad \delta = 0, \quad \text{and}$$

$$d_- = \tfrac{1}{4}a_0 = \tfrac{1}{4}a_1 = \begin{bmatrix} 0 & & & & & \\ 1 & 0 & & & 0 & \\ & 1 & 0 & & & \\ & & \cdots & & & \\ & & & \cdots & & \\ & 0 & & & \cdots & \\ & & & & 1 & 0 \end{bmatrix}, \tag{5.2}$$

each being a $k\times k$ matrix. Then from (4.8) and (4.16) we have

$$X_l^{\tau'\tau} = (l+\tfrac{1}{2}) \sum_{j=0}^{k-1} \zeta_j (d_-)^j, \tag{5.3}$$

where the ζ_j are arbitrary constants. Similarly,

$$X_l^{\tau\tau'} = (l+\tfrac{1}{2}) \sum_{j=0}^{k-1} \eta_j (d_-)^j, \tag{5.4}$$

with arbitrary η_j. We restrict our attention to cases where η_0 and ζ_0 are nonzero, so that $X_l^{\tau'\tau}$ and $X_l^{\tau\tau'}$ are not nilpotent; because we can multiply (1.1) throughout by an arbitrary constant, there is no significant loss of generality in assuming then that

$$\eta_0\zeta_0 = 1. \tag{5.5}$$

On each $2k$-dimensional subspace V_{lm} of V ($= V^\tau \oplus V^{\tau'}$),

$$V_{lm} = \begin{bmatrix} V_{lm}^{\tau} \\ V_{lm}^{\tau'} \end{bmatrix}, \quad m\in\{l, l-1, ..., -l\}, \tag{5.6}$$

the matrix of Γ_0 ($= \Gamma_0^{\tau'\tau} + \Gamma_0^{\tau\tau'}$) is

$$\Gamma_{0l} = \begin{bmatrix} 0 & X_l^{\tau\tau'} \\ X_l^{\tau'\tau} & 0 \end{bmatrix}, \tag{5.7}$$

where the zeros represent $k\times k$ blocks. It follows from (5.3)–(5.5) that Γ_{0l} has only $\pm(l+\frac{1}{2})$ as eigenvalues, but is not, in general, completely diagonalizable, depending on the values of the ζ_j and η_j.

For example, in the case that $k=3$, we can take $\zeta_0=\eta_0=1$, $\zeta_1=\eta_1=\zeta_2=\eta_2=0$, and find that for each eigenvalue $\pm(l+\frac{1}{2})$ of Γ_{0l} (and for each value of m) there are three linearly independent eigenvectors, and Γ_{0l} is diagonalizable; or take $\zeta_0=\eta_0=\zeta_2=\eta_2=1$, $\zeta_1=\eta_1=0$, and find only two linearly independent eigenvectors for each eigenvalue; or take $\zeta_0=\eta_0=\zeta_1=\eta_1=\zeta_2=\eta_2=1$ and find only one eigenvector for each eigenvalue. In these last two cases, Γ_{0l} is not diagonalizable and does not have a complete set of eigenvectors.

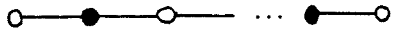
It follows that (1.1) will admit timelike solutions corresponding to at least one set of positive and one set of negative energy particles with a Majorana-type mass-spin spectrum

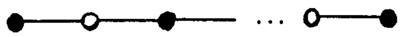
$$m_l = \kappa/(l+\tfrac{1}{2}), \quad l=\tfrac{1}{2},\tfrac{3}{2},..., \tag{5.8}$$

but that there will, in general, be enough linearly independent solutions to describe n such sets, $1\leqslant n\leqslant k$. We can expect that, in general, there will also be lightlike and spacelike solutions of (1.1), as for the case[27] of infinite-dimensional irreducible representations τ and τ'.

Similar results hold in the case that the black and white points in (5.1) are interchanged. This simply leads to an interchange of d_- and d_+ in (5.2)–(5.4).

Slightly more complicated are the cases corresponding to the diagrams

 , (5.9a)

 , (5.9b)

each with $2k+1$ points. For the diagrams (5.9a) we obtain

$$d_+ = \begin{bmatrix} & 0 \\ I_k & 0 \\ & \vdots \\ & 0 \end{bmatrix}, \quad d_- = \begin{bmatrix} 0 & 0\cdots 0 \\ & I_k \end{bmatrix}, \quad \delta=0; \tag{5.10}$$

d_+ is $k\times(k+1)$, d_- is $(k+1)\times k$, and δ is $k\times k$. Then $\frac{1}{4}a_0$ and $\frac{1}{4}a_1$ have the same form as the matrix d_- in (5.2), with a_0 being $(k+1)\times(k+1)$, and a_1 being $k\times k$. Our solution (4.8),(4.16) now gives

$$X_l^{\tau'\tau} = \begin{cases} \left(l+\dfrac{1}{2}\right)\displaystyle\sum_{j=0}^{k} \zeta_j \left(\frac{1}{4}a_0\right)^j, & \dfrac{1}{2}\leqslant l<l_1 \\ \left(l+\dfrac{1}{2}\right)\displaystyle\sum_{j=0}^{k-1} \zeta_j \left(\frac{1}{4}a_1\right)^j, & l\geqslant l_1. \end{cases} \tag{5.11}$$

We get a similar expression for $X_l^{\tau\tau'}$, with further arbitrary constants η_j replacing the ζ_j of (5.11). Again we suppose ζ_0,η_0 are nonzero, set $\eta_0\zeta_0=1$, and find that Γ_{0l} has only $\pm(l+\frac{1}{2})$ as eigenvalues. For some choices of arbitrary constants ζ_j and η_j, Γ_{0l} is diagonalizable (for every l), but for most it is not. A new feature that emerges is that for a given eigenvalue $\pm(l+\frac{1}{2})$, Γ_{0l} may have a different number of linearly independent eigenvectors for $l<l_1$ than for $l\geqslant l_1$. For example, if we set $\zeta_0=\eta_0=1$, all other ζ_j and η_j being equal to zero, then Γ_{0l} is diagonalizable, with $(k+1)$ linearly independent eigenvectors for $l<l_1$, and k for $l\geqslant l_1$. The corresponding wave equation (1.1) would admit timelike solutions capable of describing $(k+1)$ positive (or negative) energy particles with spins $\frac{1}{2},\frac{3}{2},...,l_1-1$, and k with spins $l_1, l_1+1, ...$. Again the mass-spin spectrum is of the Majorana type.

Similar remarks apply in the case of the diagram (5.9b); in this case Γ_{0l} will be $2k\times 2k$ for $l<l_1$, and $2(k+1)\times 2(k+1)$ for $l\geqslant l_1$.

For any of these straight row diagrams, we can restrict attention to the diagonalizable cases by requiring that (5.5) holds and

$$X_l^{\tau'\tau}X_l^{\tau\tau'} = (l+\tfrac{1}{2})^2 I_l, \tag{5.12}$$

where I_l is the unit matrix of the appropriate size. Then

$$(\Gamma_{0l})^2 = (l+\tfrac{1}{2})^2 (I_l \oplus I_l). \tag{5.13}$$

Some important algebraic properties of the corresponding operators Γ_μ can now be determined. It follows from (3.4) that, quite generally, the matrix of the operator $i[\Gamma_0^{\tau\tau'}\Gamma_3^{\tau'\tau} \pm \Gamma_3^{\tau\tau'}\Gamma_0^{\tau'\tau}]$ on the subspace V_{lm} of V is given by

$$\begin{aligned} &[l^2-m^2]^{1/2}\{X_{l-1}^{\tau\tau'}(X_{l-1}^{\tau'\tau}M_l^{\tau} - M_l^{\tau'}X_l^{\tau'\tau}) \pm (X_{l-1}^{\tau\tau'}M_l^{\tau'} - M_l^{\tau}X_l^{\tau\tau'})X_l^{\tau'\tau}\} - m\{X_l^{\tau\tau'}(X_l^{\tau'\tau}Z_l^{\tau} - Z_l^{\tau'}X_l^{\tau'\tau}) \\ &\pm (X_l^{\tau\tau'}Z_l^{\tau'} - Z_l^{\tau}X_l^{\tau\tau'})X_l^{\tau'\tau}\} - [(l+1)^2-m^2]^{1/2}\{X_{l+1}^{\tau\tau'}(X_{l+1}^{\tau'\tau}P_l^{\tau} - P_l^{\tau'}X_l^{\tau'\tau}) \pm (X_{l+1}^{\tau\tau'}P_l^{\tau'} - P_l^{\tau}X_l^{\tau\tau'})X_l^{\tau'\tau}\}. \end{aligned} \tag{5.14}$$

In the present context, this reduces, with the help of (2.7), (2.8), and (5.12), to

$$i[\Gamma_0,\Gamma_3]_{lm} = (F_3 - 4H_3\,\Delta_1)_{lm},$$
$$i\{\Gamma_0,\Gamma_3\}_{lm} = (F_-H_+ - H_-F_+)_{lm}. \tag{5.15}$$

Since (5.13) also holds, it is then easy to verify from (3.2) that we have the following identities on V:

$$[\Gamma_\mu,\Gamma_\nu] = -iJ_{\mu\nu} + 4i\tilde{J}_{\mu\nu}\,\Delta_1, \tag{5.16a}$$

$$\{\Gamma_\mu,\Gamma_\nu\} = 2g_{\mu\nu}(\Delta_2 + \tfrac{1}{4}I) - \{J_{\mu\sigma},J_\nu{}^\sigma\}, \tag{5.16b}$$

where, as usual,

$$J_{pq} = \epsilon_{pqr}H_r,\quad J_{0p} = -J_{p0} = F_p,$$
$$\tilde{J}_{\mu\nu} = \tfrac{1}{2}\epsilon_{\mu\nu\rho\sigma}J^{\rho\sigma}. \tag{5.17}$$

(Here latin subscripts run over 1,2,3; Greek over 0,1,2,3. We use the summation convention and set $J^{\rho\sigma} = g^{\rho\alpha}g^{\sigma\beta}J_{\alpha\beta}$. The metric tensor $g_{\mu\nu} = g^{\mu\nu}$ is diagonal, with $g_{00} = -g_{11} = -g_{22} = -g_{33} = 1$; and the alternating tensor has $\epsilon_{0123} = -1$.)

It is noteworthy that the indentities (5.16) are exactly those proved by Bracken[7] for the family of four-vector operators based on the direct sum of the irreducible representations $[\frac{1}{2},l_1]$ and $[\frac{1}{2},-l_1]$, $l_1\in\mathbb{C}$. These identities have some interesting consequences. Since Δ_1 can never vanish in the present context (as $l_0l_1\neq 0$), it follows that we never obtain the so(5,$\mathbb{C}$) commutation relations

$$[\Gamma_\mu,\Gamma_\nu] = -iJ_{\mu\nu}. \tag{5.18}$$

In fact, an analysis similar to that of Cant[8] shows that an infinite-dimensional Lie algebra will be generated by the Γ_μ in the present situation.

Supposing that (1.1) holds (with κ a number), the identity (5.16b) implies that, for sufficiently smooth ψ,

$$(\tfrac{1}{4}\partial^\mu\partial_\mu + \omega^\mu\omega_\mu)\psi = -\varkappa^2\psi, \tag{5.19}$$

where $\omega_\mu = i\tilde{J}_{\mu\nu}\,\partial^\nu$ is the Pauli–Lubanski vector operator, so that

$$\omega^\mu\omega_\mu = \tfrac{1}{2}J_{\mu\nu}J^{\mu\nu}\partial_\sigma\,\partial^\sigma - J_{\mu\sigma}J_\nu{}^\sigma\,\partial^\mu\partial^\nu$$
$$= \Delta_2\,\partial_\sigma\,\partial^\sigma - \tfrac{1}{2}\{J_{\mu\sigma},J_\nu{}^\sigma\}\partial^\mu\,\partial^\nu. \tag{5.20}$$

If ψ is a wave function for a particle with mass m_l and spin l, then we will also have

$$\partial^\mu\,\partial_\mu\psi = -\mathrm{m}_l^2\psi, \tag{5.21a}$$

$$\omega_\mu\omega^\mu\psi = -l(l+1)m_l^2\psi, \tag{5.21b}$$

and (5.19) then implies that

$$m_l^2 = \kappa^2/(l+\tfrac{1}{2})^2, \tag{5.22}$$

in agreement with (5.8). Equation (5.19) also determines the nature of generalized mass-spin relations for lightlike and spacelike solutions of (1.1). In this connection we remark that it can be seen from (5.16b) that $(\Gamma_0 + \Gamma_p)$ and Γ_p $(p = 1,2,3)$ are not diagonalizable, unlike Γ_0.

VI. CONCLUDING REMARKS

The structure of indecomposable representations of sl(2,$\mathbb{C}$) is rich and interesting from a mathematical point of view. Because of the central role played by this Lie algebra and associated group in relativistic physics, we might expect that the theory of its indecomposable representations should be of relevance to applications as well. However, it must be said that, following the present work and that of Ref. 17, it is by no means clear that relativistic wave equations of the form (1.1), based on such representations, are likely to prove useful in physics.

In the case of singular indecomposable representations, we have shown that a variety of four-vector operators and corresponding wave equations can be constructed, corresponding to the great variety of such representations, labeled by ladder diagrams as in Sec. II, and we do not claim to have exhausted the possibilities, even under the restrictive conditions (1)–(3) imposed in Sec. IV. Our main objective has been to produce illustrative examples. Only for a very restricted subclass of representations (corresponding to straight row diagrams) did we find examples with Γ_0 not nilpotent, although even then there is a considerable variety of possibilities, as we have seen in Sec. V. However, these all lead to mass-spin spectra of the Majorana type, a dissappointing result from the point of view of potential applications.

It could be that the solutions of (1.1), in cases based on indecomposable representations, ought to be interpreted, in general, in a different way than in cases based on irreducible representations. For example, we could consider ψ to belong to the representation $[\frac{1}{2}\rightarrow\frac{3}{2}]\oplus[\frac{1}{2}\rightarrow -\frac{3}{2}]$. (See Sec. II.) The subspace $U\subset V$,

$$U = \bigoplus_{l\geqslant\frac{3}{2}} V_l,$$

is then invariant under the action of the sl(2,$\mathbb{C}$) algebra. Moreover, since Γ_0 leaves each V_l invariant, U is also invariant under the action of Γ_0 and therefore, by (3.2), of all Γ_μ. It is also invariant under the action of κ if that is a multiple of the identity operator on V. The component in U of each ψ satisfying (1.1) could then be "factored out" in order to construct an unusual "gauge description" of a massive spin-$\frac{1}{2}$ particle, i.e., we could regard as physically equivalent two ψ's that differed only on U. Whether this would lead to new physics would depend on how the infinite component wave function (field) could be coupled to other fields.

Another possibility is that the equations (1.1) of interest here are not those with κ nonsingular, as usually considered, but rather ones with singular κ, associated with gauge descriptions of *massless* particles. Singular scalar operators arise naturally in the present context. For example, it follows from (2.12) that the operator whose matrix on V_{lm} is a_0 for $l<|l_1|$ and a_1 for $l\geqslant|l_1|$, is a nilpotent sl(2,$\mathbb{C}$) scalar, as is the operator whose matrix is zero for $l<|l_1|$ and δ for $l\geqslant|l_1|$.

Alternatively, these interesting representations of sl(2,$\mathbb{C}$) may of course have applications to physics, not involving relativistic wave equations (1.1) at all.[13,14,16] In any event, we hope to have made their study more accessible.

[1] I. M. Gel'fand and A. M. Yaglom, Zh. Eksp. Teor. Fiz. **18**, 703, 1096, 1105 (1948); I. M. Gel'fand, R. A. Minlos, and Z. Ya. Shapiro, *Representations of the Rotation and Lorentz Groups and their Applications* (Pergamon, Oxford, 1963).

[2]J. K. Lubanski, Physica **9**, 310 (1942); see also H. J. Bhabha, Rev. Mod. Phys. **17**, 200 (1945).
[3]Harish-Chandra, Phys. Rev. **71**, 793 (1947).
[4]A. Cant and C. A. Hurst, J. Aust. Math. Soc. B **20**, 446 (1978).
[5]F. L. Bauer, Sitzb. Bayer. Akad. Wiss, **13**, 111 (1952).
[6]M. Lorente, P. L. Huddleston, and P. Roman, J. Math. Phys. **14**, 1495 (1973).
[7]A. J. Bracken, J. Phys. A **8**, 800 (1975).
[8]A. Cant, J. Math. Phys. **23**, 354 (1982).
[9]A. Cant, J. Math. Phys. **22**, 870, 878 (1981).
[10]A. Cant and Y. Ne'eman, J. Math. Phys. **26**, 3180 (1985).
[11]A. O. Barut, C. K. E. Schneider, and R. Wilson, J. Math. Phys. **20**, 2244 (1979).
[12]As far as we are aware, representations of the group SL(2,C) have not been constructed, corresponding to the singular indecomposable representations of sl(2,C) presented in Ref. 15, except in the simplest nontrivial cases of operator-irreducible representations, as discussed, for example, in Ref. 19.
[13]P. A. M. Dirac, Proc. R. Soc. London Ser. A **183**, 284 (1945).
[14]P. A. M. Dirac, Int. J.Theor. Phys. **23**, 677 (1984).
[15]I. M. Gel'fand and V. A. Ponomarev, Usp. Mat. Nauk **23**, 3 (1968); translated in Russ. Math. Surveys **23**, 1 (1968).
[16]C. M. Bender and D. J. Griffiths, Phys. Rev. D **1**, 2335 (1970); **2**, 317 (1970); J. Math. Phys. **12**, 2151 (1971).
[17]L. Hlavaty and J. Niederle, J. Math. Phys. **22**, 1775 (1981); see also Czech. J. Phys. B **29**, 283 (1978).
[18]I. M. Gel'fand, M. I. Graev, and N. Ya. Vilenkin, *Generalized Functions* (Academic, New York, 1966), Vol. 5.
[19]W. Ruhl, *The Lorentz Group and Harmonic Analysis* (Benjamin, New York, 1970).
[20]B. Gruber, Proc. R. Irish Acad. A **82**, 13 (1972); B. Gruber and R. Lenczewski, J. Phys. A **16**, 3703 (1983); T. S. Santhanam, in *Group Theoretical Methods in Physics, Proceedings of the 14th International Colloquium*, edited by Y. M. Cho (World Scientific, Singapore, 1987).
[21]See, for example, A. Bohm, in *Lectures in Theoretical Physics*, edited by A. O. Barut and W. E. Brittin (Gordon and Breach, New York, 1968), Vol. 10B.
[22]Other matrix square roots seem either to be inconsistent with the requirements of (2.1), or to correspond to replacing the label l_1 by $-l_1$.
[23]A. J. Bracken, Phys. Rev. D **10**, 1168 (1974).
[24]A. S. Wightman, in *Invariant Wave Equations, Lecture Notes in Physics*, Vol. 73, edited by G. Velo and A. S. Wightman (Springer, New York, 1978).
[25]It would be possible also to consider the coupling of singular and nonsingular indecomposable representations. This has not been attempted.
[26]MUMATH is a trademark of The Soft Wharehouse, Hawaii.
[27]V. Bargmann, Math. Rev. **10–583**, 4 (1949).

On the degeneracy in the ground state of the $N=2$ Wess–Zumino supersymmetric quantum mechanics

Asao Arai
Department of Mathematics, Hokkaido University, Sapporo 060, Japan

(Received 2 May 1989; accepted for publication 26 July 1989)

It is known that the $N=2$ Wess–Zumino supersymmetric quantum mechanical model has $p-1$ degenerate zero-energy ground states consisting of only bosonic states, where $p\geqslant 3$ is the degree of the polynomial superpotential $V(z)\,(z\in\mathbb{C})$ of the model [Jaffe *et al.* Ann. Phys. (NY) **178**, 313 (1987)]. In this paper, the mathematical structure of the degenerate ground states is analyzed in the special case $V(z)=\lambda z^p$ $(\lambda>0)$. The following facts are discovered: (i) there exists a strongly continuous one parameter unitary group acting as a symmetry group in the quantum system under consideration; (ii) the generator of the symmetry group has infinitely many eigenspaces $\mathscr{H}_m$, $m\in\mathbb{Z}$, and the bosonic part H_+ of the Hamiltonian of the model is reduced by each of them; and (iii) there exist exactly $p-1$ $\mathscr{H}_m$'s in each of which the reduced part of H_+ has a unique zero-energy ground state. It is noted also that H_+ has infinitely many generalized eigenfunctions with eigenvalue zero. Moreover, a family of operators interrelating the zero-energy ground states is constructed. The coupling constant dependence of the nonzero eigenvalues of H_+ is exactly found.

I. INTRODUCTION

The $N=2$ Wess–Zumino supersymmetric quantum mechanics (SSQM) describes the interaction between a complex bosonic degree of freedom and two fermionic degrees of freedom,[1–4] and serves as a toy model of a supersymmetric quantum field theory.[5–8] It has been shown that the model with a polynomial superpotential $V(z)$ $(z\in\mathbb{C})$ of degree $p\geqslant 2$ has exactly $p-1$ zero-energy ground state(s) consisting of only bosonic state(s) and hence, if $p\geqslant 3$, then the ground state is degenerate (see Refs. 1 and 2 for formal discussions and Ref. 3 for a mathematically rigorous analysis). However, the origin of this degeneracy has not been clarified. The present work resulted from an attempt to understand the degeneracy in the ground state.

In SSQM, supersymmetry is said to be broken if no zero-energy states exist. It is well known that, if supersymmetry is broken, the ground state is always degenerate (e.g., Refs. 1, 3, 7, and 9). In the present case, however, supersymmetry is unbroken and hence the degeneracy in the ground state is not due to supersymmetry breaking. Therefore we infer that any mechanism different from supersymmetry breaking should exist to give rise to the degeneracy in the ground state. In Ref. 1, the following conjecture was given: The potential $|\partial V(z)|^2$, which appears as a potential term of the Hamiltonian of the model, will, in general, have $p-1$ wells; this leads to the $(p-1)$-fold degeneracy in the ground state. However, as pointed out there also, this reasoning cannot be applied, e.g., to the case $V(z)=z^p$, since in this case $|\partial V(z)|^2$ has only a single well. It seems that the structure of the degeneracy is not so simple and may vary according to the "fine structure" of the superpotential.

In this paper, as a first step towards the understanding of the degeneracy in the ground state of the Wess–Zumino model, we present a mathematically rigorous analysis of the model with $V(z)=\lambda z^p$, $\lambda>0$, concentrating our attention on the mathematical structure of the degenerate ground states. We discover the following facts: (i) there exists a strongly continuous one parameter unitary group $\{R_t=e^{itL}\,|\,t\in\mathbb{R}\}$ acting as a symmetry group in the quantum system under consideration and the generator L has infinitely many eigenspaces $\mathscr{H}_m$ $(m\in\mathbb{Z})$ decomposing the Hilbert space of bosonic states, so that the bosonic part H_+ of the total Hamiltonian of the model is reduced by each $\mathscr{H}_m$; (ii) there exist exactly $p-1$ $\mathscr{H}_m$'s in each of which H_+ has a unique ground state (see Sec. III). Thus we see that, in the present case, the degeneracy in the ground state is connected with the existence of a symmetry group. We note also that the generator L of R_t is of a form similar to that of the usual two-dimensional rotation group with spin $\frac{1}{2}$ and differs from it only in that the spin rotation part is $p-2$ times of the usual one [see (3.5)]. This suggests that there exists a degree of freedom with spin $s=(p-2)/2$ and hence any energy level of H_+ may be degenerate with multiplicity $2s+1=p-1$, which exactly coincides with the multiplicity of the degeneracy in the ground state of H_+. We note also that H_+ has infinitely many generalized eigenfunctions with eigenvalue zero. In Sec. IV, we show that H_+ has a dilation covariance, which exactly determines the λ (coupling constant) dependence of the nonzero eigenvalues of H_+. In the last section, we construct a family of operators interrelating the degenerate ground states. These operators are related to the dilation unitary group given in Sec. IV and the rotation group in $\mathbb{C}$.

We begin with a brief review of the $N=2$ Wess–Zumino model.

II. REVIEW OF THE $N=2$ WESS–ZUMINO SSQM[1–4]

The Hilbert space of state vectors of the model is realized as

$$\mathscr{H}=L^2(\mathbb{C};\mathbb{C}^4)=L^2(\mathbb{R}^2;\mathbb{C}^4)=\mathscr{H}_+\oplus\mathscr{H}_-, \tag{2.1}$$

where

 0022-2488/89/122973-05$02.50

$$\mathscr{H}_+ = \left\{ \begin{pmatrix} f_1 \\ f_2 \\ 0 \\ 0 \end{pmatrix} \middle| f_1, f_2 \in L^2(\mathbb{C}) \right\} \cong L^2(\mathbb{C};\mathbb{C}^2)$$

and

$$\mathscr{H}_- = \left\{ \begin{pmatrix} 0 \\ 0 \\ f_1 \\ f_2 \end{pmatrix} \middle| f_1, f_2 \in L^2(\mathbb{C}) \right\} \cong L^2(\mathbb{C};\mathbb{C}^2)$$

are the subspace of "bosonic" and "fermionic" states, respectively.

The fermionic degrees of freedom in the model are given by the 4×4 matrices

$$\psi_1 = \frac{1}{2}\begin{pmatrix} 0 & I+\sigma_3 \\ I-\sigma_3 & 0 \end{pmatrix},$$

$$\psi_2 = \frac{1}{2}\begin{pmatrix} 0 & i\sigma_1+\sigma_2 \\ -i\sigma_1-\sigma_2 & 0 \end{pmatrix},$$

where σ_j, $j=1, 2, 3$ are the Pauli matrices and I denotes the 2×2 identity matrix. It is straightforward to check that the ψ_j's satisfy the anticommutation relations

$$\{\psi_j, \psi_k^*\} = \delta_{jk}, \quad \{\psi_j, \psi_k\} = 0, \quad j,k = 1,2.$$

The self-adjoint supercharges of the model are defined by

$$Q_1 = i(\psi_2\bar\partial + \psi_2^*\partial) + i\{\psi_1(\partial V) - \psi_1^*(\partial V)^*\},$$

$$Q_2 = \psi_2\bar\partial - \psi_2^*\partial + \psi_1(\partial V) + \psi_1^*(\partial V)^*,$$

where $V(z)$ ($z\in\mathbb{C}$) is an entire function on $\mathbb{C}$, denoting the superpotential of the model, and $(\partial V)^*$ denotes the complex conjugate of ∂V. The operators Q_1 and Q_2 are essentially self-abjoint on C_0^∞ ($\mathbb{R}^2$; $\mathbb{C}^4$).[3,4] We denote their closures by the same symbols.

The Hamiltonian H of the model is defined by

$$H = Q_1^2.$$

It follows that

$$Q_1^2 = Q_2^2$$

and hence $D(Q_1) = D(Q_2) = D(H^{1/2})$, where $D(A)$ denotes the domain of the operator A. Further we have

$$\{Q_1, Q_2\} f = 0, \quad f\in C_0^\infty(\mathbb{R}^2; \mathbb{C}^4).$$

On $D(\partial\bar\partial)\cap D(|\partial V|^2)\cap D(\partial^2 V)$, H is expressed as

$$H = -\partial\bar\partial - \psi_1^*\psi_2(\partial^2 V)^* - \psi_2^*\psi_1\,\partial^2 V + |\partial V|^2.$$

Corresponding to the decomposition (2.1), H is decomposed as

$$H = H_+ \oplus H_-,$$

where

$$H_- = (-\partial\bar\partial + |\partial V|^2) I$$

and

$$H_+ = H_- + \begin{pmatrix} 0 & -i(\partial^2 V) \\ i(\partial^2 V)^* & 0 \end{pmatrix}. \tag{2.2}$$

The operators $H_\pm$ are essentially self-adjoint on C_0^∞($\mathbb{R}^2$; $\mathbb{C}^2$).[3,4]

The fermion number operator or the grading operator is given by

$$N_F = \begin{pmatrix} I & 0 \\ 0 & -I \end{pmatrix}.$$

The quadruple $\{\mathscr{H}, \{Q_1, Q_2\}, H, N_F\}$ satisfies the axioms of supersymmetric quantum theory.[9,10]

The following facts have been proved.

(i) No fermionic zero-energy states exist, that is,

$$\operatorname{Ker} H_- = \{0\}$$

(see Refs. 3 and 4).

(ii) If $V(z)$ is a polynomial of degree $p\geqslant 2$, then

$$\dim \operatorname{Ker} H_+ = p-1, \tag{2.3}$$

and hence, if $p\geqslant 3$, then the bosonic zero-energy state is degenerate.[3] As for the case of nonpolynomial superpotentials, we may conjecture that $\dim \operatorname{Ker} H_+ = \infty$. It has been proved in Ref. 4 that, in the special case $V(z) = \lambda e^{\alpha z}$ ($\lambda\in\mathbb{C}\setminus\{0\}$, $\alpha>0$), this is true.

III. GROUND STATES OF THE MODEL WITH A MONOMIAL SUPERPOTENTIAL

In what follows, we shall use the polar coordinate representation

$$z = re^{i\theta}, \quad r\in\mathbb{R}_+ = (0,\infty), \quad \theta\in[0,2\pi],$$

and the canonical identification

$$L^2(\mathbb{C}) = L^2(\mathbb{R}_+, d\mu)\otimes L^2(0,2\pi), \tag{3.1}$$

where $d\mu(r)$ is the measure on $\mathbb{R}_+$ given by

$$d\mu(r) = r\,dr. \tag{3.2}$$

In this section we consider the case of the monomial superpotential

$$V(z) = \lambda z^p, \quad \lambda>0, \quad p\geqslant 3, \tag{3.3}$$

so that H_+ given by (2.2) takes the form

$$H_+ = -\frac{1}{4r}\left(\frac{\partial}{\partial r}r\frac{\partial}{\partial r} + \frac{1}{r}\frac{\partial^2}{\partial\theta^2}\right) + \lambda^2 p^2 r^{2(p-1)} + \lambda i p(p-1) r^{p-2}\begin{pmatrix} 0 & -e^{i(p-2)\theta} \\ e^{-i(p-2)\theta} & 0 \end{pmatrix}, \tag{3.4}$$

on a suitable dense domain.

By (2.3), the zero-energy level of H_+ is degenerate with multiplicity $p-1\geqslant 2$. Our aim is to investigate the mathematical structure of the $p-1$ degenerate zero-energy ground states.

We first show that there exists a symmetry group acting in the quantum system governed by the bosonic Hamiltonian H_+. Let $\partial/\partial\theta$ be the generalized derivative with the periodic boundary condition in $[0,2\pi]$. Then, the operator

$$L = i\frac{\partial}{\partial\theta} + (p-2)\cdot\frac{\sigma_3}{2} \tag{3.5}$$

is self-adjoint with $D(L) = D(\partial/\partial\theta)\subset\mathscr{H}_+$ and hence

$$R_t = e^{itL}, \quad t\in\mathbb{R}, \tag{3.6}$$

generate a strongly continuous one parameter unitary group on $\mathscr{H}_+$.

Lemma 3.1: For all $t,s\in\mathbb{R}$,

$$R_t e^{isH_+} = e^{isH_+} R_t. \tag{3.7}$$

Proof: Direct computations show that, for all $t \in \mathbb{R}$, R_t takes $C_0^\infty(\mathbb{R}^2; \mathbb{C}^2)$ into $D(H_+)$ and

$$H_+ R_t \Psi = R_t H_+ \Psi, \quad \Psi \in C_0^\infty(\mathbb{R}^2; \mathbb{C}^2). \tag{3.8}$$

Since $C_0^\infty(\mathbb{R}^2; \mathbb{C}^2)$ is a core of H_+ (see Sec. II), (3.8) extends to all $\Psi \in D(H_+)$ showing at the same time that R_t takes $D(H_+)$ into $D(H_+)$. Then it follows that the resolvents, and hence the spectral projections of H_+, commute with R_t. Therefore we obtain (3.7). □

Lemma 3.1 is equivalent to saying that H_+ and L commute in the proper sense (e.g., Ref. 11, §VIII. 5) and shows that the unitary group $\{R_t\}_{t\in\mathbb{R}}$ is a symmetry group of the quantum system under consideration. We note that the generator L is only different in the coefficient of the spin rotation part $\sigma_3/2$ from that of the usual two-dimensional rotation group with spin $\frac{1}{2}$ (i.e., the case $p = 3$). Therefore, L may be regarded as the generator of a rotation group "distorted" with respect to spin degree of freedom. The form of L suggests that there exists a degree of freedom with spin $s = (p-2)/2$ and hence that any energy level of H_+ may be degenerate with multiplicity $2s + 1 = p - 1$, which exactly coincides with the multiplicity of the degeneracy in the ground state of H_+.

The spectrum $\sigma(L)$ of L is purely discrete and given as

$$\sigma(L) = \{m - (p/2) \mid m \in \mathbb{Z}\}. \tag{3.9}$$

The eigenspace of L with eigenvalue $m - (p/2)$ is given by

$$\mathscr{H}_m = \left\{ \begin{pmatrix} f \otimes e_{p-1-m} \\ g \otimes e_{1-m} \end{pmatrix} \ \middle|\ f,g \in L^2(\mathbb{R}_+, d\mu) \right\}, \tag{3.10}$$

where

$$e_n(\theta) = e^{in\theta}/\sqrt{2\pi}, \quad n \in \mathbb{Z}. \tag{3.11}$$

The Hilbert space $\mathscr{H}_+$ is decomposed as

$$\mathscr{H}_+ = \bigoplus_{m=-\infty}^{\infty} \mathscr{H}_m. \tag{3.12}$$

Lemma 3.1 implies that H_+ is reduced by each $\mathscr{H}_m$. We denote by $H_{+,m}$ the reduced part of H_+ to $\mathscr{H}_m$:

$$H_{+,m} = H_+ \restriction \mathscr{H}_m. \tag{3.13}$$

We set

$$F_\alpha(r) = \lambda r^{p-1} K_{\alpha/p}(2\lambda r^p), \quad r > 0, \quad \alpha \in \mathbb{R}, \tag{3.14}$$

where $K_\nu(z)$ is the modified Bessel function of the third kind (e.g., Ref. 12,§7.2.2). Let

$$\Phi_m = F_m \otimes e_{p-1-m}, \quad m \in \mathbb{Z}. \tag{3.15}$$

Proposition 3.2: For $m = 1,...,p-1$, $H_{+,m}$ has a unique zero-energy ground state (up to constant multiples), which is given by

$$\Omega_m = \begin{pmatrix} \Phi_m \\ -i\Phi^*_{p-m} \end{pmatrix}. \tag{3.16}$$

Proof: It follows from direct computations employing the formulas

$$zK'_\nu(z) + \nu K_\nu(z) = -zK_{\nu-1}(z)$$

and

$$K_\nu(z) = K_{-\nu}(z) \tag{3.17}$$

(e.g., Ref. 12) that, for $m = 1,...,p-1$, $\Omega_m \in D(H_{+,m})$ and

$$H_{+,m}\Omega_m = 0. \tag{3.18}$$

Combining this fact with (2.3), we conclude that Ω_m is a unique zero-energy state (up to constant multiples) of $H_{+,m}$. □

Remarks: (1) Each component of the zero-energy state Ω_m $(m = 1,...,p-1)$ is analytic in z and z^*. In fact, we have

$$\Phi_m(r,\theta) = \frac{\pi\lambda^{(p-m)/p} z^{p-1-m}}{2\sin(m\pi/p)} \sum_{n=0}^{\infty} \frac{\lambda^{2n} z^{np} z^{*np}}{n!} \times \left\{ \frac{1}{\Gamma(-(m/p)+n+1)} - \frac{\lambda^{2m/p} z^m z^{*m}}{\Gamma((m/p)+n+1)} \right\},$$

where $\Gamma(z)$ is the gamma function. This follows from the expansion of $K_\nu(x)$ in x (e.g., Ref. 12, §7.2.2).

(2) For $m \neq 1,...,p-1$, Ω_m is not in $\mathscr{H}_+$ because of the singularity at $r = 0$, but Eq. (3.18) still holds; that is, Ω_m is a generalized eigenfunction of H_+ with eigenvalue zero. Hence H_+ has infinitely many generalized eigenfunctions with eigenvalue zero. This also is a remarkable phenomenon.

IV. DILATION COVARIANCE AND THE COUPLING CONSTANT DEPENDENCE OF EIGENVALUES

Before analyzing interrelations between the zero-energy states Ω_m, we show that H_+ is dilation covariant. To express the dependence of H_+ on the coupling constant λ, we write H_+ as

$$H_+ = H_+(\lambda). \tag{4.1}$$

Let

$$(u_t f)(r) = tf(tr), \quad f \in L^2(\mathbb{R}_+, d\mu), \quad t > 0. \tag{4.2}$$

Then, it is easy to see that, for all $t > 0$, u_t is unitary on $L^2(\mathbb{R}_+, d\mu)$, strongly continuous in t, and satisfies

$$u_{t=1} = I, \quad u_t u_s = u_s u_t = u_{ts}, \quad t,s > 0, \tag{4.3}$$

where I denotes identity. Hence $\{u_t\}_{t>0}$ forms a strongly continuous unitary representation of the Abelian group (the "dilation group") $\mathbb{R}_+$. The operator u_t naturally extends to $\mathscr{H}_+$ as

$$U_t = (u_t \otimes I) \oplus (u_t \otimes I). \tag{4.4}$$

Obviously $\{U_t\}_{t>0}$ has the same properties as those of $\{u_t\}_{t>0}$ stated above.

Lemma 4.1: For all $\lambda > 0$ and $t > 0$,

$$U_t H_+(\lambda) U_t^{-1} = t^{-2} H_+(\lambda t^p). \tag{4.5}$$

Proof: This follows from the transformation properties

$$U_t \frac{\partial}{\partial r} U_t^{-1} = \frac{1}{t}\frac{\partial}{\partial r}, \quad U_t r U_t^{-1} = tr. \qquad \square$$

It is known that the spectrum $H_+(\lambda)$ is purely discrete.[3] Let

$$0 < E_1(\lambda) < E_2(\lambda) < \cdots < E_n(\lambda) < E_{n+1}(\lambda) < \cdots \tag{4.6}$$

be the nonzero eigenvalues of $H_+(\lambda)$.

Proposition 4.2: For each $n\geqslant 1$, there exists a constant $\alpha_n>0$ independent of λ such that

$$E_n(\lambda)=\alpha_n\lambda^{2/p},\quad \lambda>0. \tag{4.7}$$

Proof: By (4.5), we have

$$\sigma(H_+(\lambda))=t^{-2}\sigma(H_+(\lambda t^p)),$$

and hence, for all $n\geqslant 1$,

$$E_n(\lambda t^p)=t^2E_n(\lambda),$$

which gives

$$E_n(\lambda)/\lambda^{2/p}=E_n(\mu)/\mu^{2/p},$$

for all μ, $\lambda>0$. Therefore $\alpha_n\equiv E_n(\lambda)/\lambda^{2/p}$ is a constant independent of λ. Hence we obtain (4.7). □

Remark: The zero-energy states $\Omega_m\equiv\Omega_m(\lambda)$ are also dilation covariant:

$$U_t\Omega_m(\lambda)=\Omega_m(\lambda t^p),\quad \lambda>0,\quad t>0.$$

V. INTERRELATION BETWEEN THE ZERO-ENERGY GROUND STATES

In this section, we construct a family of operators interrelating the zero-energy states Ω_m.

For $\nu>\mu>(\nu-1)/2$, the function

$$a_{\mu\nu}(t)=t^{p(1-\nu)-1}/(t^{2p}-1)^{1+\mu-\nu},\quad t>1, \tag{5.1}$$

is integrable on $(1,\infty)$ with respect to dt. Hence

$$A_{\mu\nu}=\int_1^\infty a_{\mu\nu}(t)u_t\,dt \tag{5.2}$$

defines a bounded linear operator on $L^2(\mathbb{R}_+,d\mu)$, where the integral on the right-hand side of (5.2) is taken in the operator norm topology.

Let

$$W_\beta=\left\{f:\mathbb{R}_+\to\mathbb{C}\text{ measurable}\,\middle|\,\|f\|_\beta^2\right.$$
$$\left.\equiv\int_0^\infty(1+r^2)^\beta|f(r)|^2\,d\mu(r)<\infty\right\} \tag{5.3}$$

and

$$W_\infty=\bigcap_{\beta>0}W_\beta \tag{5.4}$$

with the Fréchet topology generated by the norms $\|\cdot\|_\beta$. We shall denote by $\mathscr{L}(W_\beta,W_\gamma)$ the family of all bounded linear operators from W_β to W_γ. Let M be the multiplication operator on $L^2(\mathbb{R}_+,d\mu)$ given by

$$(Mf)(r)=rf(r). \tag{5.5}$$

Lemma 5.1: Let $\alpha\geqslant 0$ and

$$\nu>\mu>(\nu-1)/2-\alpha/2p. \tag{5.6}$$

Then, for all $\beta\geqslant 0$, $M^\alpha A_{\mu\nu}\in\mathscr{L}(W_{\alpha+\beta},W_\beta)$. In particular, $M^\alpha A_{\mu\nu}$ takes continuously W_∞ into W_∞.

Proof: It is easy to see that, for all $t>0$, $M^\alpha u_t\in\mathscr{L}(W_{\alpha+\beta},W_\beta)$ with

$$\|M^\alpha u_tf\|_\beta\leqslant\frac{2^{\beta/2}(1+t^{2\beta})^{1/2}}{t^{\alpha+\beta}},\|f\|_{\alpha+\beta},\quad f\in W_{\alpha+\beta}.$$

Under condition (5.6), we have

$$\int_1^\infty a_{\mu\nu}(t)\frac{(1+t^{2\beta})^{1/2}}{t^{\alpha+\beta}}dt<\infty.$$

Therefore, $M^\alpha A_{\mu\nu}\in\mathscr{L}(W_{\alpha+\beta},W_\beta)$. □

We introduce the operator

$$B_{\mu\nu}=\frac{2p\lambda^{\nu-\mu}}{\Gamma(\nu-\mu)}M^{p(\nu-\mu)}A_{\mu\nu}, \tag{5.7}$$

which obeys the following composition law.

Lemma 5.2: For all $\mu<\nu<\rho$,

$$B_{\mu\nu}B_{\nu\rho}=B_{\mu\rho}. \tag{5.8}$$

Proof: The operator $B_{\mu\nu}$ is written as

$$(B_{\mu\nu}f)(r)=\int_0^\infty b_{\mu\nu}(r,s)f(s)d\mu(s),$$

with

$$b_{\mu\nu}(r,s)=\frac{2p\lambda^{\nu-\mu}}{\Gamma(\nu-\mu)}a_{\mu\nu}\left(\frac{s}{r}\right)r^{p(\nu-\mu)-2}\chi_{(r,\infty)}(s),$$

where $\chi_{(r,\infty)}$ is the characteristic function of the interval (r,∞). Then, by direct computations, we can show that

$$\int_0^\infty b_{\mu\nu}(r,s)b_{\nu\rho}(s,t)d\mu(s)=b_{\mu\rho}(r,t)$$

holds. Therefore, (5.8) follows. □

We note also that $B_{\mu\nu}$ is dilation covariant:

$$u_tB_{\mu\nu}u_t^{-1}=t^{p(\nu-\mu)}B_{\mu\nu},\quad t>0. \tag{5.9}$$

Lemma 5.3: Let F_α be given by (3.14). Then, $F_\alpha\in W_\infty$, $0\leqslant\alpha<p$, and

$$F_\alpha=B_{\alpha/p,\beta/p}F_\beta,\quad 0\leqslant\alpha<\beta<p. \tag{5.10}$$

Proof: The fact $F_\alpha\in W_\infty$ follows from the asymptotic properties of the Bessel function

$$K_\nu(x)\sim(\pi/2x)^{1/2}e^{-x},\quad\text{as }x\to\infty,$$
$$K_\nu(x)\sim\text{const }x^{-\nu},\quad\text{as }x\downarrow 0\quad(\nu>0),$$
$$K_0(x)\sim\text{const}\log x,\quad\text{as }x\downarrow 0$$

(see, e.g., Ref. 12). To prove (5.10), we note that the following formula holds [Ref. 12, 7.14.2, formula (50)]:

$$\frac{2^{\nu-\mu-1}\Gamma(\nu-\mu)}{a^{\nu-\mu}y^\mu}K_\mu(ay)$$
$$=\int_0^\infty\frac{x^{2(\nu-\mu)-1}}{(x^2+y^2)^{\nu/2}}K_\nu(a\sqrt{x^2+y^2})dx, \tag{5.11}$$

for $\operatorname{Re}\nu>\operatorname{Re}\mu$, $a>0$, and $y>0$. By direct computations we see that (5.11) with $a=2\lambda$, $y=r^p$, $\mu=\alpha/p$, and $\nu=\beta/p$ is equivalent to (5.10). □

Let u be the operator on $L^2(0,2\pi)$ defined by

$$(uf)(\theta)=e^{-i\theta}f(\theta),\quad f\in L^2(0,2\pi), \tag{5.12}$$

and $T_{\mu\nu}^\pm$ be the operators on $\mathscr{H}_+$ given by

$$T_{\mu\nu}^+=(B_{-\mu/p,-\nu/p}\otimes u^{\mu-\nu})$$
$$\oplus(B_{(p-\mu)/p,(p-\nu)/p}\otimes u^{\mu-\nu}),\quad\mu>\nu, \tag{5.13}$$

$$T_{\mu\nu}^-=(B_{\mu/p,\nu/p}\otimes u^{\mu-\nu})$$
$$\oplus(B_{(\mu-p)/p,(\nu-p)/p}\otimes u^{\mu-\nu}),\quad\mu<\nu. \tag{5.14}$$

Proposition 5.4: For all m, $n=1,\ldots,p-1$, $m<n$,

$$\Omega_n=T_{nm}^+\Omega_m,\quad\Omega_m=T_{mn}^-\Omega_n. \tag{5.15}$$

Proof: Property (3.17) implies that

$$F_\alpha = F_{-\alpha}.$$

Obviously we have

$$ue_m = e_{m-1}, \quad m\in\mathbb{Z}. \tag{5.16}$$

Combining these facts with (3.15) and (5.10), we obtain (5.15). □

Equation (5.15) shows that $T^{\pm}_{mn}$ are operators interrelating the zero-energy states Ω_m. It follows from (5.15) also that Ker H_+ has the following structures:

$$\operatorname{Ker} H_+ = \left\{\left(\beta_1 + \sum_{j=2}^{p-1} \beta_j T^+_{j1}\right)\Omega_1 \,|\, \beta_j\in\mathbb{C},\ j = 1,\dots,p-1\right\}$$

$$= \left\{\left(\sum_{j=1}^{p-2} \beta_j T^-_{j,p-1} + \beta_{p-1}\right)\Omega_{p-1} \,|\, \beta_j\in\mathbb{C},\ j = 1,\dots,p-1\right\}.$$

Finally we briefly discuss properties of $T^{\pm}_{\mu\nu}$. By (5.16), we see that, if $\mu-\nu\in\mathbb{Z}$, then $T^{\pm}_{\mu\nu}:\mathscr{H}_m \to \mathscr{H}_{m+\mu-\nu}$, $m\in\mathbb{Z}$. Further, $T^{\pm}_{\mu\nu}$ are dilation covariant:

$$U_t T^{\pm}_{\mu\nu} U_t^{-1} = t^{p(\nu-\mu)} T^{\pm}_{\mu\nu}, \quad t>0.$$

This follows from (5.9).

Remark: We have not been able to clarify the commutation relations of $T^{\pm}_{\mu\nu}$ with themselves and H_+. It seems that they are not so simple.

[1]D. Lancaster, "Supersymmetry breakdown in supersymmetric quantum mechanics," Nuovo Cimento A **79**, 28 (1984).

[2]M. Claudson and M. B. Halpern, "Supersymmetric ground state wave functions," Nucl. Phys. B **250**, 689 (1985).

[3]A. Jaffe, A. Lesniewski, and M. Lewenstein, "Ground state structure in supersymmetric quantum mechanics," Ann. Phys. (NY) **178**, 313 (1987).

[4]A. Arai, "Existence of infinitely many zero-energy states in a model of supersymmetric quantum mechanics," J. Math. Phys. **30**, 1164 (1989).

[5]J. Wess and B. Zumino, "Supergauge transformations in four dimensions," Nucl. Phys. B **70**, 39 (1974).

[6]J. Wess and B. Zumino, "A Lagrangian model invariant under supergauge transformations, Phys. Lett. B **49**, 52 (1974).

[7]E. Witten, Constraints on supersymmetry breaking," Nucl. Phys. B **202**, 253 (1982).

[8]A. Jaffe, A. Lesniewski, and J. Weitsman, "Index of a family of Dirac operators on loop space," Commun. Math. Phys. **112**, 75 (1987).

[9]A. Arai, "Supersymmetry and singular perturbations," J. Funct. Anal. **60**, 378 (1985).

[10]H. Grosse and L. Pittner, "Supersymmetric quantum mechanics defined as sesquilinear forms," J. Phys. A: Math. Gen. **20**, 4265 (1987).

[11]M. Reed and B. Simon, *Methods of Modern Mathematical Physics. Vol. I: Functional Analysis* (Academic, New York, 1972).

[12]*Higher Transcendental Functions*, Vol. II, edited by A. Erdélyi (McGraw-Hill, New York, 1953).

$N=3$ extended supersymmetric gauge theories and an explicit construction of higher conservation laws

C. Devchand
Fakultät für Physik der Universität, D-7800 Freiburg, Federal Republic of Germany

(Received 15 March 1989; accepted for publication 12 July 1989)

A Lagrangian for the superfield equations of motion for supersymmetric gauge theories with $N=3$ extended supersymmetry is presented. A novel formulation of the previously constructed infinitely many spinorial continuity equations considerably clarifies their structure.

I. INTRODUCTION

The maximally supersymmetric gauge theories in four dimensions[1,2] with $N=4$ (or equivalently $N=3$) supersymmetry have, among their many remarkable properties, the feature of ultraviolet finiteness to all orders in perturbation theory.[3–5] The conjunction of conformal invariance and solubility is a familiar feature of many quantum field theories and the strongly constrained dynamics implied by the ultraviolet finiteness suggests integrability of the field equations as a possible underlying classical precursor. The finiteness of maximal super-Yang–Mills therefore lends further promise to the notion that these are the appropriate theories for the realization of duality conjectures[6] which generalize the duality between the Thirring model and the sine–Gordon model[7] to four-dimensional gauge theories with monopole solutions. It is, however, clear that the equations of motion for the $N=4$ theory with Lagrangian[2]

$$\begin{aligned} L = (&-\tfrac{1}{4}F_{\mu\nu}F^{\mu\nu} + \tfrac{1}{2}D_\mu \Phi_{ij} D^\mu \Phi^{ij} \\ &- i\lambda^{i} \not{D}\lambda_i + \lambda^{i}[\lambda^{cj},\Phi_{ij}] \\ &- \lambda^c_i[\lambda_j,\Phi^{ij}] + \tfrac{1}{4}[\Phi_{ij},\Phi_{kl}][\Phi^{ij},\Phi^{kl}]) \end{aligned} \tag{1}$$

[where all fields are gauge algebra valued; i,j run from 1 to 4; $\Phi^{ij} = \frac{1}{2}\epsilon^{ijkl}\Phi_{kl}$ and the spinors λ_i are chiral, $\gamma_5\lambda = -i\lambda$ and $\lambda^{ci} \equiv C\lambda^{i}$ yielding $\gamma_5\lambda^c = i\lambda^c$] are *not* completely integrable in any conventional sense; nor is the S matrix of the theory trivial, as would be expected of a four-dimensional theory with higher local conserved currents.[8] Nevertheless, in the superspace formulation,[9] the classical equations of motion for this theory, similarly to the Yang–Mills self-duality equations,[10] may be formulated in a geometric way as integrability conditions for a set of linear (superfield) equations based on the Witten–Manin supertwistor correspondence.[11–13] This structural similarity to completely integrable systems with solitons has led to many (hitherto largely futile) attempts (e.g., Refs. 14–19) to obtain meaningful explicit results on classical solutions and higher conservation laws and symmetries using methods analogous to those developed to study self-dual Yang–Mills or soliton equations of the Zakharov–Shabat type. This paper is a further contribution in this direction. We present, in Sec. II, a novel Lagrangian for a reformulated set of superfield equations of motion. This Lagrangian is rather similar to the Lagrangian for self-dual Yang–Mills discussed by Leznov and Saveliev.[20] In Sec. III we discuss the infinite number of continuity equations, first introduced in Ref. 15, which in the new formulation of this paper are considerably clarified. They acquire a more explicit form that is perhaps more amenable to interpretation. In Sec. IV we comment on the infinitesimal symmetry transformations of the equations of motion and we conclude (Sec. V) with some remarks on the reduction of these equations to the supersymmetric self-duality conditions.

II. $N=3$ SUPER-YANG–MILLS EQUATIONS AND A SUPERFIELD LAGRANGIAN

N-extended complexified superspace, a supermanifold of complex dimension $(4/4N)$ with complexified space-time coordinates $x^{\alpha\dot\beta} = x^\mu \sigma_\mu^{\alpha\dot\beta}$ and the anticommuting coordinates ϑ^α_i, $\bar\vartheta^{\dot\alpha j}$, where α, $\dot\alpha$ are two-component spinor indices and $i,j, = 1,\ldots,N$ is the internal SU(N) index, the upper and lower indices referring to fundamental and conjugate representations, respectively. The supertranslation vector fields $\partial_A = (\partial_{\alpha\dot\alpha}, \mathrm{D}^i_\alpha, \mathrm{D}_{\dot\alpha j})$,

$$\mathrm{D}^i_\alpha = \frac{\partial}{\partial \vartheta^\alpha_i} + \bar\vartheta^{\dot\beta i}\partial_{\alpha\dot\beta}, \quad \mathrm{D}_{\dot\beta j} = \frac{\partial}{\partial \bar\vartheta^{\dot\beta j}} + \vartheta^\alpha_j \partial_{\alpha\dot\beta},$$

$$\partial_{\alpha\dot\beta} = \frac{\partial}{\partial x^{\alpha\dot\beta}}, \tag{2}$$

provide a nonholonomic frame for superspace and realize the superalgebra

$$\begin{aligned} &\{\mathrm{D}^i_\alpha, \mathrm{D}^j_\beta\} = 0 = \{\mathrm{D}_{\dot\alpha i}, \mathrm{D}_{\dot\beta j}\}, \\ &\{\mathrm{D}^i_\alpha, \mathrm{D}_{\dot\beta j}\} = 2\delta^i_j \partial_{\alpha\dot\beta}, \\ &[\partial_{\alpha\dot\beta}, \mathrm{D}^i_\beta] = 0 = [\partial_{\alpha\dot\beta}, \mathrm{D}_{\dot\alpha j}] = [\partial_{\alpha\dot\alpha}, \partial_{\beta\dot\beta}]. \end{aligned} \tag{3}$$

The Lie-algebra valued components of the gauge superconnection $A_A = (A_{\alpha\dot\beta}, A^i_\alpha, A_{\dot\beta j})$ transform as usual under gauge transformations

$$A_A \to e^\Lambda A_A e^{-\Lambda} + e^\Lambda \partial_A e^{-\Lambda} \tag{4}$$

a covariant superfield transforming as $\Phi \to e^\Lambda \Phi e^{-\Lambda}$, where the gauge parameter Λ is a Lie algebra valued superfield, $\Lambda = T^a \Lambda^a(x,\vartheta,\bar\vartheta)$, T^a being the Lie algebra generators acting on the gauge indices of A_A and Φ. Introducing gauge-covariant derivatives $D_A = (D_{\alpha\dot\beta}, D^i_\alpha, D_{\dot\beta j})$,

$$\begin{aligned} &D^i_\alpha \varphi = \mathrm{D}^i_\alpha \varphi + [A^i_\alpha, \varphi], \\ &D_{\dot\alpha j}\varphi = \mathrm{D}_{\dot\alpha j}\varphi + [A_{\dot\alpha j}, \varphi], \\ &D_{\alpha\dot\beta}\varphi = \partial_{\alpha\dot\beta}\varphi + [A_{\alpha\dot\beta}, \varphi], \end{aligned} \tag{5}$$

transforming as $D_A \to e^\Lambda (D_A \varphi) e^{-\Lambda}$, the superfield curvatures F_{AB} may be formed by considering the graded commu-

 0022-2488/89/122978-06$02.50

tator,

$$
\begin{aligned}
&\{D^i_{\alpha},D^j_{\beta}\}=F^{ij}_{\alpha\beta}, \quad \{D_{\dot\alpha i},D_{\dot\beta j}\}=F_{\dot\alpha i,\dot\beta j},\\
&[D_{\alpha\dot\beta},D^i_{\beta}]=F^i_{\alpha\dot\beta,\beta}, \quad [D_{\alpha\dot\beta},D_{\dot\alpha i}]=F_{\alpha\dot\beta,\dot\alpha i},\\
&[D_{\alpha\dot\beta},D_{\gamma\dot\delta}]=\epsilon_{\alpha\gamma}F_{\dot\beta\dot\delta}+\epsilon_{\dot\beta\dot\delta}F_{\alpha\gamma},\\
&\{D^i_{\alpha},D_{\dot\beta j}\}=F^i_{\alpha,\dot\beta j}+2\delta^i_j D_{\alpha\dot\beta},
\end{aligned} \tag{6}
$$

where

$$
\begin{aligned}
&F^{ij}_{\alpha\beta}=\mathrm{D}^i_{\alpha}A^j_{\beta}+\mathrm{D}^j_{\beta}A^i_{\alpha}+\{A^i_{\alpha},A^j_{\beta}\},\\
&F_{\dot\alpha i,\dot\beta j}=\mathrm{D}_{\dot\alpha i}A_{\dot\beta j}+\mathrm{D}_{\dot\beta j}A_{\dot\alpha i}+\{A_{\dot\alpha i},A_{\dot\beta j}\},\\
&F^i_{\alpha,\dot\beta j}=\mathrm{D}^i_{\alpha}A_{\dot\beta j}+\mathrm{D}_{\dot\beta j}A^i_{\alpha}+\{A^i_{\alpha},A_{\dot\beta j}\}-2\delta^i_j A_{\alpha\dot\beta},\\
&F^i_{\mu\alpha}=\partial_{\mu}A^i_{\alpha}-\mathrm{D}^i_{\alpha}A_{\mu}+[A_{\mu},A^i_{\alpha}],\\
&F_{\mu,\dot\alpha i}=\partial_{\mu}A_{\dot\alpha i}-\mathrm{D}_{\dot\alpha i}A_{\mu}+[A_{\mu},A_{\dot\alpha i}],
\end{aligned} \tag{7}
$$

and $F_{\alpha\gamma}(F_{\dot\beta\dot\delta})$ are the (anti-) self-dual parts of $F_{\mu\nu} = \partial_{\mu}A_{\nu}-\partial_{\nu}A_{\mu}+[A_{\mu},A_{\nu}]$.

All the above F_{AB}'s are Lie algebra valued superfields transforming covariantly under gauge transformations. They are clearly not independent of each other, since the generalized Jacobi identity imposes relations amongst them, e.g.,

$$
\begin{aligned}
0\equiv{}&[D^i_{\alpha},\{D_{\dot\alpha j},D_{\dot\beta k}\}]+[D_{\dot\alpha j},\{D_{\dot\beta k},D^i_{\alpha}\}]\\
&+[D_{\dot\beta k},\{D^i_{\alpha},D_{\dot\alpha j}\}]\\
={}&D^i_{\alpha}F_{\dot\alpha j,\dot\beta k}+D_{\dot\alpha j}F^i_{\alpha,\dot\beta k}+D_{\dot\beta k}F^i_{\alpha,\dot\alpha j}\\
&-2\delta^i_k F_{\alpha\dot\beta,\dot\alpha j}-2\delta^i_j F_{\alpha\dot\alpha,\dot\beta k}
\end{aligned} \tag{8}
$$

or

$$
\begin{aligned}
0\equiv{}&[D_{\alpha\dot\beta}+\{D^i_{\beta},D_{\dot\alpha j}\}]+\{D^i_{\beta},[D_{\dot\alpha j},D_{\alpha\dot\beta}]\}\\
&+\{D_{\dot\alpha j},[D^i_{\beta},D_{\alpha\dot\beta}]\}\\
={}&D_{\alpha\dot\beta}F^i_{\beta,\dot\alpha j}-D^i_{\beta}F_{\alpha\dot\beta,\dot\alpha j}-D_{\dot\alpha j}F^i_{\alpha\dot\beta,\beta}\\
&+2\delta^i_j(\epsilon_{\alpha\beta}F_{\dot\alpha\dot\beta}+\epsilon_{\dot\alpha\dot\beta}F_{\alpha\beta}).
\end{aligned} \tag{9}
$$

Nevertheless, the set of super Yang–Mills fields F_{AB} defined by Eq. (7) clearly has an enormous number of degrees of freedom; and in order to minimize the number of component fields to just those required for the construction of an irreducible representation of the supersymmetry algebra, one usually imposes the covariant (under both supersymmetry and gauge transformations) constraint equations,[9]

$$F^{(ij)}_{\alpha\beta}=0=F_{\dot\alpha(i,\dot\beta j)}, \tag{10a}$$

$$F^i_{\alpha,\dot\beta j}=0. \tag{10b}$$

For $N<3$, these equations do not imply any equations in x space. They are therefore a suitable representation condition for a minimal multiplet of component fields. For $N\geqslant 3$, on the other hand, the constraints (10) do have dynamical content and do not yield an off-shell representation of the theory with only the minimal physical fields (viz., one spin 1, four spin $\frac{1}{2}$, and six spin 0). Remarkably, for the case of $N=3$, the constraints (10) are precisely equivalent to the equations of motion for the component fields.[21]

Equations (10a) have the solution

$$F^{ij}_{\alpha\beta}=\epsilon_{\alpha\beta}W^{ij}, \quad W^{ij}=\epsilon^{ijk}W_k, \tag{11a}$$

$$F_{\dot\alpha i,\dot\beta j}=\epsilon_{\dot\alpha\dot\beta}W_{ij}, \quad W_{ij}=\epsilon_{ijk}W^k, \tag{11b}$$

whereas the diagonal parts of (10b) are the conventional equations expressing the vector potential $A_{\alpha\dot\beta}$ entirely in terms of A^i_{α} and $A_{\dot\alpha i}$ in virtue of (7),

$$A_{\alpha\dot\beta}=\tfrac{1}{6}(\mathrm{D}^i_{\alpha}A_{\dot\beta i}+\mathrm{D}_{\dot\beta i}A^i_{\alpha}+\{A^i_{\alpha},A_{\dot\beta i}\}). \tag{12}$$

The Bianchi identities can now be used to express the theory entirely in terms of the superfields $F_{\alpha\beta}$, $F_{\dot\alpha\dot\beta}$, W^k, W_k, $D^i_{\alpha}W_i$, $D_{\dot\alpha i}W^i$, $\epsilon_{ijk}D^i_{\alpha}W^j$, $\epsilon^{ijk}D_{\dot\alpha i}W_j$, and covariant derivatives thereof. The leading components (in a power series expansion in ϑ, $\bar\vartheta$) of $F_{\alpha\beta}$, $F_{\dot\alpha\dot\beta}$ yield the field strength components of the component vector field $A_{\alpha\dot\beta}(x)$, whereas the leading components of the other superfields yield the remaining component fields of the theory: two SU(3)-triplets of scalar fields (W^i,W_i), an SU(3)-scalar Majorana spinor $(\lambda_{\alpha},\lambda_{\dot\alpha})$, and an SU(3)-triplet of spinors $(\chi_{\alpha i},\chi^i_{\dot\alpha})$. These component fields have the following equations of motion [corresponding to the $N=3$ version of the theory (1)]:

The Dirac equations

$$\epsilon^{\alpha\beta}D_{\alpha\dot\beta}\lambda_{\beta}+[\chi^i_{\dot\beta},W_i]=0, \tag{13a}$$

$$\epsilon^{\dot\alpha\dot\beta}D_{\alpha\dot\alpha}\lambda_{\dot\beta}+[\chi_{i\alpha},W^i]=0, \tag{13b}$$

$$\epsilon^{\alpha\beta}D_{\alpha\dot\beta}\chi_{i\beta}+[\chi^j_{\dot\beta},W^k]\epsilon_{ijk}-[\lambda_{\dot\beta},W_i]=0, \tag{13c}$$

$$\epsilon^{\dot\alpha\dot\beta}D_{\alpha\dot\alpha}\chi^j_{\dot\beta}+[\chi_{i\alpha},W_k]\epsilon^{ijk}-[\lambda_{\alpha},W^j]=0, \tag{13d}$$

the scalar field equations

$$
\begin{aligned}
\epsilon^{\alpha\beta}\epsilon^{\dot\alpha\dot\beta}&D_{\alpha\dot\alpha}D_{\beta\dot\beta}W_j+2[[W^i,W_j],W_i]\\
&-[[W^i,W_i],W_j]+\epsilon^{\alpha\beta}\{\chi_{\alpha j},\lambda_{\beta}\}\\
&-\tfrac{1}{2}\epsilon_{ijk}\epsilon^{\dot\alpha\dot\beta}\{\chi^i_{\dot\alpha},\chi^k_{\dot\beta}\}=0
\end{aligned} \tag{13e}
$$

(similarly for W_i), and the Yang–Mills equation

$$
\begin{aligned}
\epsilon^{\alpha\beta}&D_{\alpha\dot\beta}F_{\gamma\beta}+\epsilon^{\dot\alpha\dot\gamma}D_{\gamma\dot\alpha}F_{\dot\gamma\dot\beta}+\{\chi_{\gamma k},\chi^k_{\dot\beta}\}+\{\lambda_{\gamma},\lambda_{\dot\beta}\}\\
&+[W_i,D_{\gamma\dot\beta}W^i]+[W^i,D_{\gamma\dot\beta}W_i]=0.
\end{aligned} \tag{13f}
$$

The equivalence proof of Ref. 21 eliminates the ϑ dependence of the gauge transformations (4) by subjecting the superconnection to the "transverse" gauge condition,

$$\vartheta^{\alpha}_{i}A^i_{\alpha}+\bar\vartheta^{\dot\alpha j}A_{\dot\alpha j}=0, \tag{14}$$

which effectively eliminates all ϑ-dependent gauge transformations, while posing no restriction on x-dependent gauge transformations of the component fields. This allows the construction of a unique correspondence between a superconnection constrained by (10) and a component multiplet solving (13). For the $N=3$ theory, therefore, the constraint equations (10) are just a compact way of writing the field equations (13). These constraints equations, remarkably, also correspond to the vanishing of the supercurvature along supernull lines.[11] Consider the equations for the covariant constancy of sections along the direction in superspace given by $\xi^A=(\mu^{\alpha}\lambda^{\dot\alpha},\mu^{\alpha},\lambda_{\dot\alpha})$,

$$\mu^{\alpha}D^i_{\alpha}\Phi=0=\lambda^{\dot\alpha}D_{\dot\alpha j}\Phi, \tag{15a}$$

$$\mu^{\alpha}\lambda^{\dot\alpha}D_{\alpha\dot\alpha}\Phi=0, \tag{15b}$$

where $v^{\alpha\dot\alpha}=\mu^{\alpha}\lambda^{\dot\alpha}$ is a null vector in x space and ξ^A is defined to be a supernull vector. Integrability of (15) (i.e., the path independence of Φ) requires $\xi^A A_A=(\mu^{\alpha}\lambda^{\dot\alpha}A_{\alpha\dot\alpha}, \mu^{\alpha}A^i_{\alpha},\lambda^{\dot\alpha}A_{\dot\alpha i})$ to be a flat connection. This is tantamount to requiring that the supercurvature components satisfy (10), and allows the identification[11] of bundles with connections satisfying constraints (10) over super-Minkowski space with bundles over the supertwistor space of supernull lines with triviality conditions over certain $CP^1 \times CP^1$ submanifolds. It is the tantalizing similarity of (15) to Lax-type linear systems for soliton equations which led to previous attempts[14–19] to understand these theories using solitonic methods.

Following our previous approach,[15] we note that the following subset of constraints (10):

$$F^{ij}_{11} = 0 = F^{ij}_{22}, \tag{16}$$

$$F_{\dot{1}i,\dot{1}j} = 0 = F_{\dot{2}i,\dot{2}j}, \tag{17}$$

$$F^{i}_{1,\dot{1}j} = 0 = F^{i}_{2,\dot{2}j}, \tag{18}$$

are equivalent to writing the spinor potentials and two components of the vector potential in pure gauge form,

$$A^i_1 = g^{-1}D^i_1 g, \quad A_{\dot{1}i} = g^{-1}D_{\dot{1}i}g, \quad A_{1\dot{1}} = g^{-1}\partial_{1\dot{1}}g, \tag{19}$$

$$A^i_2 = h^{-1}D^i_2 h, \quad A_{\dot{2}i} = h^{-1}D_{\dot{2}i}h, \quad A_{2\dot{2}} = h^{-1}\partial_{2\dot{2}}h. \tag{20}$$

Introducing the matrix

$$B = gh^{-1}, \tag{21}$$

the remaining components of the vector potential then take the form

$$A_{1\dot{2}} = g^{-1}\partial_{1\dot{2}}g + \tfrac{1}{6}g^{-1}D^i_1(B\,D_{\dot{2}i}B^{-1})g \tag{22a}$$

$$= h^{-1}\partial_{1\dot{2}}h + \tfrac{1}{6}h^{-1}D_{\dot{2}i}(B^{-1}D^i_1 B)h, \tag{22b}$$

$$A_{2\dot{1}} = g^{-1}\partial_{2\dot{1}}g + \tfrac{1}{6}g^{-1}D_{\dot{1}i}(B\,D^i_2 B^{-1})g \tag{23a}$$

$$= h^{-1}\partial_{2\dot{1}}h + \tfrac{1}{6}h^{-1}D^i_2(B^{-1}D_{\dot{1}i}B)h \tag{23b}$$

in virtue of (10b).

A gauge transformation (4) of these potentials (19)–(23) corresponds to the transformation

$$g \to Uge^{-\Lambda}, \quad h \to Vhe^{-\Lambda}, \tag{24}$$

where Λ is an arbitrary matrix superfield in the (complexified) gauge algebra, which we take to be $gl(n,\mathbb{C})$, and the matrices U and V are $GL(n,\mathbb{C})$ matrices satisfying

$$D^i_1 U = 0 = D_{\dot{1}i}U, \quad \partial_{1\dot{1}}U = 0, \tag{25a}$$

$$D^i_2 V = 0 = D_{\dot{2}i}V, \quad \partial_{2\dot{2}}V = 0. \tag{25b}$$

The remaining constraints in (10) may now be multiplied by g on the left and g^{-1} on the right to yield the equivalent equations,

$$D^{(i}_1 B\,D^{j)}_2 B^{-1}) = 0, \tag{26a}$$

$$D_{\dot{1}(i}(B\,D_{\dot{2}j)}B^{-1}) = 0, \tag{26b}$$

$$D^i_1(B\,D_{\dot{2}j}B^{-1}) - 2\delta^i_j g D_{1\dot{2}}g^{-1} = 0, \tag{26c}$$

$$D_{\dot{1}i}(B\,D^j_2 B^{-1}) - 2\delta^j_i g D_{2\dot{1}}g^{-1} = 0. \tag{26d}$$

These equations transform covariantly under the U transformation in (24), being manifestly invariant under the V- and Λ-dependent parts of (24). Equations (26) may be solved by writing

$$B\,D^i_2 B^{-1} = D^i_1 x, \tag{27a}$$

$$B\,D_{\dot{2}j}B^{-1} = D_{\dot{1}j}y, \tag{27b}$$

where x and y are matrix superfields transforming under the gauge transformation (24) as

$$x \to UxU^{-1}, \quad y \to UyU^{-1}. \tag{28}$$

These superfields satisfy the equations

$$D_{\dot{1}j}x = 0 = D^i_1 y \tag{29}$$

in order to satisfy the nondiagonal parts of (26c), (26d) yielding the following expressions for the potentials $A_{1\dot{2}}, A_{2\dot{1}}$ of (22), (23):

$$A_{1\dot{2}} = g^{-1}(\partial_{1\dot{2}} + \partial_{1\dot{1}}y)g, \tag{30a}$$

$$A_{2\dot{1}} = g^{-1}(\partial_{2\dot{1}} + \partial_{1\dot{1}}x)g. \tag{30b}$$

Similarly multiplying the curvature components in (16)–(18) by g on the left and g^{-1} on the right, we see that the first equality in (16)–(18) is identically satisfied, leaving the following forms of the right-hand side equations:

$$D^{(i}_2(B\,D^{j)}_2 B^{-1}) + \{B\,D^i_2 B^{-1}, B\,D^j_2 B^{-1}\} = 0, \tag{31a}$$

$$D_{\dot{2}(i}(B\,D_{\dot{2}j)}B^{-1}) + \{B\,D_{\dot{2}i}B^{-1}, B\,D_{\dot{2}j}B^{-1}\} = 0, \tag{31b}$$

$$\begin{aligned} &D^i_2(B\,D_{\dot{2}j}B^{-1}) + D_{\dot{2}j}(B\,D^i_2 B^{-1}) \\ &\quad + \{B\,D^i_2 B^{-1}, B\,D_{\dot{2}j}B^{-1}\} \\ &\quad = 2\delta^i_j g D_{2\dot{2}}g^{-1}. \end{aligned} \tag{31c}$$

Inserting the solution (27) into these equations yields the equations

$$D^{(i}_2 D^{j)}_1 x + \{D^i_1 x, D^j_1 x\} = 0, \tag{32a}$$

$$D_{\dot{2}(i}D_{\dot{1}j)}y + \{D_{\dot{1}i}y, D_{\dot{1}j}y\} = 0, \tag{32b}$$

$$\begin{aligned} &D^i_2 D_{\dot{1}j}y + D_{\dot{2}j}D^i_1 x + \{D^i_1 x, D_{\dot{1}j}y\} \\ &\quad = 2\delta^i_j g D_{2\dot{2}}g^{-1}. \end{aligned} \tag{32c}$$

The last of these may be written, using (29),

$$\begin{aligned} &D^i_1 D_{\dot{2}j}x + D_{\dot{1}j}D^i_2 y + D_{\dot{1}j}[D^i_1 x, y] \\ &\quad = 2\delta^i_j(\partial_{1\dot{2}}x + \partial_{2\dot{1}}y + [\partial_{1\dot{1}}x, y] - gD_{2\dot{2}}g^{-1}). \end{aligned} \tag{33}$$

The nondiagonal parts of this equation are satisfied if the superfields x and y, in addition to (29), satisfy

$$D_{\dot{2}j}x = 0, \tag{34a}$$

$$D^i_2 y + [D^i_1 x, y] = 0, \tag{34b}$$

yielding

$$A_{2\dot{2}} = g^{-1}(\partial_{2\dot{2}} + \partial_{1\dot{2}}x + \partial_{2\dot{1}}y + [\partial_{1\dot{1}}x, y])g. \tag{35}$$

In terms of the matrix superfields x and y satisfying (29), (34), the equations of motion for the theory therefore take the remarkably simple form of Eq. (32a), (32b).

Introducing the symmetric products of pairs of derivatives (2),

$$\begin{aligned} D^{ij} &\equiv D^{\alpha i}D^j_\alpha = D^{ji}, \\ \overline{D}_{ij} &\equiv D_{\dot{\alpha}i}D^{\dot{\alpha}}_j = \overline{D}_{ji}, \end{aligned} \tag{36}$$

Eqs. (32a), (32b) may be obtained by varying the Lagrangian density

$$\begin{aligned} L = \mathrm{tr}\{&\overline{D}_{ij}(\tfrac{1}{2}D^i_1 x D^j_2 x + x D^i_1 x D^j_1 x) \\ &+ D^{ij}(\tfrac{1}{2}D_{\dot{1}i}y D_{\dot{2}j}y + y D_{\dot{1}i}y D_{\dot{1}j}y)\}. \end{aligned} \tag{37}$$

This Lagrangian can be considered to be an integral over a subspace of the odd part of superspace, since up to a total x derivative, a supercovariant derivative D^i_α is equivalent to an an ordinary spinor derivative $\partial/\partial\vartheta^{\alpha}_{i}$, which in turn is equivalent to a ϑ integration (since $\int d\vartheta\,\vartheta = 1$). The Lagrangian (37) is therefore a sum of integrals over a two-dimensional ϑ subspace and a two-dimensional $\bar{\vartheta}$ subspace. The functional (37) is rather reminiscent of the superinvariants constucted in Ref. 22 by integrating over even-dimensional submanifolds of extended superspace. The constraints (29), (34) on x and y may be incorporated using Lagrange multipliers.

III. A CONSTRUCTION OF CONSERVED CURRENTS

Equations (32) are the consistency conditions for the following set of equations:

$$N^{i}\psi \equiv (\mathrm{D}^{i}_{1} + \mu\mathrm{D}^{i}_{2} + \mu\mathrm{D}^{i}_{1}x)\psi = 0, \tag{38a}$$

$$M_{i}\psi \equiv (\mathrm{D}_{\dot{1}i} + \lambda\,\mathrm{D}_{\dot{2}i} + \lambda\,\mathrm{D}_{\dot{1}i}y)\psi = 0, \tag{38b}$$

$$\begin{aligned}Z\psi \equiv (&\partial_{1\dot{1}} + \lambda(\partial_{1\dot{2}} + \partial_{1\dot{1}}y) + \mu(\partial_{2\dot{1}} + \partial_{1\dot{1}}x)\\ &+ \lambda\mu(\partial_{2\dot{2}} + \partial_{1\dot{2}}x + \partial_{2\dot{1}}y + [\partial_{1\dot{1}}x,y]))\psi = 0,\end{aligned} \tag{38c}$$

where ψ is a matrix superfield in the gauge group depending on both parameters μ and λ. Consistency of the above equations is tantamount to requiring that (N^{i},M_{i},Z) satisfy the quantum-mechanical supersymmetry algebra,

$$\{N^{i},N^{j}\} = 0 = \{M_{i},M_{j}\}, \tag{39a}$$

$$\{N^{i},M_{j}\} = 2\delta^{i}_{j}Z, \tag{39b}$$

$$[N^{i},Z] = 0 = [M_{i},Z]\,. \tag{39c,d}$$

The brackets (39a) yield relations (32a), (32b) and the relation (39b) corresponds to Eqs. (29), (34). The system (38) may be obtained from (15) by multiplying the latter by g^{-1} on the right and g on the left, inserting the solution (27), and denoting by parameters μ and λ the ratios of the components of the spinors μ^{α} and $\lambda^{\dot{\alpha}}$, $\mu^{2}/\mu^{1} = \mu$, $\lambda^{\dot{2}}/\lambda^{\dot{1}} = \lambda$.

Now we introduce the generating function for superfields $x^{(n)},y^{(n)}$ $(n\geqslant 0, x^{(0)} = x, y^{(0)} = y)$, $\ln\psi$, where

$$\psi = \lim_{N\to\infty}\psi_{N};$$

$$\begin{aligned}\psi_{N} &= e^{-\lambda^{N}y^{(N-1)}}e^{-\mu^{N}x^{(N-1)}}e^{-\lambda^{N-1}y^{(N-2)}}\\ &\quad\times\cdots e^{-\mu^{2}x^{(1)}}e^{-\lambda y}e^{-\mu x}.\end{aligned} \tag{40}$$

Requiring that this ψ satisfies (38) immediately yields an infinite number of conserved supercurrents. From (38) we obtain

$$\psi(\mathrm{D}^{i}_{1+\mu 2})\psi^{-1} \equiv \psi(\mathrm{D}^{i}_{1} + \mu\mathrm{D}^{i}_{2})\psi^{-1} = \mu\mathrm{D}^{i}_{1}x, \tag{41a}$$

$$\psi(\mathrm{D}_{\dot{1}+\lambda\dot{2},i})\psi^{-1} \equiv \psi(\mathrm{D}_{\dot{1}i} + \lambda\,\mathrm{D}_{\dot{2}i})\psi^{-1} = \lambda\,\mathrm{D}_{\dot{1}i}y, \tag{41b}$$

$$\begin{aligned}&\psi(\partial_{1\dot{1}} + \mu\partial_{2\dot{1}} + \lambda\partial_{1\dot{2}} + \mu\lambda\partial_{2\dot{2}})\psi^{-1}\\ &\quad = \mu\partial_{1\dot{1}}x + \lambda\partial_{1\dot{1}}y + \mu\lambda(\partial_{1\dot{2}}x + \partial_{2\dot{1}}y + [\partial_{1\dot{1}}x,y]).\end{aligned} \tag{41c}$$

Equations (41a)–(41b) immediately yield

$$\mathrm{D}^{(i}_{1}(\psi\mathrm{D}^{j)}_{1+\mu 2}\,\psi^{-1}) = 0, \tag{42a}$$

$$\mathrm{D}_{\dot{1}(i}(\psi\mathrm{D}_{\dot{1}+\lambda\dot{2},j)}\psi^{-1}) = 0. \tag{42b}$$

Expanding (42) in μ,λ yields the Nth continuity equation as the trace [over the SU (3) indices] of the coefficient of μ^{N+1} and λ^{N+1} in (42a) and (42b), respectively. We consider

$$\psi_{3} = e^{-\lambda^{3}y^{(2)}}e^{-\mu^{3}x^{(2)}}e^{-\lambda^{2}y^{(1)}}e^{-\mu^{2}x^{(1)}}e^{-\lambda y}e^{-\mu x}.$$

Expanding $\psi_{3}\,\mathrm{D}^{i}_{1+\mu 2}\psi_{3}^{-1}$ up to order $\mu^{3}\lambda^{3}$ yields

$$\begin{aligned}&\mathrm{D}^{i}_{1+\mu 2} + \lambda\,\mathrm{D}^{i}_{1+\mu 2} + (\lambda^{2}/2)[\mathrm{D}^{i}_{1+\mu 2}y,y] + (\lambda^{3}/6)[[\mathrm{D}^{i}_{1+\mu 2}y,y],y] + \mu\mathrm{D}^{i}_{1+\mu 2}x + \lambda\mu[\mathrm{D}^{i}_{1+\mu 2}x,y]\\
&\quad + (\lambda^{2}\mu/2)[[\mathrm{D}^{i}_{1+\mu 2}x,y],y] + (\lambda^{3}\mu/6)[[[\mathrm{D}^{i}_{1+\mu 2}x,y],y],y] + (\mu^{2}/2)[\mathrm{D}^{i}_{1+\mu 2}x,x] + (\lambda\mu^{2}/2)[[\mathrm{D}^{i}_{1+\mu 2}x,x],y]\\
&\quad + (\lambda^{2}\mu^{2}/4)[[[\mathrm{D}^{i}_{1+\mu 2}x,x]y],y] + (\lambda^{3}\mu^{2}/12)[[[[\mathrm{D}^{i}_{1+\mu 2}x,x],y],y],y] + (\mu^{3}/6)[[\mathrm{D}^{i}_{1}x,x],x]\\
&\quad + (\lambda\mu^{3}/6)[[[\mathrm{D}^{i}_{1}x,x],x],y] + (\lambda^{2}\mu^{3}/12)[[[[\mathrm{D}^{i}_{1}x,x],x],y],y] + (\lambda^{3}\mu^{3}/36)[\ldots[\mathrm{D}^{i}_{1}x,x],x],y],y],y]\\
&\quad + \mu^{2}\mathrm{D}^{i}_{1+\mu 2}x^{(1)} + \lambda\mu^{2}[\mathrm{D}^{i}_{1+\mu 2}y,x^{(1)}] + (\lambda^{2}\mu^{2}/2)[[\mathrm{D}^{i}_{1+\mu 2}y,y],x^{(1)}] + (\lambda^{3}\mu^{2}/6)[[[\mathrm{D}^{i}_{1+\mu 2}y,y],y],x^{(1)}]\\
&\quad + \mu^{3}[\mathrm{D}^{i}_{1+\mu 2}x,x^{(1)}] + \lambda^{2}\mathrm{D}^{i}_{1+\mu 2}y^{(1)} + \lambda^{3}[\mathrm{D}^{i}_{1+\mu 2}y,y^{(1)}] + \mu\lambda^{2}[\mathrm{D}^{i}_{1+\mu 2}x,y^{(1)}] + \mu\lambda^{3}[[\mathrm{D}^{i}_{1+\mu 2}x,y],y^{(1)}]\\
&\quad + (\lambda^{2}\mu^{2}/2)[[\mathrm{D}^{i}_{1+\mu 2}x,x],y^{(1)}] + (\lambda^{3}\mu^{2}/2)[[[\mathrm{D}^{i}_{1+\mu 2}x,x],y],y^{(1)}] + (\lambda^{2}\mu^{3}/6)[[[\mathrm{D}^{i}_{1}x,x],x],y^{(1)}]\\
&\quad + (\lambda^{3}\mu^{3}/6)[[[[\mathrm{D}^{i}_{1}x,x],x],y],y^{(1)}] + \lambda^{2}\mu^{2}[\mathrm{D}^{i}_{1+\mu 2}x^{(1)},y^{(1)}] + \lambda^{3}\mu^{2}[[\mathrm{D}^{i}_{1+\mu 2}y,x^{(1)}],y^{(1)}]\\
&\quad + \mu^{3}\mathrm{D}^{i}_{1}x^{(2)} + \mu^{3}\lambda\,[\mathrm{D}^{i}_{1}y,x^{(2)}] + (\mu^{3}\lambda^{2}/2)[[\mathrm{D}^{i}_{1}y,y],x^{(2)}] + (\mu^{3}\lambda^{3}/6)[[[\mathrm{D}^{i}_{1}y,y],y],x^{(2)}]\\
&\quad + \mu^{3}\lambda^{2}[\mathrm{D}^{i}_{1}y^{(1)},x^{(2)}] + \mu^{3}\lambda^{3}[\mathrm{D}^{i}_{1}y,y^{(1)}] + \lambda^{3}\mathrm{D}^{i}_{1+\mu 2}y^{(2)} + \lambda^{3}\mu[\mathrm{D}^{i}_{1+\mu 2}x,y^{(2)}] + (\lambda^{3}\mu^{2}/2)[[\mathrm{D}^{i}_{1+\mu 2}x,x],y^{(2)}]\\
&\quad + (\lambda^{3}\mu^{3}/6)[[[\mathrm{D}^{i}_{1}x,x]x],y^{(2)}] + \lambda^{3}\mu^{2}[\mathrm{D}^{i}_{1+\mu 2}x^{(1)},y^{(2)}] + \mu^{3}\lambda^{3}[\mathrm{D}^{i}_{1}x^{(2)},y^{(2)}].\end{aligned} \tag{43a}$$

Similarly expanding $\psi_{3}\,\mathrm{D}_{\dot{1}+\lambda\dot{2},i}\psi_{3}^{-1}$ yields [we suppress the SU(3) index i in this formula]

$$\begin{aligned}&\mathrm{D}_{\dot{1}+\lambda\dot{2}} + \lambda\,\mathrm{D}_{\dot{1}+\lambda\dot{2}} + (\lambda^{2}/2)[\mathrm{D}_{\dot{1}+\lambda\dot{2}}y,y] + (\lambda^{3}/6)[[\mathrm{D}_{\dot{1}+\lambda\dot{2}}y,y],y] + \mu\mathrm{D}_{\dot{1}+\lambda\dot{2}}x + \mu\lambda\,[\mathrm{D}_{\dot{1}+\lambda\dot{2}}x,y]\\
&\quad + (\mu\lambda^{2}/2)[[\mathrm{D}_{\dot{1}+\lambda\dot{2}}x,y],y] + (\mu\lambda^{3}/6)[[[\mathrm{D}_{\dot{1}+\lambda\dot{2}}x,y],y],y] + (\mu^{2}/2)[\mathrm{D}_{\dot{1}+\lambda\dot{2}}x,x] + (\lambda\mu^{2}/2)[[\mathrm{D}_{\dot{1}+\lambda\dot{2}}x,x],y]\\
&\quad + (\lambda^{2}\mu^{2}/4)[[[\mathrm{D}_{\dot{1}+\lambda\dot{2}}x,x],y],y] + (\lambda^{3}\mu^{2}/12)[[[[\mathrm{D}_{\dot{1}+\lambda\dot{2}}x,x],y],y],y] + (\mu^{3}/6)[[\mathrm{D}_{\dot{1}+\lambda\dot{2}}x,x],x]\\
&\quad + (\lambda\mu^{3}/6)[[[\mathrm{D}_{\dot{1}+\lambda\dot{2}}x,x],x],y] + (\lambda^{2}\mu^{3}/12)[[[[\mathrm{D}_{\dot{1}+\lambda\dot{2}}x,x],x],y],y] + (\lambda^{3}\mu^{3}/36)[[[[[\mathrm{D}_{\dot{1}+\lambda\dot{2}}x,x]x]y]y]y]\\
&\quad + \mu^{2}\mathrm{D}_{\dot{1}+\lambda\dot{2}}x^{(1)} + \mu^{2}\lambda\,[\mathrm{D}_{\dot{1}+\lambda\dot{2}}y,x^{(1)}] + (\mu^{2}\lambda^{2}/2)[[\mathrm{D}_{\dot{1}+\lambda\dot{2}}y,y],x^{(1)}] + (\mu^{2}\lambda^{3}/6)[[[\mathrm{D}_{\dot{1}+\lambda\dot{2}}y,y],y],x^{(1)}]\\
&\quad + \mu^{3}[\mathrm{D}_{\dot{1}+\lambda\dot{2}}x,x^{(1)}] + \lambda\mu^{3}[[\mathrm{D}_{\dot{1}+\lambda\dot{2}}x,y],x^{(1)}] + (\mu^{3}\lambda^{2}/2)[[[\mathrm{D}_{\dot{1}+\lambda\dot{2}}x,y],y],x^{(1)}]\\
&\quad + (\mu^{3}\lambda^{3}/6)[[[[\mathrm{D}_{\dot{1}+\lambda\dot{2}}x,y],y],y],x^{(1)}] + \lambda^{2}\mathrm{D}_{\dot{1}+\lambda\dot{2}}y^{(1)} + \lambda^{3}[\mathrm{D}_{\dot{1}}y,y^{(1)}] + \lambda^{2}\mu[\mathrm{D}_{\dot{1}+\lambda\dot{2}}x,y^{(1)}]\end{aligned}$$

$$+\lambda^3\mu[[D_{\mathrm{i}+\lambda\dot{2}}x,y],y^{(1)}]+(\mu^2\lambda^2/2)[[D_{\mathrm{i}+\lambda\dot{2}}x,x],y^{(1)}]+(\mu^2\lambda^3/2)[[[D_{\mathrm{i}+\lambda\dot{2}}x,x],y],y^{(1)}]$$
$$+(\mu^3\lambda^2/6)[[[D_{\mathrm{i}+\lambda\dot{2}}x,x],x],y^{(1)}]+(\mu^3\lambda^3/6)[[[[D_{\mathrm{i}+\lambda\dot{2}}x,x],x],y],y^{(1)}]+\lambda^2\mu^2[D_{\mathrm{i}+\lambda\dot{2}}x^{(1)},y^{(1)}]$$
$$+\mu^2\lambda^3[[D_{\mathrm{i}+\lambda\dot{2}}y,x^{(1)}],y^{(1)}]+\lambda^2\mu^3[[D_{\mathrm{i}+\lambda\dot{2}}x,x^{(1)}],y^{(1)}]+\lambda^3\mu^3[[[D_{\mathrm{i}}x,y]x^{(1)}],y^{(1)}]+\mu^3D_{\mathrm{i}+\lambda\dot{2}}x^{(2)}$$
$$+\mu^3\lambda[D_{\mathrm{i}+\lambda\dot{2}}y,x^{(2)}]+(\mu^3\lambda^2/2)[[D_{\mathrm{i}+\lambda\dot{2}}y,y],x^{(2)}]+(\mu^3\lambda^3/6)[[[D_{\mathrm{i}}y,y],y],x^{(2)}]$$
$$+\mu^3\lambda^2[D_{\mathrm{i}+\lambda\dot{2}}y^{(1)},x^{(2)}]+\mu^3\lambda^3[[D_{\mathrm{i}}y,y^{(1)}],x^{(2)}]+\lambda^3D_{\mathrm{i}+\lambda\dot{2}}y^{(2)}+\lambda^3\mu[D_{\mathrm{i}}x,y^{(2)}]$$
$$+(\lambda^3\mu^2/2)[[D_{\mathrm{i}}x,x],y^{(2)}]+(\mu^3\lambda^3/6)[[[D_{\mathrm{i}}x,x],x],y^{(2)}]+\lambda^3\mu^2[D_{\mathrm{i}}x^{(1)},y^{(2)}]$$
$$+\mu^3\lambda^3[[D_{\mathrm{i}}x,x^{(1)}],y^{(2)}]+\lambda^3\mu^3[D_{\mathrm{i}}x^{(2)},y^{(2)}] \tag{43b}$$

and

$$\psi_2(\partial_{1\mathrm{i}}+\mu\partial_{2\mathrm{i}}+\lambda\partial_{1\dot{2}}+\mu\lambda\partial_{2\dot{2}})\psi_2^{-1}$$
$$=\mu\partial_{1\mathrm{i}}x+\lambda\partial_{1\mathrm{i}}y+\mu\lambda(\partial_{1\dot{2}}x+\partial_{2\mathrm{i}}y+[\partial_{1\mathrm{i}}x,y])$$
$$+\mu^2(\partial_{2\mathrm{i}}x+\tfrac{1}{2}[\partial_{1\mathrm{i}}x,x]+\partial_{1\mathrm{i}}x^{(1)})+\lambda^2(\partial_{1\dot{2}}y+\tfrac{1}{2}[\partial_{1\mathrm{i}}y,y]+\partial_{1\mathrm{i}}y^{(1)})+\mu\lambda^2(\partial_{2\dot{2}}y+\tfrac{1}{2}[\partial_{2\mathrm{i}}y,y]+[\partial_{1\dot{2}}x,y]$$
$$+\tfrac{1}{2}[[\partial_{1\mathrm{i}}x,y],y]+\partial_{2\mathrm{i}}y^{(1)}+[\partial_{1\mathrm{i}}x,y^{(1)}])+\lambda\mu^2(\partial_{2\dot{2}}x+\tfrac{1}{2}[\partial_{1\dot{2}}x,x]+[\partial_{2\mathrm{i}}x,y]$$
$$+\tfrac{1}{2}[[\partial_{1\mathrm{i}}x,x],y]+\partial_{2\mathrm{i}}x^{(1)}+[\partial_{1\mathrm{i}}y,x^{(1)}])+O(\mu^2\lambda^2). \tag{43c}$$

Therefore, Eqs. (41) contain Eqs. (29), (34),

$$D_{\alpha j}x=0, \quad (29a), (34a),$$
$$D^i_{\dot{1}}y=0, \quad (29b),$$
$$D^i_{\dot{2}}y+[D^i_{\dot{1}}x,y]=0 \quad , \quad (34b),$$

as well as the relations

$$D^i_{\dot{1}}x^{(1)}+D^i_{\dot{2}}x+\tfrac{1}{2}[D^i_{\dot{1}}x,x]=0, \tag{44a}$$
$$D^i_{\dot{1}}x^{(2)}+D^i_{\dot{2}}x^{(1)}+[D^i_{\dot{1}}x,x^{(1)}]$$
$$+\tfrac{1}{2}[D^i_{\dot{2}}x,x]+\tfrac{1}{6}[[D^i_{\dot{1}}x,x],x]=0, \tag{44b}$$
$$D^i_{\dot{1}}y^{(n)}=0, \quad n\geqslant 0, \quad y^{(0)}=y, \tag{44c}$$
$$D^i_{\dot{2}}y^{(n)}+[D^i_{\dot{1}}x,y^{(n)}]=0, \quad n\geqslant 0, \tag{44d}$$
$$D^i_{\dot{1}}[x^{(1)},y]=0, \tag{44e}$$

as coefficients of μ^2, μ^3, λ^{n+1}, $\mu\lambda^{n+1}$, $\mu^2\lambda$, respectively, in (41a), and

$$D_{1i}y^{(1)}+D_{\dot{2}i}y+\tfrac{1}{2}[D_{1i}y,y]=0, \tag{45a}$$
$$D_{1i}y^{(2)}+D_{\dot{2}i}y^{(1)}+[D_{1i}y,y^{(1)}]$$
$$+\tfrac{1}{2}[D_{\dot{2}i}y,y]+\tfrac{1}{6}[[D_{1i}y,y],y]=0, \tag{45b}$$
$$D_{1i}x^{(n)}=0, \quad n\geqslant 0, \quad x^{(0)}=x, \tag{45c}$$
$$D_{\dot{2}i}x^{(n)}+[D_{1i}y,x^{(n)}]=0, \quad n\geqslant 0, \tag{45d}$$

as coefficients of λ^2, λ^3, μ^{n+1}, $\mu^{n+1}\lambda$, respectively, in (41b). Equation (41c) yields the relations

$$\partial_{2\mathrm{i}}x+\tfrac{1}{2}[\partial_{1\mathrm{i}}x,x]+\partial_{1\mathrm{i}}x^{(1)}=0, \tag{46a}$$
$$\partial_{2\dot{2}}x+\tfrac{1}{2}[\partial_{1\dot{2}}x,x]+\partial_{1\dot{2}}x^{(1)}+\partial_{1\mathrm{i}}[y,x^{(1)}]=0, \tag{46b}$$

which together with (29), (34), (44a), and (44e) may be used to verify the integrability condition (39c), as well as the relations

$$\partial_{1\dot{2}}y+\tfrac{1}{2}[\partial_{1\mathrm{i}}y,y]+\partial_{1\mathrm{i}}y^{(1)}=0, \tag{46a$'$}$$
$$\partial_{2\dot{2}}y+\tfrac{1}{2}[\partial_{2\mathrm{i}}y,y]+\partial_{2\mathrm{i}}y^{(1)}+[\partial_{1\dot{2}}x,y]$$
$$+\tfrac{1}{2}[[\partial_{1\mathrm{i}}x,y],y]+[\partial_{1\mathrm{i}}x,y^{(1)}]=0, \tag{46b$'$}$$

which together with (29), (34), and (45a) may similarly be used to verify the integrability condition (39d).

Now defining

$$J^{(n)i}_{\dot{2}}=-D^i_{\dot{1}}x^{(n)}, \tag{47a}$$
$$J^{(n)}_{2i}=-D_{1i}y^{(n)}, \tag{47b}$$

we clearly have

$$D^{(i}_{\dot{1}}J^{(n)j)}_{\dot{2}}=0=D_{1(i}J^{(n)}_{2j)}, \tag{48}$$

and, tracing over the SU (3) indices, yields infinitely many solutions of the conservation equation

$$D^i_{\dot{1}}J_{2i}-D_{1i}J^i_{\dot{2}}=0. \tag{49}$$

Thus the coefficient of μ^{N+1} (λ^{N+1}) in (41a) [(41b)] yields an expression for $J^{(N)i}_{\dot{2}}$ ($J^{(N)}_{2i}$) in terms of $x^{(n)}$ ($y^{(n)}$), $n<N$, and its derivatives. Therefore, given superfields x,y solving Eqs. (32), (29), (34), we need to perform no further integrations in order to successively construct all the $J^{(n)i}_{\dot{2}}$'s and $J^{(n)}_{2i}$'s explicitly just by taking spinorial derivatives and commutators. From (44), (45), the components of the first two currents, for instance, are

$$J^{(1)i}_{\dot{2}}=D^i_{\dot{2}}x+\tfrac{1}{2}[D^i_{\dot{1}}x,x], \tag{50a}$$
$$J^{(2)i}_{\dot{2}}=D^i_{\dot{2}}x^{(1)}+[D^i_{\dot{1}}x,x^{(1)}]$$
$$+\tfrac{1}{2}[D^i_{\dot{2}}x,x]+\tfrac{1}{6}[[D^i_{\dot{1}}x,x]x], \tag{50b}$$
$$J^{(1)}_{2i}=D_{\dot{2}i}x+\tfrac{1}{2}[D_{1i}y,y], \tag{50c}$$
$$J^{(2)}_{2i}=D_{\dot{2}i}y^{(1)}+[D_{1i}y,y^{(1)}]$$
$$+\tfrac{1}{2}[D_{\dot{2}i}y,y]+\tfrac{1}{6}[[D_{1i}y,y]y]. \tag{50d}$$

IV. INFINITESIMAL SYMMETRY TRANSFORMATIONS

Analogously to the infinitesimal symmetry transformations obtained previously in Ref. 15, the equations of motion (32), (29), (34) are left invariant, to first order in the variation, under the transformations

$$\delta x=-\mu^{-1}(\psi T\psi^{-1})+\xi, \tag{51a}$$
$$\delta y=-\lambda^{-1}(\psi T\psi^{-1})+\eta, \tag{51b}$$

where ψ satisfies (38), and T is a matrix in the gauge algebra satisfying

$$(\mathrm{D}_1^i + \mu \mathrm{D}_2^i)T = 0, \tag{52a}$$

$$(\mathrm{D}_{\dot 1 i} + \lambda\, \mathrm{D}_{\dot 2 i})T = 0, \tag{52b}$$

$$(\partial_{1\dot 1} + \lambda\partial_{1\dot 2} + \mu\partial_{2\dot 1} + \lambda\mu\partial_{2\dot 2})T = 0, \tag{52c}$$

and the matrix superfields ξ and η are defined to be solutions of the following equations:

$$\begin{aligned} &\mathrm{D}_1^i\ \xi = 0 = \mathrm{D}_{\dot 1 i}\eta, \\ &\mathrm{D}_{\dot\alpha j}\xi = \mu^{-1}\mathrm{D}_{\dot\alpha j}S, \\ &\mathrm{D}_1^i\ \eta = \lambda^{-1}\,\mathrm{D}_1^i\ S, \\ &\mathrm{D}_2^i\ \eta = \mu^{-1}[\mathrm{D}_1^i\ S,y] - \lambda^{-1}\,\mathrm{D}_1^i\ S - [\mathrm{D}_1^i\ x,\eta], \end{aligned} \tag{53}$$

where

$$S \equiv \psi T \psi^{-1}. \tag{54}$$

It is clear that (29), (34) are left invariant under the variations (51), whereas the invariance of Eq. (32a) requires the vanishing of

$$\mathrm{D}_2^{(i}\mathrm{D}_1^{j)}\, S + \{\mathrm{D}_1^{(i}x, \mathrm{D}_1^{j)}\, S\}. \tag{55}$$

Taking the second term in (55)

$$\begin{aligned} \{\mathrm{D}_1^{(i}x,\, \mathrm{D}_1^{j)}\, S\} &= -\mathrm{D}_1^{(i}[\mathrm{D}_1^{j)}\, x,S] \\ &= \mu^{-1}\,\mathrm{D}_1^{(i}\mathrm{D}_1^{j)}\, S + \mathrm{D}_1^{(i}\mathrm{D}_2^{j)}\, S, \end{aligned}$$

using (38a).

Thus (55) vanishes. The invariance of (32b) may similarly be verified using Eq. (38b).

The algebra of these infinitesimal transformations as well as their integrability to finite (Bäcklund) transformations will be discussed elsewhere.

V. SUPERSYMMETRIC SELF-DUALITY CONDITIONS

The supersymmetric self-duality equations arise as integrability conditions if, in addition to (38), the matrix superfield ψ is required to be μ independent,

$$\frac{\partial}{\partial\mu}\psi = 0. \tag{56}$$

This is tantamount to all the $x^{(n)}$'s, $n \geqslant 0$, in (40) being set to zero. Then Eqs. (38) reduce to

$$\mathrm{D}_1^i\ \psi = 0 = (\mathrm{D}_2^i + \mathrm{D}_1^i x)\psi, \tag{57a}$$

$$(\mathrm{D}_{\dot 1 i} + \lambda\,\mathrm{D}_{\dot 2 i} + \lambda\,\mathrm{D}_{\dot 1 i}y)\psi = 0, \tag{57b}$$

$$[\partial_{1\dot 1} + \lambda(\partial_{1\dot 2} + \partial_{1\dot 1}y)]\psi = 0, \tag{57c}$$

$$[\partial_{2\dot 1} + \lambda(\partial_{2\dot 2} + \partial_{2\dot 1}y)]\psi = 0, \tag{57d}$$

a supersymmetrization of the linear system (57c), (57d) for the form of the self-duality equation

$$y_{y\bar y} + y_{z\bar z} = [y_y, y_z], \tag{58}$$

[where we use Yang's variables $x_{\alpha\beta} = \left[\begin{smallmatrix} y & -\bar z \\ z & \bar y \end{smallmatrix}\right]$, used by Leznov and Saveliev[20] in their reformulation of the conservation laws for which the generating function $\psi|_{\mu=0}$ in (40) is identical to that found in Ref. 23. Similarly, imposing λ independence on ψ yields the anti-self-duality equations.

ACKNOWLEDGMENT

I would like to thank Alan Chodos and David Fairlie for discussions on Ref. 20.

[1]F. Gliozzi, J. Scherk, and D. Olive, Nucl. Phys. B **122**, 253 (1977).
[2]L. Brink, J. Schwarz, and J. Scherk, Nucl. Phys. B **121**, 77 (1977).
[3]S. Mandelstam, Nucl. Phys. B **213**, 149 (1983); L. Brink, O. Lindgren, and B. Nilsson, *ibid.* **212**, 401 (1983); Phys. Lett. B **123**, 323 (1983).
[4]P. Howe, K. Stelle, and P. Townsend, Nucl. Phys. B **214**, 519 (1983); **236**, 123 (1984).
[5]M. Sohnius and P. C. West, Phys. Lett. B **100**, 245 (1981).
[6]P. Goddard, J. Nuyts, and D. Olive, Nucl. Phys. B **125**, 1 (1977); C. Montonen and D. Olive, Phys. Lett. B **72**, 117 (1977).
[7]S. Coleman, Phys. Rev. D **11**, 2088 (1975).
[8]S. Coleman and G. Mandula, Phys. Rev. **159**, 1251 (1967).
[9]R. Grimm, M. Sohnius, and J. Wess, Nucl. Phys. B **133**, 275 (1978); M. Sohnius, *ibid.* **136**, 461 (1978).
[10]R. S. Ward, Phys. Lett. A **61** 31 (1977); M. F. Atiyah and R. S. Ward, Commun. Math. Phys. **55** 117 (1977).
[11]E. Witten, Phys. Lett. B **77** 394 (1978).
[12]Yu. I. Manin, J. Sov. Math. **2**, 465 (1983).
[13]Yu. I. Manin, *Gauge Field Theory and Complex Geometry* (Grundlehren der mathematischen Wissenschaften, Vol. 289) (Springer, Berlin, 1988).
[14]I. V. Volovich, Lett. Math. Phys. **7**, 517 (1983); Phys. Lett. B **129**, 429 (1983); I. Ya. Aref'eva and I. V. Volovič, Lett. Math. Phys. **9**, 231 (1985).
[15]C. Devchand, Nucl. Phys. B **238**, 333 (1984); in *Field Theory, Quantum Gravity and Strings*, edited by H. de Vega and N. Sanchez, *Lecture Notes in Physics*, Vol. 246 (Springer, Berlin, 1986).
[16]L.-L. Chau, Ge Mo-Lin, and Z. Popowicz, Phys. Rev. Lett. **52**, 1940 (1984); L.-L. Chau, *Proceedings of the Workshop on Vertex Operators in Mathematics and Physics*, Berkeley, 1983, edited by J. Lepowsky, S. Mandelstam, and I. M. Singer (Springer, Berlin, 1984).
[17]J. Harnad and M. Jacques, J. Math. Phys. **27**, 2394 (1986).
[18]J. Tafel, J. Math. Phys. **28** 240 (1987).
[19]N. Suzuki, Commun. Math. Phys. **113**, 155 (1987).
[20]A. N. Leznov and M. V. Saveliev, Serpukhov preprint, IHEP 87-195; A. N. Leznov, Theor. Math. Fiz. **73**, 302 (1987).
[21]J. Harnad, J. Hurtubise, M. Legare, and S. Shnider, Nucl. Phys. B **256**, 609 (1985).
[22]P. Howe, K. Stelle, and P. Townsend, Nucl. Phys. B **191**, 445 (1981).
[23]A. Chodos and D. B. Fairlie, unpublished.

Vertex operator representations of basic affine superalgebras

L. Frappat[a)]

Laboratoire d'Annecy-6-Vieux de Physique des Particules, B.P. 909, 74019 Annecy-le-Vieux, France

A. Sciarrino

Departimento di Scienze Fisiche, 80125 Napoli, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Napoli, Italy

P. Sorba

Laboratoire d'Annecy-6-Vieux de Physique des Particules, B.P. 909, 74019 Annecy-le-Vieux, France

(Received 12 May 1989; accepted for publication 9 August 1989)

Using the basic construction of the affine superalgebra $\mathrm{OSp}(M|N)^{(1)}$ in terms of vertex operators, vertex operator representations of a large class of untwisted and twisted affine basic superalgebras of the unitary and orthosymplectic series are found. The construction uses two methods: first, the regular embeddings of semisimple superalgebras (untwisted case) and, second, the folding method developed in a previous paper [Frappat *et al.*, Nucl. Phys. B **305**, 164 (1988)] to construct twisted vertex operators for Lie algebras with fermionic elementary fields (twisted case).

I. INTRODUCTION

The theory of superalgebras today constitutes a natural extension of the known algebraic structures with supersymmetry appearing as a privileged component in physics. Vertex operators of affine simple Lie algebras play a fundamental role in string theories and more generally in two-dimensional conformal field theories via the boson–fermion equivalence.[1]

Recently, a current algebra treatment of symplectic bosons, which arise in constructing superconformal ghosts of fermionic string theories, naturally led the authors of Ref. 2 to use affine simple superalgebras. This was in some sense a first step in the study of conformal field theories based on affine superalgebras. Therefore one may hope that a construction of vertex operators for affine superalgebras, which will involve the boson–boson equivalence, could help us to go further in the study of conformal theories. Moreover such a framework also may be useful in the construction of extended superconformal theories.

In a previous publication, the realization of the basic orthosymplectic affine superalgebra $\mathrm{OSp}(M|N)^{(1)}$ via vertex operators has been considered by one of us.[3] The first fundamental ingredient in this construction is the use, following Ref. 4, of an affine Clifford algebra to obtain the $k=1$ level of $\mathrm{SO}\,(M)^{(1)}$ and of an affine Weyl algebra to generate the $k=-1$ level of $\mathrm{SP}(N)^{(1)}$, both showing up in the $k=1$ representation of $\mathrm{OSp}(M|N)^{(1)}$. The second step consists in expressing such fermionic (resp. bosonic) elementary fields in terms of vertex operators, such a program being achieved owing to the fermion (resp. symplectic boson) bosonization. A summary of this method is given in Sec. III A.

The generalization of this vertex construction to all the untwisted as well as twisted affine basic superalgebras for unitary and orthosymplectic series has been made possible owing to general structure properties of affine superalgebras we have developed in Ref. 5. First, an extensive study of the extended Dynkin diagrams (EDD's) associated to a finite simple superalgebra has led us to deduce the regular subsuperalgebras of a basic superalgebra. Second, the use of the symmetries of the affine Dynkin diagrams allows us to determine by folding the twisted affine superalgebras. It is this kind of technique, i.e., embedding and folding, that is used to achieve the vertex construction for classical affine superalgebras.

II. A REMINDER ABOUT SUPERALGEBRAS [5,6]

Let us recall rapidly the Z_2-grading structure and the root systems of the unitary and orthosymplectic superalgebras: we will denote by $\mathscr{G}_0$ (resp. $\mathscr{G}_1$) the bosonic (resp. fermionic) part of the superalgebra $\mathscr{G}$, and by Δ_0 (resp. Δ_1) the set of even (resp. odd) roots. The roots can be expressed in terms of $e_1,\ldots,e_m,f_1,\ldots,f_n$.

For the unitary series $A(m-1,n-1)$ or $\mathrm{Sl}\,(m|n)$,

$$\mathscr{G}_0=\mathrm{Sl}(m)\times\mathrm{Sl}(n)\times\mathrm{U}(1),$$

$$\mathscr{G}_1=(m,\bar{n})+(\bar{m},n)\,, \tag{2.1}$$

$$\Delta_0=\{e_i-e_j;f_i-f_j\},\quad \Delta_1=\{\pm(e_i-f_j)\}\,. \tag{2.2}$$

In the case $m=n$, the bosonic part reduces to $\mathrm{Sl}(m)\times\mathrm{Sl}(m)$.

For $B(m,n)$ or $\mathrm{OSp}(2m+1/2n)$ with $m\neq 0$,

$$\mathscr{G}_0=\mathrm{O}(2m+1)\times\mathrm{Sp}(2n),\quad \mathscr{G}_1=(2m+1,2n)\,, \tag{2.3}$$

$$\Delta_0=\{\pm e_i\pm e_j;\pm e_i;\pm f_i\pm f_j;\pm 2f_i\}\quad (i\neq j),$$

$$\Delta_1=\{\pm f_i;\pm e_i\pm f_j\}\,. \tag{2.4}$$

For $B(0,n)$ or $\mathrm{OSp}(1|2n)$,

$$\mathscr{G}_0=\mathrm{Sp}(2n),\quad \mathscr{G}_1=(2n)\,, \tag{2.5}$$

$$\Delta_0=\{\pm f_i\pm f_j;\pm 2f_i\}\quad (i\neq j),\quad \Delta_1=\{\pm f_i\}\,. \tag{2.6}$$

For $D(m,n)$ or $\mathrm{OSp}(2m|2n)$ with $m\neq 1$,

$$\mathscr{G}_0=\mathrm{O}(2m)\times\mathrm{Sp}(2n),\quad \mathscr{G}_1=(2m,2n)\,, \tag{2.7}$$

[a)] Also at Université de Savoie, 74000 Annecy, France.

$$\Delta_0 = \{\pm e_i \pm e_j; \pm f_i \pm f_j; \pm 2f_i\} \quad (i \neq j),$$
$$\Delta_1 = \{\pm e_i \pm f_j\}. \tag{2.8}$$

For $C(n+1)$ or $\mathrm{OSp}(2|2n)$,

$$\mathscr{G}_0 = \mathrm{O}(2)\times \mathrm{Sp}(2n), \quad \mathscr{G}_1 = (2n) + (2n), \tag{2.9}$$
$$\Delta_0 = \{\pm f_i \pm f_j; \pm 2f_i\} \quad (i \neq j), \quad \Delta_1 = \{\pm e \pm f_i\}. \tag{2.10}$$

III. VERTEX OPERATORS FOR THE UNTWISTED AFFINE SUPERALGEBRA

In a previous paper,[3] we constructed a level 1 representation in terms of vertex operators of the affine superalgebra $\mathrm{OSp}(M|N)^{(1)}$, using fermionic fields and the bosonic ghost fields from string theories. Let us remind the readers of the basic results about this construction, in particular, the form of the vertex operators.

A. Vertex operator representation of affine untwisted orthosymplectic superalgebra OSp($M/2n$) with $M>2$

In order to construct a vertex operator representation for $\mathrm{OSp}(2m|2n)^{(1)}$ one starts with m Fubini–Veneziano fields

$$Q^i(z) = q^i - ip^i \ln z + i \sum_{m\neq 0} \frac{\alpha^i_m}{m} z^{-m}, \tag{3.1}$$

with

$$[\alpha^i_m, \alpha^j_m] = m\delta_{m+n}\delta_{ij}, \tag{3.2}$$
$$[q^i, p^j] = i\delta_{ij} \quad (\alpha^i_o = p^i). \tag{3.3}$$

If e_i is a weight vector of the tensor coset $\mathbb{Z}^m$ of the weight lattice Λ_w of $\mathrm{SO}(2m)$, one defines the vertex operator as

$$U(\pm e_i, z) = z^{1/2} :\exp \pm i e_i Q^i(z):. \tag{3.4}$$

Notice that the elementary fields $E(\pm e_i,z) = U(\pm e_i,z) c_{\pm e_i}$, the cocycle operator $c_{\pm e_i}$ being defined on the lattice Λ_w, are fermionic fields. The Cartan subalegbra generators

$$P^i(z) = iz\,\partial Q^i(z) = \alpha^i_o + \sum_{m\neq 0} \alpha^i_m z^{-m} \tag{3.5}$$

and the step operators associated to the roots $\pm e_i \pm e_j$ with $1 < i \neq j \leqslant m$,

$$E(\pm e_i \pm e_j, z) = :U(\pm e_i,z)c_{\pm e_i} U(\pm e_j,z) c_{\pm e_j}:$$
$$= U(\pm e_i \pm e_j, z) c_{\pm e_i \pm e_j}, \tag{3.6}$$

generate a level 1 representation of the $\mathrm{SO}(2m)^{(1)}$ algebra. For the symplectic part $\mathrm{Sp}(2n)$, one has to consider n pairs of Fubini–Veneziano fields

$$\phi^i(z) = j^i_q - j^i_o \ln z + \sum_{m\neq 0} \frac{j^i_m}{m} z^{-m}, \tag{3.7}$$

$$\chi^i(z) = h^i_q - h^i_o \ln z + \sum_{m\neq 0} \frac{h^i_m}{m} z^{-m}, \tag{3.8}$$

such that

$$[j^i_m, j^j_n] = -m\delta_{m+n}\delta_{ij}, \tag{3.9}$$
$$[h^i_m, h^j_m] = m\delta_{m+n}\delta_{ij}. \tag{3.10}$$

We define the vertex operators as

$$U(f_i,z) = z^{1/2} :\exp \phi^i(z): :\exp -\chi^i(z):, \tag{3.11}$$
$$U(-f_i,z) = z^{1/2} :\exp -\phi^i(z): \partial :\exp \chi^i(z): \tag{3.12}$$

Now these elementary vertex operators behave rather as bosonic fields, but with a fermionic propagator. The Cartan subalgebra generators

$$P^i(z) = -z\,\partial\phi^i(z) = j^i_o + \sum_{m\neq 0} j^i_m z^{-m} \tag{3.13}$$

and the step operators associated to the roots $\pm f_i \pm f_j$, with $1 \leqslant i,j \leqslant m$,

$$E(\pm f_i \pm f_j, z) = :U(\pm f_i,z) U(\pm f_j,z): \tag{3.14}$$

generate a level -1 representation of the $\mathrm{Sp}(2n)^{(1)}$ algebra. The fermionic generators of $\mathrm{OSp}(2m|2n)^{(1)}$ are constructed as

$$E(\pm e_i \pm f_j, z) = U(\pm e_i,z) c_{\pm e_i} U(\pm f_j,z). \tag{3.15}$$

In the case of $\mathrm{OSp}(2m+1|2n)^{(1)}$, one has to construct the vertex operators associated to the (bosonic) roots $\pm e_i$ of $\mathrm{SO}(2m+1)^{(1)}$, one needs a supplementary fermionic field $\Gamma(z)$ such that

$$E(\pm e_i, z) = U(\pm e_i,z) c_{\pm e_i} \Gamma(z). \tag{3.16}$$

The step operators associated to the fermionic roots $\pm f_i$ are defined by

$$E(\pm f_j, z) = U(\pm f_j,z)\Gamma(z). \tag{3.17}$$

One can notice that the explicit expression of this fermionic auxiliary field in terms of vertex operators can be obtained by folding the DD (or EDD) of $\mathrm{OSp}(2m+2|2n)$. One obtains

$$\Gamma(z) = (1/\sqrt{2})\big(U(e_{2m+2},z)c_{e_{2m+2}} + U(-e_{2m+2},z)c_{-e_{2m+2}}\big). \tag{3.18}$$

B. "Bosonic" representation of $\mathrm{OSp}(2|2n)^{(1)}$ and $\mathrm{OSp}(1|2n)^{(1)}$

The cases of $\mathrm{OSp}(2|2n)^{(1)}$ and $\mathrm{OSp}(1|2n)^{(1)}$ require special attention because of the nature of their bosonic part.

In the case of $\mathrm{OSp}(2|2n)^{(1)}$, the bosonic part is $\mathrm{SO}(2)\times\mathrm{Sp}(2n)$ and the fermionic part reduces to twice the fundamental representation of $\mathrm{Sp}(2n)$ of dimension $2n$. The generators are constructed following Eqs. (3.13)–(3.15), i.e., for the bosonic roots $\pm f_i \pm f_j$ and $\pm 2f_i$ of $\mathrm{Sp}(2n)$,

$$E(\pm f_i \pm f_j, z) = :U(\pm f_i,z) U(\pm f_j,z):;$$

and, for the fermionic roots $\pm f_i \pm e$,

$$E(\pm f_i \pm e, z) = U(\pm f_i,z) U_\pm(z),$$

where $U_\pm(z)$ is a fermionic field constructed from a Fubini–Veneziano field $Q(z)$ as Eqs. (3.1) and (3.4):

$$U_\pm(z) = z^{1/2} :\exp \pm Q(z):.$$

The SO(2) generator is simply given by the $U(1)$ current associated to the vertex operators $U_\pm(z)$, i.e., $P(z) = iz\,\partial Q(z)$.

In the case of $\mathrm{OSp}(1|2n)$, there is no orthogonal part for the bosonic underlying algebra. The generators are constructed following Eqs. (3.14) and (3.17), i.e., for the fermionic roots $\pm f_i$ [representation $(2n)$ of $\mathrm{Sp}(2n)$],

$$E(\pm f_i,z) = U(\pm f_i,z)\Gamma(z)\,, \tag{3.19}$$

where $\Gamma(z)$ is an auxiliary fermionic field, given by [from Eq. (3.18)]

$$\Gamma(z) = (1/\sqrt{2})(U_+(z) + U_-(z))\,. \tag{3.20}$$

Notice that in this case the central charge in the bosonic part $\mathrm{Sp}(2n)$ of the algebra has the value -1, so that the vertex operator representation is a nonstandard representation of level -1.

C. "Fermionic" representation of $\mathrm{OSp}(2|2n)^{(1)}$ and $\mathrm{OSp}(1|2n)^{(1)}$

One first constructs, a level 1 representation of the bosonic part $\mathrm{Sp}(2n)$ in the usual way, using auxiliary fields. The vertex operators associated to the long roots $\pm 2e_i$ are

$$E(\pm 2e_i,z) = U(\pm 2e_i,z)c_{\pm 2e_i}\,. \tag{3.21}$$

The vertex operators associated to the long roots $\pm e_i \pm e_j$ are

$$E(\pm e_i \pm e_j,z) = U(\pm e_i \pm e_j,z)c_{\pm e_i \pm e_j}\Gamma_{ij}(z), \tag{3.22}$$

with

$$\Gamma_{ij}(z) = (1/\sqrt{2})(U(\xi_i - \xi_j,z)c_{\xi_i-\xi_j} + U(\xi_j - \xi_i,z)c_{\xi_j-\xi_i})\,. \tag{3.23}$$

The vectors ξ_i are orthogonal to each other ($\xi_i \cdot \xi_j = 0$) and also orthogonal to the e_i's ($e_i \cdot \xi_i = 0$). The vectors e_i and ξ_i are normalized to $e_i^2 = \xi_i^2 = \frac{1}{2}$. The auxiliary fields $\Gamma_{ij}(z)$ depend only on the orbit $\Omega = \{\pm e_j \pm e_j\}$ to which the short root belongs. The cocycle operators $c_{\pm e_i \pm e_j}$ are constructed on the root lattice of $\mathrm{Sp}(2n)$. The ξ_i's constitute a basis of the rescaled root lattice $(1/\sqrt{2})/\Lambda_R(\mathrm{SU}(2n))$, on which the cocycle operators $c_{\xi_i-\xi_j}$ can be constructed.

In the case of $\mathrm{OSp}(2|2n)^{(1)}$, the fermionic roots are $\pm e_i \pm f$. One defines the vertex operators associated to the fermionic roots $\pm e_i \pm f$ by

$$E(\pm e_i \pm f,z) = U(\pm e_i,z)c_{\pm e_i}\Gamma_i^{\pm}(z)\,, \tag{3.24}$$

with

$$\Gamma_i^{\pm}(z) = (1/\sqrt{2})(U(\xi_i,z)c_{\xi_i}U(-f,z) \pm U(-\xi_i,z)c_{-\xi_i}U(f,z)), \tag{3.25}$$

where $U(\pm f,z)$ are bosonic fields constructed as in Eqs. (3.11) and (3.12):

$$U(f,z) = z^{1/2} :\exp \phi(z): :\exp -\chi(z):,$$

$$U(-f,z) = z^{1/2} :\exp -\phi(z): \partial :\exp \chi(z):.$$

The $\mathrm{SO}(2)$ generator is give by the $U(1)$ current associated to the vertex operators $U(\pm f,z)$, i.e. $P(z) = -z\,\partial\phi(z)$.

In the case of $\mathrm{OSp}(1|2n)^{(1)}$, the fermionic roots are $\pm e_i$. One defines the vertex operators associated to the fermionic roots $\pm e_i$ by

$$E(\pm e_i,z) = U(\pm e_i,z)c_{\pm e_i}\Gamma_i^B(z)\,, \tag{3.26}$$

with

$$\Gamma_i^B(z) = \tfrac{1}{2}\big((U(\xi_i,z)c_{\xi_i} + U(-\xi_i,z)c_{-\xi_i})U(f,z) + (U(\xi_i,z)c_{\xi_i} - U(-\xi_i,z)c_{-\xi_i})U(-f,z)\big)\,, \tag{3.27}$$

the cocycle operators $c_{\pm e_i}$ and $c_{\pm \xi_i}$ being extended to two independent lattices $(1/\sqrt{2})\mathbf{Z}^n$.

D. Vertex operator construction of the unitary affine superalgebra

The vertex construction for $\mathrm{Sl}(m|n)^{(1)}$ is obtained by using the results on the regular semisimple subsuperalgebras of a superalgebra.[5] Actually one can show that $\mathrm{Sl}(m|n)$ is a regular subsuperalgebra of $\mathrm{OSp}(2m|2n)$. In fact the bosonic part of $\mathrm{Sl}(m|n)$ is

$$\mathrm{Sl}(m)\times\mathrm{Sl}(n)\times\mathrm{U}(1)\subset\mathrm{O}(2m)\times\mathrm{Sp}(2n) \tag{3.28}$$

and the fermionic part is

$$(\bar{m},n) + (m,\bar{n}) \subset (2m,2n) \tag{3.29}$$

[compare also Eqs. (2.1) and (2.3) with Eqs. (2.7) and (2.8)]. Therefore, one has, for the bosonic roots $\pm(e_i - e_j)$ $(1\leqslant i\neq j\leqslant m)$ and $\pm(f_i - f_j)$ $(1\leqslant i\neq j\leqslant n)$,

$$E(\pm(e - e_j),z) = :U(\pm e_i,z)c_{\pm e_i}U(\mp e_j,z)c_{\mp e_j}: = U(\pm(e_i - e_j),z)c_{\pm(e_i-e_j)}\,, \tag{3.30}$$

$$E(\pm(f_i - f_j),z) = :U(\pm f_i,z)U(\neq f_j,z):; \tag{3.31}$$

and, for the fermionic roots $\pm(e - f_j)$ $(1\leqslant i \leqslant m$ and $1\leqslant j\leqslant n)$,

$$E(\pm(e_i - f_j),z) = U(\pm e_i,z)c_{\pm e_i}U(\neq f_j,z)\,. \tag{3.32}$$

Notice that the symmetry of the indices m and n in the structure of $\mathrm{Sl}(m|n)^{(1)}$ implies that one can construct two different vertex operator representations: the first one corresponding to $k=1$ for $\mathrm{Sl}(m)$ and $k=-1$ for $\mathrm{Sl}(n)$ and the second one to $k=-1$ for $\mathrm{Sl}(m)$ and $k=1$ for $\mathrm{Sl}(n)$.

IV. VERTEX OPERATORS FOR THE TWISTED AFFINE SUPERALGEBRA

Now following the method developed in Ref. 7, we want to construct the vertex operators for the twisted affine superalgebras (or T.S.A.). Table I shows the different T.S.A.,

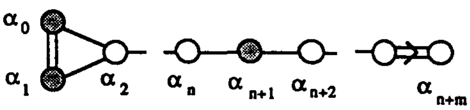
TABLE I. The different twisted affine superalgebras, their invariant sub-(super-) algebra, and the superalgebra they come from via the folding method.

Twisted S. A.	Invariant S. A.	Folded S. A.
$\mathrm{OSp}(2m\|2n)^{(2)}$ $m\geqslant 1$	$\mathrm{OSp}(2m-1\|2n)$	$\mathrm{OSp}(2m+1\|2n)^{(1)}$
$\mathrm{Sl}(1\|2m)^{(2)}$ $n\geqslant 2$	$\mathrm{OSp}(1\|2m)$	$\mathrm{Sl}(2\|2m)^{(2)}$
$\mathrm{Sl}(1\|2m)^{(4)}$ $n\geqslant 2$	$\mathrm{SO}(2m)$	$\mathrm{Sl}(2\|2m)^{(2)}$
$\mathrm{Sl}(2\|2m+1)^{(2)}$ $n\geqslant 2$	$\mathrm{OSp}(2m+1\|2)$	$\mathrm{Sl}(2\|2m+2)^{(2)}$
$\mathrm{Sl}(2\|2m+1)^{(4)}$ $n\geqslant 2$	$\mathrm{OSp}(2\|2m)$	$\mathrm{Sl}(2\|2m+2)^{(2)}$
$\mathrm{Sl}(1\|2m+1)^{(4)}$ $n\geqslant 2$	$\mathrm{OSp}(1\|2m)$	$\mathrm{Sl}(1\|2m+2)^{(2)}$
$\mathrm{Sl}(1\|2m+1)^{(4)}$ $n\geqslant 2$	$\mathrm{SO}(2n+1)$	$\mathrm{Sl}(1\|2m+2)^{(2)}$

their invariant sub-(super-)algebra, and the superalgebra they come from via the folding method.

A. Construction of OSp$(2m/2n)^{(2)}$ (with $m>1$)

One starts from the untwisted affine S.A. $\mathrm{OSp}(2m+1|2n)^{(1)}$ and one considers a particular EDD that exhibits a Z_2 symmetry transforming the affine root into a horizontal root:



OSp(2m+1/2n)$^{(1)}$

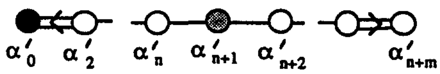
which corresponds to the simple root system

$$\Delta=\{\alpha_0=\delta-e_1-f_1, \alpha_1=e_1-f_1, \alpha_2=f_1-f_2, \ldots,$$
$$\alpha_n=f_{n-1}-f_n,$$
$$\alpha_{n+1}=f_n-e_2, \alpha_{n+2}=e_2-e_3, \ldots,$$
$$\alpha_{n+m-1}=e_{m-1}-e_m, \alpha_{n+m}=e_m\}. \tag{4.1}$$

The outer automorphism τ of order 2 corresponding to the Z_2 symmetry of the diagram above is defined by

$$\tau(\alpha_0)=\alpha_1, \quad \tau(\alpha_i)=\alpha_i \quad (2\leqslant i\leqslant n+m). \tag{4.2}$$

The folded EDD is



OSp(2m/2n)$^{(2)}$

with the simple root system

$$\Delta'=\{\alpha'_0=\delta/2-f_1, \ \alpha'_2=f_1-f_2, \ldots,$$
$$\alpha'_n=f_{n-1}-f_n, \ \alpha'_{n+1}=f_n-e_2,$$
$$\alpha'_{n+2}=e_2-e_3, \ldots,$$
$$\alpha'_{n+m-1}=e_{m-1}-e_m, \ \alpha'_{n+m}=e_m\}, \tag{4.3}$$

which corresponds to the twisted superalgebra $\mathrm{OSp}(2m|2n)^{(2)}$. The invariant integral subsuperalgebra is $\mathrm{OSp}(2m-1|2n)$ and the twisted part is the fundamental representation of $\mathrm{Osp}(2m-1|2n)$ of dimension $2m+2n-1$. Therefore if

$$\Delta'_0=\{\pm e_i\pm e_j, \ 2\leqslant i\neq j\leqslant m,$$
$$\pm f_i\pm f_j, \ \pm 2f_i, \ 1\leqslant i\neq j\leqslant n$$
$$\pm e_i\pm f_i, \ 2\leqslant i\leqslant m, \ 1\leqslant j\leqslant n\}$$

and

$$\Delta'_1=\{\pm e_i, \ 2\leqslant i\leqslant m, \ \pm f_j, \ 1\leqslant j\leqslant n\}, \tag{4.4}$$

the roots of Δ'_0 and Δ'_1 appear at each integral level, whereas the roots of Δ'_1 appear at each half-integral level.

The construction of the vertex operators is given as follows. For the invariant part, the generators associated to the bosonic roots are

$$E(\pm e_i\pm e_j,z)=U(\pm e_i,z)c_{\pm e_i}U(\pm e_j,z)c_{\pm e_j}, \tag{4.5}$$

$$E(\pm e_i,z)=U(\pm e_i,z)c_{\pm e_i}\Gamma_F(z), \tag{4.6}$$

$$E(\pm f_i\pm f_j,z)= \ :U(\pm f_i,z)U(\pm f_j,z):, \quad 1\leqslant i,j\leqslant n, \tag{4.7}$$

where $\Gamma_F(z)$ is the auxiliary fermionic field needed for $\mathrm{OSp}(2m-1|2n)$,

$$\Gamma(z)=(1/\sqrt{2})(U(e_{2m},z)c_{e_{2m}}+U(-e_{2m},z)c_{-e_{2m}});$$

and the generators associated to the fermionic roots are

$$E(\pm e_i\pm f_j,z)=U(\pm e_i,z)c_{\pm e_i}U(\pm f_j,z). \tag{4.8}$$

For the twisted part, the generators at half-integral levels are constructed using a Lorentzian lattice $\overline{\Lambda}_{\mathrm{SO}}$ by extending the root lattice Λ_{SO} of $\mathrm{SO}(2m)$ by the isotropic direction δ: if (e_i) is a basis of the $\mathrm{SO}(2m)$ lattice Λ_{SO}, one extends it with the conditions

$$e_i\cdot\delta=0, \quad \delta^2=0. \tag{4.9}$$

The Fubini–Veneziano fields $Q^i(z)$ are extended from $1\leqslant i\leqslant m$ to $\dim\overline{\Lambda}_{\mathrm{SO}}$ such that

$$\alpha^i_m\cdot\delta=0 \quad (m\neq 0) \quad \text{and} \quad \alpha_0\cdot\delta=\alpha_0\cdot p=1. \tag{4.10}$$

One uses an auxiliary fermionic field constructed by folding and defined by

$$\Gamma'_F(z)=\frac{1}{\sqrt{2}}\left(U\left(e_1+\frac{\delta}{2},z\right)c_{e_1}+U\left(-e_1+\frac{\delta}{2},z\right)c_{-e_1}\right). \tag{4.11}$$

The generators associated to the bosonic short roots $\pm e_i$ at half-integral levels are defined by

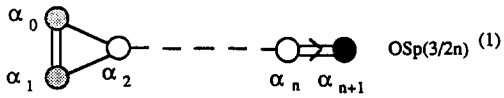
$$E(\pm e_i,z)=U(\pm e_i,z)c_{\pm e_i}\Gamma'_F(z) \tag{4.12}$$

and the generators associated to the fermionic roots $\pm f_j$ at half-integral levels by

$$E(\pm f_j,z)=U(\pm f_j,z)\Gamma'_F(z). \tag{4.13}$$

B. Construction of OSp$(2|2n)^{(2)}$

The construction is very similar to the last one, the only difference being in the fact that the invariant integral subsuperalgebra is now $OSp(1|2n)$, which bosonic part reduces to $Sp(2n)$, and the twisted part is the fundamental representation of $Osp(1|2n)$ of dimension $2n+1$. One starts from the EDD of $OSp(3|2n)^{(1)}$,



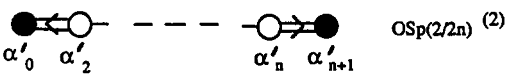
with the simple root system

$$\Delta' = \{\alpha_0 = \delta - e - f_1,\ \alpha_1 = e - f_1,\ \alpha_2 = f_1 - f_2,\ \ldots,$$
$$\alpha_n = f_{n-1} - f_n,\ \alpha_{n+1} = f_n\}. \tag{4.14}$$

The automorphism τ associated to the Z_2 symmetry is defined by

$$\tau(\alpha_0) = \alpha_1, \quad \tau(\alpha_i) = \alpha_i \quad (2 \leqslant i \leqslant n+1). \tag{4.15}$$

The folded EDD is



with the simple root system

$$\Delta' = \{\alpha'_0 = \delta/2 - f_1,\ \alpha'_2 = f_1 - f_2,\ \ldots,$$
$$\alpha'_n = f_{n-1} - f_n,\ \alpha'_{n+1} = f_n\}. \tag{4.16}$$

Therefore the generators are for the invariant part,

$$E(\pm f_i \pm f_j, z) = :U(\pm f_i, z) U(\pm f_j, z):, \quad 1 \leqslant i,j \leqslant n, \tag{4.17}$$

$$E(\pm f_i, z) = U(\pm f_i, z)\Gamma_F(z), \tag{4.18}$$

where $\Gamma_F(z)$ is the auxiliary fermionic field needed in $OSp(1|2n)$; and, for the twisted part, one uses another auxiliary fermionic field constructed by folding and defined by

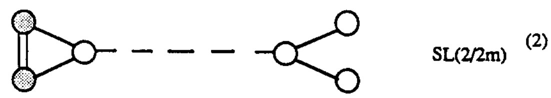
$$\Gamma'_F(z) = (1/\sqrt{2})\big(U(e + \delta/2, z) + U(-e + \delta/2, z)\big). \tag{4.19}$$

The generators associated to the fermionic roots $\pm f_j$ at half-integral levels are

$$E(\pm f_j, z) = U(\pm f_j, z)\Gamma'_F(z). \tag{4.20}$$

C. Construction of $Sl(2|2m)^{(2)}$

Consider the EDD of the twisted superalgebra $Sl(2|2m)^{(2)}$,



associated to two different root systems, corresponding to different end points of the EDD for the affine root. These two root systems are

$$\Delta = \{\alpha_0 = \delta/2 - e_1 - f,\ \ \alpha_1 = e_1 - e_2,\ \ldots,$$
$$\alpha_{m-1} = e_{m-1} - e_m,\ \ \alpha_m = e_{m-1} + e_m,$$
$$\alpha_{m+1} = -e_1 + f\} \tag{4.21}$$

or

$$\Delta = \{\alpha_1 = e_1 - f,\ \ \alpha_2 = e_2 - e_1,\ \ldots,$$
$$\alpha_m = e_m - e_{m-1},\ \ \alpha_{m+1} = e_1 + f,$$
$$\alpha_0 = \delta/2 - e_m - e_{m-1}\}. \tag{4.22}$$

These two root systems correspond to different gradings of the twisted superalgebra. In the first case, the invariant subsuperalgebra is $OSp(2m|2)$, whereas in the second case the invariant subsuperalgebra is $OSp(2|2m)$. Only the last case will keep our attention. The roots

$$\Delta'_0 = \{\pm e_i \pm e_j,\quad \pm 2e_i,\quad \pm e_i \pm f,\ \ 1 \leqslant i \neq \leqslant j \leqslant m\} \tag{4.23}$$

appear at each integral level (invariant part), and the roots

$$\Delta'_1 = \{\pm e_i \pm e_j,\quad \pm e_i \pm f,\ \ 1 \leqslant i \neq j \leqslant m\} \tag{4.24}$$

appear at each half-integral level (twisted part).

The construction of the vertex operators is as follows: for the invariant part,

$$E(\pm 2e_i, z) = U(\pm 2e_i, z) c_{\pm 2e_i} \tag{4.25}$$

$$E(\pm e_i \pm e_j, z) = U(\pm e_i \pm e_j, z) c_{\pm e_j} \Gamma_{ij}(z), \tag{4.26}$$

with $\Gamma_{ij}(z)$ given by Eq. (3.23),

$$E(\pm e_i \pm f, z) = U(\pm e_i, z) c_{\pm e_i} \Gamma_i^{\pm}(z),$$

with

$$\Gamma_i^{\pm}(z) = (1/\sqrt{2})\big(U(\xi_i, z) c_{\xi_i} U(-f, z)$$
$$\pm U(-\xi_i, z) c_{\xi_i} U(f, z)\big); \tag{4.27}$$

and for the twisted part,

$$E'(\pm e_i \pm e_j, z) = U(\pm e_i \pm e_j, z) c_{\pm e_i + e_j} \Gamma'_{ij}(z), \tag{4.28}$$

$$E'(\pm e_i \pm f, z) = U(\pm e_i, z) c_{\pm e_i} \Gamma_i'^{\pm}(z),$$

with

$$\Gamma'_{ij}(z) = (1/\sqrt{2})\big(U(\xi_i + \xi_j + \delta/2, z) c_{\xi_i - \xi_j}$$
$$+ U(-\xi_j - \xi_i + \delta/2, z) c_{\xi_j - \xi_i}\big), \tag{4.29}$$

$$\Gamma_i'^{\pm}(z) = (1/\sqrt{2})\big(U(\xi_i + \delta/2, z) c_{\xi_i} U(f, z)$$
$$\pm U(-\xi_i + \delta/2, z) c_{\xi_i} U(-f, z)\big). \tag{4.30}$$

D. Construction of $Sl(2|2m-1)^{(4)}$

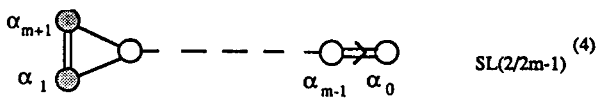
The folding of the EDD of $Sl(2|2m)^{(2)}$ in $Sl(2|2m-1)^{(k)}$ leads to two twisted superalgebras, namely, $\mathrm{Sl}(2|2m-1)^{(2)}$ with invariant subsuperalgebra $\mathrm{OSp}(2m-1|2)$ and $\mathrm{Sl}(2|2m-1)^{(4)}$ with invariant subsuperalgebra $\mathrm{OSp}(2|2m-2)$. We will construct a vertex operator representation for the last one. One considers the EDD of $\mathrm{Sl}(2|2m)^{(2)}$ associated to the root system

$$\Delta=\{\alpha_1=e_1+f,\quad \alpha_2=e_2-e_1,\ \ldots,\quad \alpha_m=e_m-e_{m-1},\quad \alpha_{m+1}=e_1+f,\quad \alpha_0=\delta/2-e_m-e_{m-1}\}\,. \tag{4.31}$$

The automorphism τ associated to the Z_2 symmetry is defined by

$$\tau(\alpha_0)=\alpha_m,\quad \tau(\alpha_i)=\alpha_i\quad (1\leqslant i\leqslant m-1)\,. \tag{4.32}$$

The folded diagram is



associated to the root system

$$\Delta=\{\alpha_1=e_1+f,\quad \alpha_2=e_2-e_1,\ \ldots,\quad \alpha_{m+1}=e_1-f,\quad \alpha_0'=\delta/4-e_{m-1}\}\,. \tag{4.33}$$

The roots

$$\Delta_0'=\{\pm e_i\pm e_j,\quad \pm 2e_i,\quad \pm e_i\pm f,\quad 1\leqslant i\neq\leqslant j\leqslant m-1\} \tag{4.34}$$

appear at each integral level (invariant part); the roots

$$\Delta_1'=\{\pm e_i\pm e_j,\quad \pm e_i\pm f,\quad 1\leqslant i\neq j\leqslant m-1\} \tag{4.35}$$

appear at each half-integral level (twisted part); and the roots

$$\Delta_2'=\{\pm e_i,\quad 1\leqslant i\leqslant m-1\} \tag{4.36}$$

appear at each level $\mathbb{Z}+\frac{1}{4}$ and $\mathbb{Z}+\frac{3}{4}$ (twisted part).

The vertex operators are constructed as follows: for the invariant part, the generators associated to the roots at level $\mathbb{Z}$ are given by the formulas (4.32)–(4.34), with $1\leqslant i\neq j\leqslant m-1$; and, for the twisted part, the generators associated to the roots at level $\mathbb{Z}+\frac{1}{2}$ are given by the formulas (4.35)–(4.37), with $1\leqslant i\neq j\leqslant m-1$. The generators associated to the roots e_i at level $\mathbb{Z}+\frac{1}{4}$ are given by

$$E''(\pm e_i,z)=\tfrac{1}{2}U(\pm e_i,z)\Gamma_i''^{B}(z)\,, \tag{4.37}$$

with

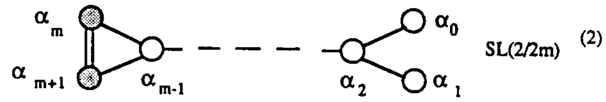
$$\Gamma_i''^{B}(z)=\big(U(e_m+\delta/4,z)\Gamma_{im}(z)+U(-e_m-\delta/4,z)\Gamma_{im}'(z)\big)\,, \tag{4.38}$$

and the generators associated to the roots e_i at level $\mathbb{Z}+\frac{3}{4}$ are obtained by action of the generators at level $\mathbb{Z}+\frac{1}{2}$ on the previous generators at level $\mathbb{Z}+\frac{1}{4}$.

E. Construction of $\mathrm{Sl}(1|2m)^{(2)}$

The folding of the EDD of $\mathrm{Sl}(2|2m)^{(2)}$ in $\mathrm{Sl}(1|2m)^{(k)}$ leads to two twisted superalgebras, namely, $\mathrm{Sl}(1|2m)^{(2)}$ with invariants subsuperalgebra $\mathrm{OSp}(1|2m)$ and $\mathrm{Sl}(1|2m)^{(4)}$ with invariant subsuperalgebra $\mathrm{SO}(2m)$. We will only treat the case of $\mathrm{Sl}(1|2m)^{(2)}$.

One starts from the twisted affine superalgebra $\mathrm{Sl}(2|2m)^{(2)}$ with EDD



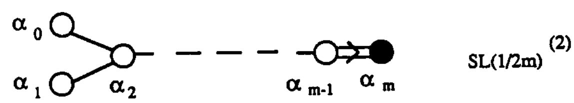
associated to the root system

$$\Delta=\{\alpha_0=\delta/2-e_1-e_2,\quad \alpha_1=e_1-e_2,\ \ldots,\quad \alpha_{m-1}=e_{m-1}-e_m,\quad \alpha_m=e_m-f,\quad \alpha_{m+1}=e_m+f\}\,. \tag{4.39}$$

The automorphism that defines the folding is

$$\tau(\alpha_m)=\alpha_{m+1},\quad \tau(\alpha_i)=\alpha_i\quad (0\leqslant i\leqslant m-1)\,. \tag{4.40}$$

The folded EDD is



with the corresponding simple root system

$$\Delta'=\{\alpha_0=\delta/2-e_1-e_2,\quad \alpha_1=e_1-e_2,\ \ldots,\quad \alpha_{m-1}=e_{m-1}-e_m,\quad \alpha_m'=e_m\}\,. \tag{4.41}$$

One obtains the twisted superalgebra $\mathrm{Sl}(1|2m)^{(2)}$. The invariant integral subalgebra is $\mathrm{OSp}(1|2m)$. The roots

$$\Delta_0'=\{\pm e_i\pm e_j,\quad \pm 2e_i,\quad \pm e_i,\quad 1\leqslant i\neq\leqslant j\leqslant m\} \tag{4.42}$$

appear at each integral level (invariant part), and the roots

$$\Delta_1'=\{\pm e_i\pm e_j,\pm e_i,\quad 1\leqslant i\neq j\leqslant m\} \tag{4.43}$$

appear at each half-integral level (twisted part).

The construction of the vertex operators is given by, for the invariant part, the generators associated to the bosonic roots are

$$E(\pm 2e_i,z)=U(\pm 2e_i,z)c_{\pm 2e_i}\,, \tag{4.44}$$

$$E(\pm e_i\pm e_j,z)=U(\pm e_i\pm e_j,z)c_{\pm e_i\pm e_j}\Gamma_{ij}(z)\,, \tag{4.45}$$

with

$$\Gamma_{ij}(z)=(1/\sqrt{2})\big(U(\xi_i-\xi_j,z)c_{\xi_i-\xi_j}+U(\xi_j-\xi_i,z)c_{\xi_j-\xi_i}\big)\,.$$

The generators associated to the fermionic roots are

$$E(\pm e_i,z)=(1/\sqrt{2})U(\pm e_i,z)c_{e_i}\Gamma_i^{B}(z)\,, \tag{4.46}$$

with

$$\Gamma_i^{B}(z)=\tfrac{1}{2}\big((U(\xi_i,z)c_{\xi_i}+U(-\xi_i,z)c_{-\xi_i})U(f,z)+(U(\xi_i,z)c_{\xi_i}-U(-\xi_i,z)c_{-\xi_i}U(-f,z)\big)\,; \tag{4.47}$$

and, for the twisted part, the generators associated to the bosonic roots are

$$E'(\pm e_i\pm e_j,z)=U(\pm e_i\pm e_j,z)c_{\pm e_i\pm e_j}\Gamma_{ij}'(z)\,, \tag{4.48}$$

with

$$\Gamma'_{ij}(z) = (1/\sqrt{2})(U(\xi_i + \xi_j + \delta/2,z)c_{\xi_i+\xi_j} + U(-\xi_j - \xi_i + \delta/2,z)c_{-\xi_j-\xi_i}) . \tag{4.49}$$

The generators associated to the fermionic roots are

$$E'(\pm e_i,z) = (1/\sqrt{2})U(\pm e_i,z)c_{e_i}\Gamma_i'^{B}(z) ,$$

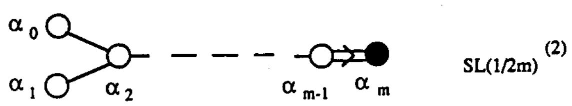
with

$$\Gamma_i'^{B}(z) = \tfrac{1}{2}((U(\xi_i + \delta/2,z)c_{\xi_i} + U(-\xi_i + \delta/2,z)c_{-\xi_i})U(-f,z) + (U(\xi_i + \delta/2,z)c_{\xi_i} - U(-\xi_i + \delta/2,z)c_{-\xi_i}) \times U(f,z)) . \tag{4.50}$$

F. Construction of $\mathrm{Sl}(1|2m-1)^{(4)}$

The twisted superalgebra $\mathrm{Sl}(1|2m+1)^{(4)}$ can be constructed with two different $\mathbf{Z}_2$ gradings, corresponding to opposite end points of the Dynkin diagram for the affine root. In the first case, the integral invariant subsuperalgebra is $\mathrm{OSp}(1|2m-2)$, whereas in the second case, it is $\mathrm{SO}(2m-1)$. We will concentrate on the first case.

One starts from the twisted affine superalgebra $\mathrm{Sl}(1|2m)^{(2)}$ with EDD



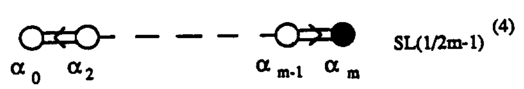
associated to the simple root system

$$\Delta' = \{\alpha_0 = \delta/2 - e_1 - e_2, \quad \alpha_1 = e_1 - e_2, \ ..., \quad \alpha_{m-1} = e_{m-1} - e_m, \quad \alpha_m = e_m\} . \tag{4.51}$$

The automorphism that defines the folding is

$$\tau(\alpha_0) = \alpha_1, \quad \tau(\alpha_i) = \alpha_i \quad (2 \leqslant i \leqslant m) . \tag{4.52}$$

The folded EDD is



with the corresponding simple root system

$$\Delta' = \{\alpha_0' = \delta/4 - e_2, \quad \alpha_2 = e_2 - e_3, \ ..., \quad \alpha_{m-1} = e_{m-1} - e_m, \quad \alpha_m = e_m\} . \tag{4.53}$$

One obtains the twisted superalgebra $\mathrm{Sl}(1|2m-1)^{(4)}$. The invariant integral subalgebra is $\mathrm{OSp}(1|2m-2)$. The roots

$$\Delta_0' = \{\pm e_i \pm e_j, \quad \pm 2e_i, \quad 2 \leqslant i \neq \leqslant j < m\} \tag{4.54}$$

appear at each integral level (invariant part); the roots

$$\Delta_1' = \{\pm e_i \pm e_j, \quad \pm e_i, \quad 2 \leqslant i \neq j \leqslant m\} \tag{4.55}$$

appear at each half-integral level (twisted part); and the roots

$$\Delta_1'' = \{\pm e_i, 2 \leqslant i \leqslant m\} \tag{4.56}$$

appear at each level $\mathbf{Z} + \frac{1}{4}$ and $\mathbf{Z} + \frac{3}{4}$ (twisted part). Notice that the generators $E(e_i,z)$ and $E'(e_i,z)$ associated to the roots e_i at integral and half-integral levels, respectively, have a fermionic nature, whereas the generators $E''(e_i,z)$ associated to the roots e_i at the levels $\mathbf{Z} + \frac{1}{4}$ and $\mathbf{Z} + \frac{3}{4}$ have a bosonic nature.

The construction of the vertex operators is as follows: for the invariant part, the generators associated to the bosonic roots $\pm 2e_i$ and $\pm e_i \pm e_j$ are given by the formula (4.62) and the generators associated to the fermionic ones $\pm e_i$ by the formula (4.64), with $2 \leqslant i \neq j \leqslant m$; for the twisted part at level $\mathbf{Z} + \frac{1}{2}$, the generators associated to the bosonic roots $\pm e_i \pm e_j$ are given by the formula (4.66) and the generators associated to the fermionic ones $\pm e_i$ by the formula (4.68), with $2 \leqslant i \neq j \leqslant m$; and for the twisted part at level $\mathbf{Z} + \frac{1}{4}$, the generators associated to the bosonic roots $\pm e_i$ at level $\mathbf{Z} + \frac{1}{4}$ are given by

$$E''(\pm e_i,z) = \tfrac{1}{2}U(\pm e_i,z)\Gamma_i''^{B}(z) , \tag{4.57}$$

with

$$\Gamma_i''^{B}(z) = (U(e_1 + \delta/4,z)\Gamma_{1i}(z) + U(-e_1 - \delta/4,z)\Gamma'_{1i}(z)) . \tag{4.58}$$

The generators associated to the bosonic roots $\pm e_i$ at level $\mathbf{Z} + \frac{3}{4}$ are obtained by action of the generators associated to the short roots $\pm e_i \pm e_j$ at level $\mathbf{Z} + \frac{1}{2}$ on the generators associated to the bosonic roots $\pm e_i$ at level $\mathbf{Z} + \frac{1}{4}$.

In summary, Fig. 1 gives the different foldings with invariant subsuperalgebras $\mathscr{G}_0$ in the case of the unitary superalgebras.

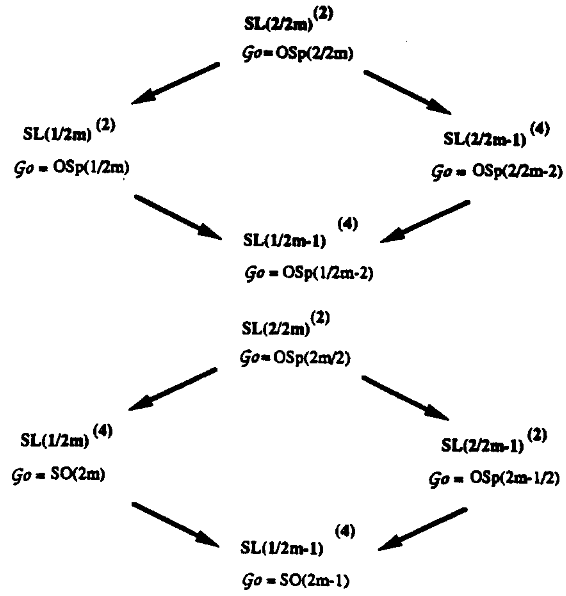


FIG. 1. The different foldings with invariant subsuperalgebras $\mathscr{G}_0$ in the case of unitary superalgebras.

[1]P. Goddard and D. Olive, Int. J. Mod. Phys. **1**, 304 (1986), and references therein.
[2]P. Goddard, D. Olive, and G. Waterson, Commun. Math. Phys. **112**, 591 (1987).
[3]L. Frappat, Int. J. Mod. Phys. **3**, 2545 (1988).
[4]I. B. Frenkel and A. Feingold, Adv. Math. **56**, 117 (1985).
[5]L. Frappat, A. Sciarrino, and P. Sorba, Commun. Math. Phys. **121**, 457 (1989).
[6]T. Jarvis and H. Green, J. Math. Phys. **20**, 10 (1979).
[7]L. Frappat, A. Sciarrino, and P. Sorba, Nucl. Phys. B **305**, 164 (1988).

Solution of Dyson equation for elastic membranes

H. Kleinert
Institut für Theorie der Elementarteilchen, Freie Universität Berlin, Arnimallee 14, D-1000 Berlin 33, Federal Republic of Germany

(Received 16 June 1989; accepted for publication 2 August 1989)

The Dyson equation for the renormalization of the curvature stiffness by elastic fluctuations into a simple differential equation is transformed and solved iteratively as well as numerically. If the tension parameter falls below some critical value, the inverse two-point function develops discontinuously a minimum at a nonzero momentum p_0. This should be relevant for understanding a transition in elastic membranes.

Although elastic forces do not contribute to the ultraviolet divergencies in the extrinsic curvature stiffness of membrane undulations,[1] they have important effects in the infrared.[2] To see this most easily, the elastic energy ($u;j$ = strain tensor)

$$E_{\rm el} = \int d^2\xi \sqrt{g}\left[\mu\, u_i j^2 + \left(\frac{\lambda}{2}\right) u_i\, i^2\right] \tag{1}$$

of an almost flat surface with tangential and vertical displacements $\tau^i(\xi)$ and $u(\xi)$,

$$x^1 = \xi^1 + \tau^1(\xi), \quad x^2 = \xi^2 + \tau^2(\xi), \quad x^3(\xi) = u(\xi),$$

is approximated by (we set $u_i \equiv \partial_i u$)

$$E_{\rm el} = \int d^2\xi \left[\frac{\mu}{4}(\partial_i\tau_j + \partial_j\tau_i - u_i u_j)^2 + \frac{\lambda}{2}\left(\partial_i\tau_i - \frac{1}{2}u_i u_i\right)^2\right]. \tag{2}$$

An auxiliary fluctuating stress field σ_{ij} is introduced to rewrite this in a canonical form

$$E_{\rm el} = \int d^2\xi \left[\frac{1}{4\mu}\left[\sigma_{ij}^2 - \frac{\nu}{(1+\nu)}\sigma_{ii}^2\right] + \frac{i}{2}\sigma_{ij}[\partial_i\tau_j + \partial_i\tau_i - u_i u_j]\right], \tag{3}$$

where $\nu = \lambda/(\lambda + 2\mu)$ is the Poisson ratio. Integrating out the τ_i fields gives $\partial_i\sigma_{ij} = 0$, to be enforced via a stress gauge field χ by setting $\sigma_{ij} \equiv \epsilon_{ik}\epsilon_{jl}\,\partial_k\,\partial_l\chi$ and $E_{\rm el}$ becomes

$$\begin{aligned} E_{\rm el} &= \int d^2\xi \left[\frac{1}{4\mu(1+\nu)}(\partial^2\chi)^2 + \frac{i}{2}\chi\,\epsilon_{ik}\epsilon_{jl}\,\partial_i\,\partial_j u_k\, u_l\right] \\ &= \frac{\mu(1+\nu)}{4}\int d^2\xi (P_{kl}u_k u_l)^2, \end{aligned} \tag{4}$$

where $P_{ij} = \epsilon_{ik}\epsilon_{jl}\,\partial_k\,\partial_l/\partial^2$ is the transverse projection. The quantity $\epsilon_{ik}\epsilon_{jl}\,\partial_i\,\partial_j u_k u_l = 2\det(\partial_i\,\partial_j u) = R$ is the scalar (= twice Gaussian) curvature of the surface {recall that $R = C^2 - C_i^{\,j} C_j^{\,i}$, where $C_i^{\,j}$ is the extrinsic curvature matrix $C_i^{\,j} = \partial_i[(1+(\partial_l u)^2)^{-1/2}\,\partial_j u]$}.

As pointed out in Ref. 2, a phonon renormalizes the bending and stretching energy

$$E_0 = \frac{1}{2}\int d^2\xi\,[k(\partial^2 u)^2 + r(\partial u)^2] \tag{5}$$

and brings it to a momentum-dependent form $E = \frac{1}{2}\int d^2\xi\, u\Gamma(-i\partial)u$, where $\Gamma(p)$ is determined by the Dyson equation

$$\Gamma(p) = kp^4 + rp^2 + \mu\,(1+\nu)T\left(\frac{1}{4}\right)\int\left(\frac{d^2q}{(2\pi)^2}\right) \times [p_j p_i P_{ij}(q)]^2/\Gamma(q+p)\,. \tag{6}$$

The purpose of this paper is to solve this equation and find a phase transition that may be related to transitions in physical membranes. In Ref. 2 it was suggested that a transition in the model $E_0 + E_{\rm el}$ may be related to the crumpling transition seen in Monte Carlo simultions,[3] but this seems dubious since the model surface under consideration lacks the rotational invariance of the simulated system.

Certainly, the solution to the equation (6) to be presented here is not the solution of the entire one-loop renormalization problem. For this, the renormalization of the elastic energy[4] will also have to be considered. The solution of Eq. (6), however, is an interesting problem in its own right and presents a necessary first step. In particular, we find it worth communicating that the complicated looking integral [Eq. (6)] can be transformed into a very simple differential equation. This transformation goes as follows: We observed that after replacing $q + p \to q$, the numerator in the integrand of (6) becomes $(1-\cos^2\theta)/[1 + (q/p)^2 - 2(q/p)\cos\theta]^2$, where $\cos\theta = p\cdot q/\sqrt{p^2q^2}$. The integral over θ gives simply $(3\pi/4)q^4$ for $q<p$ and $(3\pi/4)p^4$ for $q>p$. Hence we can rewrite (6) as

$$\Gamma(p) = kp^4 + rp^2 + \mu(1+\nu)T\left(\frac{3}{128\pi}\right) \times \left[\int_0^{p^2} dp^2\, q^4\Gamma^{-1}(q) + p^4\int_{p^2}^{\infty} dq^2\,\Gamma^{-1}(q)\right]. \tag{7}$$

Introducing the variable $s = p^2$ and setting $\mu(1+\nu)T(3/128\pi) = t/2$, we obtain from this the simple differential equation ($' \equiv d/ds$),

$$(\Gamma'(s)/s)' + t/\Gamma(s) = -r/s^2. \tag{8}$$

We now seek for a solution, which for large s behaves like

$$\Gamma(s) = rs + ks^2 + t\Sigma(s), \tag{9}$$

with a self-energy $\Sigma(s)$, which grows weaker than ks^2 and satisfies the equation

$$\Sigma'' - \Sigma'/s + s/(r + ks^2 + t\Sigma) = 0. \tag{10}$$

We now insert the ansatz

$$t\Sigma(s) = \frac{t}{k}s - \frac{t}{2k^3}(rk + t)\ln s + s\sum_{j=2}^{\infty}\sum_{j=0}^{i} c_{ij}s^{-i}\ln^j s, \tag{11}$$

and work out the coefficients c_{ij} iteratively, with the result shown in Table I.

For small s, we use the ansatz

$$t\Sigma = cs^2 - \frac{t}{2r}s^2\ln s + s^2\sum_{i=1}^{\infty}\sum_{j=0}^{i} a_{ij}s^i\ln^j s, \tag{12}$$

and find the coefficients shown in Table II, where we have replaced $k + c$ by k'.

The solutions for small and large s can be matched with each other by an appropriate choice of $c<0$ and a numerical solution of the differential equation (9) in the intermediate regime. For a typical configuration of parameters $k = 1$, $t = 1$, the curves $\Gamma(s)$ approach, for large s, rapidly the leading terms in (9) and (11),

$$\Gamma(s) \sim (r + t/k)s + ks^2 - (t/2k^2)(r + t/k)\ln s \tag{13}$$

(see the dotted curves in Fig. 1). For the small s series (12), the radius of convergence is, for small r, very small since the coefficients a_{ij} involve inverse powers of r. In particular, the first two terms in (12) give

$$\Gamma(s) = rs + k's^2 - (t/2r)s^2\ln s, \tag{14}$$

and this is applicable only very near $s = 0$. It can be used to fix the initial value and the slope of $\Gamma(s)$ for the numerical integration of the differential equation from small to large s. For $r \gtrsim 0$, there is a phase transition at which the flat ground state destabilizes, going over into a wrinkled state of nonzero wave vector $p_0{}^2 \sim 0.9$, not far from the lowest estimate $p_0{}^2 \approx e^{-1/2}$, which can be obtained from (12). This situation is similar to the transitions of spontaneous strings in large dimensions,[5] the main difference being the absence of cubic field terms in the present case.

For $r = 0$, there exists an exact solution (8),

TABLE I. Large p^2 expansion of inverse undulation correlation $\Gamma(p^2)$; $\Gamma(s) = ks^2 + (r + t/k)s - (\frac{1}{2})(t/k^3)(rk + t)\log s + \Sigma_{i=2}^{\infty}\Sigma_{j=0}^{i} c(i,j)s^i\log^j s$.

$C(2,0) = -k^{-5}t(\frac{1}{3}k^2r^2 + \frac{8}{9}krt + \frac{5}{9}t^2)$
$C(2,1) = \frac{1}{6}k^{-5}t^2(kr + t)$
$C(3,0) = k^{-7}t(\frac{1}{8}k^3r^3 + \frac{41}{96}k^2r^2t$
$\quad + \frac{251}{576}krt^2 + \frac{77}{576}t^3)$
$C(3,1) = -k^{-7}t^2(\frac{1}{8}k^2r^2 + \frac{11}{48}krt + \frac{5}{48}t^2)$
$C(4,0) = -k^{-9}t(\frac{1}{15}k^4r^4 + \frac{481}{1800}k^3r^3t$
$\quad + \frac{12943}{36000}k^2r^2t^2$
$\quad + \frac{40541}{216000}krt^3 + \frac{6203}{216000}t^4)$
$C(4,1) = k^{-9}t^2(\frac{1}{10}k^3r^3 + \frac{517}{1800}k^2r^2t$
$\quad + \frac{331}{1200}krt^2 + \frac{319}{3600}t^3)$
$C(4,2) = -k^{-9}t^3(\frac{1}{60}k^2r^2 + \frac{1}{30}krt + \frac{1}{60}t^2)$

TABLE II. Small p^2 expansion of inverse undulation correlation $\Gamma(p^2)$; $\Gamma(s) = rs + k's^2 - (t/2r)s^2\log s + \Sigma_{i=1}^{\infty}\Sigma_{j=0}^{i} a(i,j)s^{i+2}\log^j s$.

$a(1,0) = r^{-3}t(\frac{1}{3}k'r + \frac{2}{9}t)$
$a(1,1) = -\frac{1}{6}r^{-3}t^2$
$a(2,0) = -r^{-5}t(\frac{1}{8}k'^2r^2 + \frac{5}{96}k'rt - \frac{37}{2304}t^2)$
$a(2,1) = r^{-5}t^2(\frac{1}{8}k'r + \frac{5}{192}t)$
$a(2,2) = -\frac{1}{32}r^{-5}t^3$
$a(3,0) = r^{-7}t(\frac{1}{15}k'^3r^3 + \frac{1}{1800}k'^2r^2t$
$- \frac{1421}{36000}k'rt^2 - \frac{40927}{4320000}t^3)$
$a(3,1) = -r^{-7}t^2(\frac{1}{10}k'^2r^2 + \frac{1}{1800}k'rt - \frac{1421}{72000}t^2)$
$a(3,2) = r^{-7}t^3(\frac{1}{20}k'r + \frac{1}{7200}t)$
$a(3,3) = -\frac{1}{120}r^{-7}t^4$

$$\Gamma_0 = \sqrt{4t/3}\, s^{3\ 2}. \tag{15}$$

This, however, is useles for membranes since it corresponds to a vanishing curvature stiffness k. The small s behavior, however, also agrees with Γ_0 for $k \neq 0$. In fact, as r passes from 0^+ to 0^-, the solution flips from the upper to the lower branch of the square root. The zero in $\Gamma(p^2)$ at $p^2 = s_0 \neq 0$ is an essential singularity. As Γ passes through zero, s increases like $s - s_0 \approx (2ts_0)^{-1/2}(-\ln\Gamma)^{-1/2}\Gamma$, which explains the infinite slopes of $\Gamma(s)$ at s_0.

The coefficients of the large s expansion at $r = 0$ can be taken from Table II.

The renormalization resulting from the nonlinearities in the curvature terms at the one-loop level ($u_{ij} \equiv \partial_i\partial_j u$)

$$E_{\rm nl\,curv} = \frac{1}{2}\int d^2\xi\left[-\frac{1}{2}u_{ii}{}^2u_j{}^2 - 2u_{ii}u_ku_lu_{kl} + \cdots\right] \tag{16}$$

gives only a rather trivial change of $\Gamma(p^2)$,

$$\Delta\Gamma(p^2) = -(3T/2)Lp^4 - (3T/2)Qp^2, \tag{17}$$

where $\quad L \equiv \int (dp/2\pi)^2)(p^2/\Gamma(p)), \quad Q \equiv \int(d^2p/(2\pi)^2)$

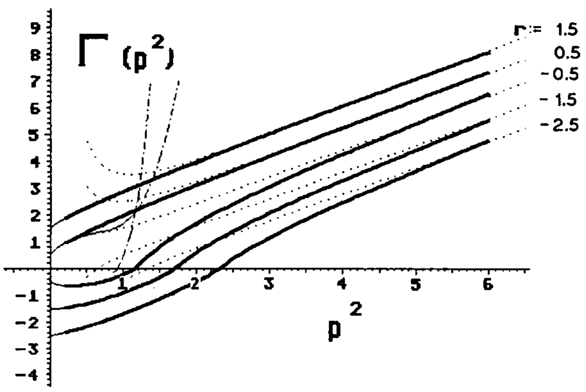


FIG. 1. The dispersion curves of undulations in elastic membranes (divided by p^2, to show the asymptotic behavior kp^4 as a straight line). We have chosen $k = 1$ and varied the tension parameter r from 1.5 to -2.5. At $r = 0$, the curve jumps from an initial behavior $\sqrt{4t/3}\, s^{3/2}$ to $-\sqrt{4t/3}\, s^{3/2}$ and a new minimum of Γ arises at $s \sim 0.9$, where $\Gamma_{\min}$ jumps from 0 to ≈ -0.2. The dotted lines are the large p^2 approximation (11). The dashed-dotted lines starting from $p^2 = 0$ show the small s solution (12); for $r \gtrsim 1.5$ they lie on top of the flat lines up to $p^2 \sim 1.5$. For $r = \pm 0.5$ the convergence breaks down earlier.

$\times\langle(p^4/\Gamma(p))\rangle$ are constants, with a divergence that can be absorbed into the initial values of k and r. The differential equation (8) will therefore remain unchanged and the solutions, when parametrized directly by the k, r values observed at large s, will be the same as before.

Let us end by noting that if the elastic membrane has undergone a continuous premelting transition to a phase with only orientational order (hexatic phase), the Dyson equation contains one extra factor of q^2 in the integrand (6)[2,6] and we find quite similarly, the integral equation

$$\Gamma(p^2) = kp^4 + rp^2 + \frac{t}{2}p^2\left[\int_0^{p^2} dq^2\, q^4\left(3 - \frac{q^2}{p^2}\right)\Gamma^{-1}(q) + p^4\int_{p^2}^{\infty} dp^2\left(3\frac{q^2}{p^2} - 1\right)\Gamma^{-1}(q)\right], \tag{18}$$

where $t \equiv K_A T/32\pi^2$ and K_A is the angular stiffness. This equation can be converted into the differential equation

$$\Gamma'''' = 3t\Gamma^{-1}; \tag{19}$$

it can be solved by similar methods.

ACKNOWLEDGMENTS

The author thanks Prof. Dr. Nelson, Dr. S. Ami, and F. Langhammer for discussion at the initial stage of this work.

This work was supported in part by the Deutsche Forschungsgemeinschaft under Grant No. Kl 256.

[1]S. Ami and H. Kleinert, Phys. Lett. A **120**, 207 (1987). Notice that the intrinsic curvature energy *does* receive substantial ultraviolet renormalization from elasticity.

[2]D. R. Nelson and L. Peliti, J. Phys. (Paris) **48**, 1085 (1987); also see the discussion in infinite dimensions by F. David and D. Guitter, Europhys. Lett. **5**, 709 (1988).

[3]J. Kantor and D. R. Nelson, Phys. Rev. A **36**, 4020 (1987).

[4]H. Aronowitz and T. C. Lubensky, University of Pennsylvania preprint.

[5]H. Kleinert, Phys. Lett. B **197**, 125 (1987).

[6]F. David, L. Peliti, and E. Guitter, J. Phys. (Paris) **48**, 2059 (1987).

Equilibrium states of a class of quantum mean-field theories

Pavel Bóna
Department of Theoretical Physics, Faculty of Mathematics and Physics, Comenius University, 842 15 Bratislava, Czechoslovakia

(Received 15 September 1988; accepted for publication 28 June 1989)

The general structure and existence of equilibrium and ground states of infinite spin systems with the mean-field type interaction of Hepp and Lieb [Helv. Phys. Acta **46**, 573 (1973)] is investigated. The form and the role of the "Bogoliubov–Haag" Hamiltonians is cleared. The analysis is based on a recent mathematically correct definition of the corresponding time automorphism group τ^Q (specified by a classical Hamiltonian Q) of an appropriately defined "physical" C^*-algebra of observables $\mathscr{C}$ containing an abelian subalgebra $\mathscr{N}$ of classical macroscopic quantities as well as the conventional "bare" quasilocal (sub-) algebra $\mathscr{A}$, $\mathscr{C} = \mathscr{A} \otimes \mathscr{N}$ [P. Bóna, J. Math. Phys. **29**, 2223 (1988)]. A Lie group G of "hidden symmetries" acting nontrivially (also) on $\mathscr{N}$ plays an essential role in the presented theory. The applications to the quasispin strong-coupling version of the BCS model of superconductivity, and to the corresponding model of the Josephson junction illustrate possibilities of the developed formalism as well as an easy manipulation with it. A simple mathematical description of "symmetry breaking" is sketched for the considered class of mathematical models of large quantal systems.

I. INTRODUCTION

Mathematically clear description of macroscopic properties of equilibrium states of large physical systems (such as phase transitions) is available only in the framework of the theory of infinite systems.[1,2] The first mathematically correct proof of possibility of description of phase transitions in the framework of equilibrium statistical physics of infinite systems was the Onsager's calculation[3] of thermodynamic potentials of the two-dimensional Ising model. In trials to overcome difficulties with Hilbert space descriptions of quantum field theory[4] a representation-independent formalism of quantum theory of infinite system was developed,[5] i.e., the formalism of (C^*-) algebraic quantum theory generalizing the elementary Hilbert space formalism in a natural way. The main mathematical object in algebraic formulation of quantum theory of (spatially, or in other variables) extended systems is a "quasilocal algebra of observables" $\mathscr{A}$ built of a net of its "local subalgebras" $\mathscr{A}^\Lambda$ associated with members $\Lambda \in \mathscr{F}_\Pi$ of the set of "finite regions" of some infinite "configuration set" Π. In the framework of algebraic formalism, models for quantum-theoretic description of interesting collective many-constituents behavior like various types of phase transitions were constructed.[1a, 2,6,7]

The usual physically intuitive method of theoretical investigation of thermal equilibria of infinite systems consists in taking some—for the considered physical situation most appropriate—kind of thermodynamic limit[7] of local Gibbs states. This method is particularly convenient in investigation of surface effects[8] and stability properties[9,10] of the system. It is almost inevitable to use this method in investigation of infinite *classical* lattice systems.[11] In the last case, there is not known any way of direct (i.e., without going to a form of "quantization" of the considered classical system) determination of time evolution: We have not defined any canonical formalism associating some dynamics with a given Hamiltonian for classical lattice systems.

Infinite *quantum* systems afford also an alternative method of investigation of equilibrium states. It is applicable, strictly speaking, only in the cases in which there is given a mathematically correct definition of dynamics of the whole infinite system as a one-parameter group of automorphisms of the C^*-algebra $\mathscr{A}$ of observables. This method uses the KMS condition (after Kubo, Martin, and Schwinger) introduced into the algebraic formulation of quantum theory by Haag, Hugenholtz, and Winnink.[12] The KMS condition was extensively used[6] for investigation of spin systems with short-range interactions, in which cases the correct definition of the global time evolution was known for a long time.[13] In some cases of long-range interactions, a reformulation of the KMS condition not using the global dynamics of the infinite system was introduced: It was formulated in terms of limits of expressions containing derivations of local time evolutions[10] instead. In these last mentioned cases, however, the question was left open as to the existence of some connection of such a reformulated KMS condition with some one-parameter group of (global-automorphic) time translations. Long-range interactions were studied almost exclusively in terms of local Hamiltonians and corresponding local dynamics. Some very interesting results were obtained in this way, especially in investigations of the mathematical structure of appropriately simplified versions of the Bardeen–Cooper–Schrieffer (BCS) model of superconductivity.[14–17] Clear interrelations between different observed phenomena like spontaneous symmetry breaking, phase transitions, and the possibility of a description of dynamics of local perturbations of some stationary states by the linearized ("Bogoliubov–Haag") form of Hamilton operator remained, however, missing.

The KMS states of a class of infinite quantal spin systems with a mean-field type (hence long-range) interactions are investigated in this paper. The considered class of systems include the arbitrary polynomial mean-field systems introduced by Hepp and Lieb,[18] which in turn include, e.g.,

 0022-2488/89/122994-14$02.50

the quadratic models of the BCS theory in its quasispin formulation in the strong coupling limit,[15] including a model of the Josephson junction.[18,19] A mathematically correct (global) dynamics of these models was formulated in a recent paper[20]; that paper also contains a brief description of the history of the problem with citations of some relevant papers.[21] Let us sketch here a formulation of the problem for the case of polynominal models.

Let $\mathscr{A}^p$ ($p\in\Pi$) be copies of the algebra $\mathscr{A}^0 := \mathscr{L}(\mathscr{H}_0)$ of all bounded operators on the finite-dimensional complex Hilbert space $\mathscr{H}_0$, and let these algebras be connected by isomorphisms

$$\pi_p:\quad \mathscr{A}^0\to\mathscr{A}^p,\quad x\mapsto\pi_p(x),\quad x\in\mathscr{A}^0. \tag{1.1}$$

The conventional C^*-algebra $\mathscr{A}$ of quasilocal observables (we shall call it the "bare algebra"—it is chosen independently of interactions) is defined as the C^*-inductive limit[22] of finite tensor products of local algebras

$$\mathscr{A}^\Lambda := \bigotimes_{p\in\Lambda}\mathscr{A}^p,\quad \Lambda\in\mathscr{F}_\Pi. \tag{1.2}$$

Here, $\mathscr{A}$ can be considered as an infinite tensor product, $\mathscr{A}=\bigotimes_{p\in\Pi}\mathscr{A}^p$.

Let a polynomial Q,

$$Q(F_1,F_2,\dots,F_n) := \sum_{m=1}^{q}\sum_{j_1j_2\cdots j_m} c^{(m)}_{j_1j_2\cdots j_m}F_{j_1}F_{j_2}\cdots F_{j_m}, \tag{1.3}$$

of n variables and of degree q be given in such a way that the operators Q^Λ defined below are all self-adjoint. Let $X_j\in\mathscr{A}^0$ ($j=1,2,\dots,n$) be arbitrary chosen (mutually linearly independent) self-adjoint matrices, and define the local intensive observables $X_{j\Lambda}$ by the formula

$$X_{j\Lambda} := \frac{1}{|\Lambda|}\sum_{p\in\Lambda}\pi_p(X_j),\quad \Lambda\in F_\Pi, \tag{1.4}$$

where $|\Lambda| :=$ card Λ. The local Hamiltonians are defined by

$$Q^\Lambda := |\Lambda|\,Q(X_{1\Lambda},X_{2\Lambda},\dots,X_{n\Lambda}). \tag{1.5}$$

The local time evolution of the algebra $\mathscr{A}^\Lambda$ is given by its automorphism group τ^Λ determined in the usual way:

$$\tau_t^\Lambda(y) := \exp(itQ^\Lambda)\,y\,\exp(-itQ^\Lambda),\quad t\in\mathbb{R},\quad y\in\mathscr{A}^\Lambda. \tag{1.6}$$

We have shown in Ref. 20 in what sense a limiting global time evolution τ^Q exists for an arbitrary (not only polynomial) Q. It was shown in Corollary 4.4 of Ref. 20 that, roughly speaking, the net of local evolutions (1.6) leads to an automorphism group τ^Q of the "bare" algebra $\mathscr{A}$ iff Q is a linear function. Hence for a description of dynamics of the nontrivial mean-field-type interactions, one has to extend the algebra $\mathscr{A}$ by adjoining to it other algebraic elements, to include also the time transforms of its (original) elements. The minimal extension of $\mathscr{A}$ suitable to a description of τ^Q for a general Q is a C^*-algebra $\mathscr{C}$ isomorphic to the tensor product $\mathscr{A}\otimes\mathscr{N}$, where $\mathscr{N}$ is an abelian C^*-algebra of classical macroscopic quantities.

The description of the limiting evolution τ^Q in the ("physical", or "dressed") C^*-algebra $\mathscr{C}$ rests heavily on a certain ("sufficiently continuous" as well as "sufficiently discontinuous") action $\sigma(G)$ of a Lie group G on the "bare" algebra $\mathscr{A}$. A convenient choice of the group G—a group of "hidden symmetries" of the system—is given by its Lie algebra $\mathfrak{g}$ generated by the matrices X_j ($j=1,2,\dots,n$) in $\mathscr{L}(\mathscr{H}_0)$ occurring in (1.5). All considered macroscopic phenomena are expressed in terms of the group G (and of related mathematical objects).

The definition of kinematics of described systems together with a brief determination of the evolution $\tau^Q\subset{}^*$-aut $\mathscr{C}$ for a general $\mathbb{R}$-valued differentiable function Q defined on the dual $\mathfrak{g}^*$ of the Lie algebra[23] $\mathfrak{g}$ of G is reviewed and supplemented in Sec. II; that section also contains an investigation of the "Bogoliubov–Haag" Hamiltonians. Section III is devoted to general investigations of KMS states of such a general class of systems. Illustrative examples of Sec. IV include the strong coupling version of the BCS model, and also a corresponding model of the Josephson junction.

II. THE MATHEMATICAL STRUCTURE OF MEAN-FIELD THEORIES

We use here some concepts of global differential calculus on manifolds[24,25] to formulate connections of microscopic dynamics τ^Q with classical Hamiltonian dynamics generated by the Hamiltonian function Q. The theory of C^*-algebras[6,22,26] is extensively used in this paper. The final results are formulated, however, in a way suitable for an easy use in applications.

We shall start with a general setting, and later on we shall specify the formalism to the situations described in Sec. I.

A. Macroscopic observables and classical properties of states

Let G by an arbitrary connected Lie group, let $\mathfrak{g}$ be its Lie algebra and let $\mathfrak{g}^*$ be the dual space of $\mathfrak{g}$. Let Ad$^*(G)$ be the coadjoint representation[24,25,20] of G in $\mathfrak{g}^*$. Let $[\xi,\eta]\in\mathfrak{g}$ be the Lie bracket in $\mathfrak{g}$ ($\xi,\eta\in\mathfrak{g}$), and let f_ξ be the linear functions on the linear space $\mathfrak{g}^*$ defined by the relation:

$$f_\xi(F) := F(\xi),\quad \text{for all}\quad \xi\in\mathfrak{g},\quad \text{and all}\quad F\in\mathfrak{g}^*.$$

Here $F(\xi)\in\mathbb{R}$ is the value of the linear functional F taken on the vector ξ. The Poisson bracket $\{f,h\}$ is determined[20] for any differentiable functions f and h on an Ad*-invariant domain $E\subset\mathfrak{g}^*$ [i.e., Ad$^*(g)F\in E$ for all $F\in E$] by the formula

$$\{f_\xi,f_\eta\}(F) := -F([\xi,\eta]),\text{for all } F\in E,\xi,\eta\in\mathfrak{g}. \tag{2.1}$$

This bracket endows E with the Poisson structure[27] of a (generalized) classical phase space. It enables us to associate the (generalized classical Hamiltonian) dynamics[20] on E with any differentiable function Q on E.

Let $\mathscr{A}$ be a simple C^*-algebra with unit element $\mathbf{1}$. Assume that a strongly continuous representation $\sigma(G)\subset{}^*$-aut $\mathscr{A}$ is given. Each $\sigma_g := \sigma(g)\in\sigma(G)$ ($g\in G$) can be canonically extended to an automorphism of the double dual $\mathscr{A}^{**}$ to $\mathscr{A}$. The space $\mathscr{A}^{**}$ is considered here as a von Neumann algebra containing $\mathscr{A}$ as a weakly dense C^*-algebra in a canonical way.[28] The center $\mathscr{Z}$ of $\mathscr{A}^{**}$ is a $\sigma(G)$-invariant subset of $\mathscr{A}^{**}$. Although the center of $\mathscr{A}$ is trival, $\mathscr{Z}$ can be (and typically is) a rather huge commutative von Neumann algebra. The formulation of our assumption of "sufficient discontinuity" of the action of $\sigma(G)$ on $\mathscr{A}^{**}$ consists in an assumption of "sufficient nontriviality" of the action[29] of $\sigma(G)$ on $\mathscr{Z}$. Let us introduce first a basic mathematical object.[30]

Definition 2.1: Let $\mathscr{B}$ be the σ-algebra of all Borel subsets of $\mathfrak{g}^*$. Then a *G-measure* $E_\mathfrak{g}$ [or $\sigma(G)$*-measure* $E_\mathfrak{g}$] is a projection-valued measure $E_\mathfrak{g}$ on $\mathfrak{g}^*$ with values in $\mathscr{Z}$ which is G-equivariant. This means that $E_\mathfrak{g}$ is a σ-additive mapping from $\mathscr{B}$ into the Boolean lattice of projections in $\mathscr{Z}$ satisfying the relation

$$\sigma_g(E_\mathfrak{g}(B)) = E_\mathfrak{g}(\mathrm{Ad}^*(g)B), \text{ for all } g\in G, \text{ and all } B\in\mathscr{B}. \tag{2.2}$$

The G-measure $E_\mathfrak{g}$ is *nontrivial* if its support[20] $E := \mathrm{supp}\, E_\mathfrak{g} \subset g^*$ contains at least one two-point transitive set [i.e., there are two distinct $F^{(j)}\in E, j = 1,2$, connected by the transformation: $\mathrm{Ad}^*(g_{12})F^{(1)} = F^{(2)}$, for some $g_{12}\in G$]. Let $P_\mathfrak{g} := E_\mathfrak{g}(\mathfrak{g}^*) = E_\mathfrak{g}(E)$. ■

In the considered cases E will be compact. The above-mentioned (dis-)continuity of $\sigma(G)$ is now formulated in the following assumption.

Assumption 2.2: $\sigma(G) \subset {}^*\text{-aut}\, \mathscr{A}^{**}$ is such that there is at least one nontrivial G-measure $E_\mathfrak{g}$ in the considered system $(\mathscr{A},\sigma(G))$. ■

The albegra $\mathscr{C}$ of observables of $\sigma(G)$-mean-field theories[20] is isomorphic to the C^*-algebra $C(E,\mathscr{A})$ of $\mathscr{A}$-valued norm-continuous functions on the compact $E := \mathrm{supp}\, E_\mathfrak{g}$, which in turn is isomorphic to the C^*-tensor product $\mathscr{A} \otimes C(E)$. It can be embedded into $\mathscr{A}^{**}$ by the isomorphism $E_\mathfrak{g}: C(E,\mathscr{A}) \to \mathscr{A}^{**}$ expressed by the formula[20]

$$E_\mathfrak{g}: \hat{f} \mapsto E_\mathfrak{g}(\hat{f}) := \int \hat{f}(F) E_\mathfrak{g}(dF), \quad \hat{f}\in C(E,\mathscr{A}). \tag{2.3}$$

The algebras $\mathscr{A}$, resp. $C(E)$ are considered here as the subalgebras of $C(E,\mathscr{A})$ consisting of constant, resp. $\mathbb{C}$-valued functions. The image of $\mathscr{A}$ by (2.3) is denoted here again by $\mathscr{A}$ (it belongs, strictly speaking, to the subalgebra $P_\mathfrak{g}\mathscr{A}^{**}$), and $E_\mathfrak{g}(C(E)) \subset \mathscr{Z}$ will be denoted by $\mathscr{N}$. The embeddding of $\mathscr{C} = \mathscr{A} \otimes \mathscr{N}$ into $\mathscr{A}^{**}$ is useful due to the existence of canonical extensions of mappings defined on $\mathscr{A}$ to (normal) mappings on $\mathscr{A}^{**}$. We obtain in this way, e.g., the natural extension of $\sigma(G) \subset {}^*\text{-aut}\, \mathscr{A}$ to the group of autmosphisms of $\mathscr{C} = \mathscr{A} \otimes \mathscr{N}$ leaving the subalgebras $\mathscr{A}$ and $\mathscr{N}$ invariant; moreover, the group of automorphisms of $\mathscr{N}$ constructed in this way coincides with the natural action of $\mathrm{Ad}^*(G)$ on $C(E)$:

$$\sigma_g(E_\mathfrak{g}(f)) = E_\mathfrak{g}(f_g),$$

where

$$f_g(F) := f(\mathrm{Ad}^*(g^{-1})F).$$

We obtain also the canonical extensions to $\mathscr{C}$ of those states $\omega\in S(\mathscr{A})$ ($:=$ the set of all states on $\mathscr{A}$) the central covers[28] of which are majorized by $P_\mathfrak{g}$. "Natural" extensions to $\mathscr{C}$ of other states on $\mathscr{A}$ are nonunique, and they depend on chosen way of taking "the" thermodynamic limit.[31] Let us formulate now a connection between the states $\omega\in S(\mathscr{C})$ and the corresponding classical states μ_ω on $\mathscr{N}$.

Proposition 2.3: The restriction of any state ω on $\mathscr{C} := \mathscr{A} \otimes \mathscr{N}$ to the subalgebra $\mathscr{N} \sim C(E)$ is given by a probability measure μ_ω on E. If ω is a factor state, the measure μ_ω is concentrated at a point $F_\omega\in E$, i.e., $\mu_\omega = \delta_{F_\omega}$ $:=$ the corresponding Dirac measure. Each point $F_0\in E$ corresponds to some factor state ω on $\mathscr{C}$ with $F_\omega = F_0$. ■

Proof: Any state on $C(E)$ is described by a probability measure on E, by the Riesz–Markov theorem. Since $C(E)$ is the center of $\mathscr{C}$, the one-point support property of μ_ω for a factor state ω follows. The last assertion of the proposition is a consequency of the product form[32] of the pure states on $\mathscr{A} \otimes \mathscr{N}$: For any pure state $\omega_\mathscr{A}$ on $\mathscr{A}$ and any Dirac measure δ_{F_0} on E there is a pure (hence factor) state on $\mathscr{C}$ with restrictions to $\mathscr{A}$, resp. to $\mathscr{N}$, equal to $\omega_\mathscr{A}$, resp. to δ_{F_0}.
Q.E.D.

Let us consider now a quasilocal structure[33] on the "bare" algebra $\mathscr{A}$. Let $\mathscr{A} := \otimes_{p\in\Pi}\mathscr{A}^p$ be the C^*-algebra defined in Sec. I. Let $U(G)$ be a norm-continuous unitary representation of G in $\mathscr{H}_0$, and let $X_\xi\in\mathscr{A}^0$ ($\xi\in\mathfrak{g}$) be the generators of one-parameter groups $t\mapsto U(\exp(t\xi)) = \exp(-itX_\xi)$. Let $X_{\xi\Lambda}$ be defined according to (1.4),

$$X_{\xi\Lambda} := \frac{1}{|\Lambda|}\sum_{p\in\Lambda} \pi_p(X_\xi), \quad \xi\in\mathfrak{g}. \tag{2.4}$$

Then $X_{\xi\Lambda}\in\mathscr{A}$ ($\Lambda\in\mathscr{F}_\Pi$) is a uniformly bounded net. It was shown in the proof of Proposition 2.9 in Ref. 20 that there is an absorbing directed subset $\{\Lambda(j): j\in J\} \subset \mathscr{F}_\Pi$ such that all the subnets $\{X_{\xi\Lambda(j)}; j\in J\}$ ($\xi\in\mathfrak{g}$) converge in the w^*-topology of $\mathscr{A}^{**}$. This limit could be denoted by $X_{\xi\Pi}$ ($\in\mathscr{Z}$). Clearly, all the numerical nets $\{\omega(X_{\xi\Lambda(j)}): j\in J\}$ (with $\omega\in S(\mathscr{A})$, $\xi\in\mathfrak{g}$) are convergent:

$$F_\omega(\xi) := \lim_{j\in J} \omega(X_{\xi\Lambda(j)}) = \omega(X_{\xi\Pi}). \tag{2.5}$$

For states $\omega\in P_\mathfrak{g} S_*(\mathscr{A}^{**})$, with $P_\mathfrak{g} := E_\mathfrak{g}(\mathfrak{g}^*)$ as defined in Sec. II A of Ref. 20, and S_* denoting the set of normal states, the limits in (2.5) are independent of the choice of taking the thermodynamic limit.[20] The functions F_ω depend linearly on $\xi\in\mathfrak{g}$, since the association $\xi\mapsto X_\xi$ is a representation of the Lie algebra $\mathfrak{g}$. Hence $F_\omega\in\mathfrak{g}^*$. We shall consider $\omega\in P_\mathfrak{g} S_*(\mathscr{A}^{**})$ as a state on $\mathscr{C} \subset P_\mathfrak{g}\mathscr{A}^{**}$. One can choose different $P_\mathfrak{g}$ depending on the way of taking the thermodynamic limit. Let the minimal $P_\mathfrak{g}$ considered here be that one constructed in Sec. II A of Ref. 20, i.e., the maximal of such projections $P_\mathfrak{g}$ $\in\mathscr{Z}$ for which the limits for the nets (2.4) (with $\Lambda\in\mathscr{F}_\Pi$)

$$\text{s-}\lim_{\Lambda\nearrow\Pi} X_{\xi\Lambda}P_\mathfrak{g} = E_\mathfrak{g}(f_\xi)\in\mathscr{Z}, \quad \xi\in\mathfrak{g}, \tag{2.6}$$

exist. Let us deonte by conv $B \equiv \mathrm{conv}(B)$ the closed *convex hull* of $B \subset \mathfrak{g}^*$. We shall now describe the structure of $\mathrm{supp}\, E_\mathfrak{g}$.

Proposition 2.4: The subset of $E \subset \mathfrak{g}^*$ determined by the relation

$$E := \{F_\omega\in\mathfrak{g}^*: \ \omega\in S(\mathscr{C}),\ F_\omega(\xi) := \omega(E_\mathfrak{g}(f_\xi)) \text{ for } \xi\in\mathfrak{g}\} \tag{2.7}$$

is the convex hull of the set

$$E_0 := \{F_\psi\in\mathfrak{g}^*: \ F_\psi(\xi) = (\psi, X_\xi\psi), \ \xi\in\mathfrak{g}, \ \psi\in\mathscr{H}_0, \ \|\psi\| = 1\}.$$

The set E is compact. It is identical with $\mathrm{supp}\, E_\mathfrak{g}$:

$$\mathrm{supp}\, E_\mathfrak{g} = E = \mathrm{conv}\, E_0. \quad ■$$

Proof: For any basis $\{\xi_j : j = 1,2,...,n\}$ of $\mathfrak{g}$ the numbers $F_\omega(\xi_j)$ are coordinates of $F_\omega \in \mathfrak{g}^*$ in the dual basis. Supposed boundedness of the generators X_ξ implies boundedness of E. Convexity follows from the convexity in $\mathscr{C}^*$ (: = the topological dual to $\mathscr{C}$) of the set $S(\mathscr{C})$, as well as from the obvious convexity of the mapping $\omega \mapsto F_\omega$. Since $\mathscr{C}$ is unital, $1 \in \mathscr{C}$, the set $S(\mathscr{C})$ is w^*-compact[22] in $\mathscr{C}^*$. Due to its convexity and boundedness, the convex mapping $\omega \mapsto F_\omega$ is continuous.[34] But a continuous image of a compact set is compact. Hence E is convex and compact. Since the C^*-algebra $\mathscr{C}$ is antiliminal,[35] the set $\mathscr{E}S(\mathscr{C})$ of its pure states is w^*-dense in $S(\mathscr{C})$, and its (continuous) image is dense in E. The equality conv (supp $E_\mathfrak{g}$) $= E$ is valid due to the validity of

$$F_\omega(\xi) = \omega(\mathfrak{g}(f_\xi)) = \int F(\xi)\mu_\omega(dF).$$

To obtain any $F_\omega \in \text{supp } E_\mathfrak{g}$ it suffices to choose for μ_ω the corresponding Dirac measure. We can conclude now that also supp $E_\mathfrak{g}$ is convex.[30] Any product state ω_ψ on $\mathscr{A}$ given for an arbitrary normalized $\psi \in \mathscr{H}_0$ by

$$\omega_\psi(\pi_{p_1}(x_1) \otimes \pi_{p_2}(x_2) \otimes \cdots \otimes \pi_{p_m}(x_m))$$
$$:= \prod_{j=1}^{m} (\psi, x_j, \psi), \quad x_j \in \mathscr{L}(\mathscr{H}_0)$$

belongs to $P_\mathfrak{g} S(\mathscr{A})$. One has for any given $\xi \in \mathfrak{g}$, $\psi \in \mathscr{H}_0$, $\|\psi\| = 1$:

$$\omega_\psi(X_{\xi\Lambda}) = (\psi, X_\xi \psi), \quad \text{for all} \quad \Lambda \in \mathscr{F}_{\Pi}.$$

Since the extension of ω_ψ to the state on $\mathscr{C}$ given by (2.5) is pure, one has $E_0 \in \text{supp } E_\mathfrak{g}$ as a consequence of Proposition 2.3. The construction of the G-measure $E_\mathfrak{g}$ implies that the spectrum $\text{sp}(X_{\xi\Pi})$ of the self-adjoint element $X_{\xi\Pi} \in \mathscr{C}$ is $\text{sp}(X_{\xi\Pi}) = \{F(\xi) \in \mathbb{R} : F \in E\}$. But from the properties of spectra of functions of operators acting on a tensor product space, as well as from the convergence properties of the spectra of convergent nets of operators[36] one can conclude that

$$\text{sp}(X_{\xi\Pi}) = \text{conv sp}(X_\xi), \quad \xi \in \mathfrak{g}. \tag{2.8}$$

From the definition of F_ω in (2.5)–(2.7) we can see that E coincides with conv E_0.[37] Q.E.D.

Remark 2.5: Each state $\omega \in S(\mathscr{A})$ has many different extensions to states on $\mathscr{C}$. These extensions differ one from another, roughly speaking, by different possible distributions of values of macroscopic quantities.[38] We have associated classical states μ_ω with states $\omega \in S(\mathscr{C})$. If one chooses a unique "physically acceptable" extension of any state $\omega \in S(\mathscr{A})$ to a state $\omega \in S(\mathscr{C})$, than one can associate the corresponding classical state directly to the state $\omega \in S(\mathscr{A})$ on the "bare" C^*-algebra $\mathscr{A}$. We have seen in Proposition 2.9 of Ref. 20 that such a natural choice of the extensions of states is possible. ■

B. Dynamics of mean-field theories

The Poisson structure[27] on the compact $E := \text{supp } E_\mathfrak{g}$ introduced in (2.1) associates[20] with any differentiable function $Q \in C^\infty(E)$ on E the classical ("Poisson", or equivalently "generalized Hamilton") flow φ^Q on E. With $f_t := \varphi_t^{Q*} f := f \circ \varphi_t^Q$, the dynamics on E is expressed by Eq. (2.20) of Ref. 20. The φ^Q-invariance of the $\text{Ad}^*(G)$-orbits allows us to introduce[20] the differentiable cocycle $g_Q : \mathbb{R} \times E \to G$ such that the relation

$$\varphi_t^Q(F) = \text{Ad}^*(g_Q(t,F))F, \quad t \in \mathbb{R}, \quad F \in E, \tag{2.9}$$

is valid. The cocycle g_Q is given by a solution of an ordinary nonautonomous differential equation on the group manifold G, or by a time-dependent Schrödinger-like linear equation.[39] It is specified,[20] moreover, by an arbitrary function $\xi^0 : E \to \mathfrak{g}^*$, $F \mapsto \xi_F^0$, such that $\text{Ad}^*(\exp(t\xi_F^0))F \equiv F(\forall F \in E, \forall t \in \mathbb{R})$, cf. (2.14).

Now we can express the time evolution τ^Q in the system $(\mathscr{C}, \sigma(G))$ described in Sec. II A with a help of the cocycle g_Q. Let $\hat{f} \in C(E, \mathscr{A}) \sim \mathscr{C}$. We define then $\hat{f}_t \in C(E, \mathscr{A})$ by the formula[20]

$$\hat{f}_t(F) := \sigma(g_Q^{-1}(t,F))(\hat{f}(\varphi_t^Q F)), \quad t \in \mathbb{R}, \quad F \in E. \tag{2.10}$$

After the embedding of $C(E, \mathscr{A})$ into $P_\mathfrak{g} \mathscr{A}^{**}$ via the mapping $E_\mathfrak{g}$ from (2.3) we obtain the expression of time-evolved elements $\tau_t^Q(y)$ $(y \in \mathscr{C})$ in the form

$$\tau_t^Q(E_\mathfrak{g}(\hat{f})) = \int \hat{f}_t(F) E_\mathfrak{g}(dF). \tag{2.11}$$

The elements of $\mathscr{A}$ are identified with the constant functions in $C(E, \mathscr{A})$. We see that this subalgebra $\mathscr{A}$ of $\mathscr{C}$ is not invariant[40] with respect of τ^Q if some constants are transformed by (2.10) into some nontrivially varying functions, i.e., if the function

$$F \mapsto \sigma(g_Q^{-1}(t,F))(x)$$

depends nontrivially on $F \in E$ for some $x \in \mathscr{A}$ and for some fixed $t \in \mathbb{R}$. Let us formulate now some of the main results of Ref. 20.

Theorem 2.6: With the above definitions and notation, let $\sigma(G) \subset {}^*\text{-aut } \mathscr{A}$ be strongly continuous, and let $E = \text{supp } E_\mathfrak{g}$ be compact. Then

(i) τ^Q is a strongly continuous one-parameter group of $*$-automorphisms of $\mathscr{C}$.

(ii) Let $\delta_\sigma(\xi)$, $\xi \in \mathfrak{g}$, be the derivation of the one-parameter automorphism group $t \mapsto \sigma(\exp(t\xi))$ of $\mathscr{A}$. Let $\xi^0 \equiv 0$. The derivation δ_Q of τ^Q can be expressed by the formula

$$\delta_Q(E_\mathfrak{g}(\hat{f})) = \int \sum_{j=1}^{n} \left(\{Q, F_j\}(F) \frac{\partial \hat{f}(F)}{\partial F_j} - \frac{\partial Q(F)}{\partial F_j} \right.$$
$$\left. \times \delta_\sigma(\xi_j)(\hat{f}(F)) \right) E_\mathfrak{g}(dF), \tag{2.12}$$

for any "sufficiently differentiable"[41] $\hat{f} \in C(E, \mathscr{A}) \sim \mathscr{A} \otimes \mathscr{N}$.

(iii) The subalgebra $\mathscr{N} := E_\mathfrak{g}(C(E))$ is τ^Q-invariant, and the restriction of τ^Q to $\mathscr{N}$ coincides with the classical evolution φ^Q, i.e., $\tau_t^Q(E_\mathfrak{g}(f)) = E_\mathfrak{g}(\varphi_t^{Q*} f)$. ■

We can be more specific in the case if the quasilocal structure of $\mathscr{A}$ as described in Secs. I and II A is introduced. Let us quote[20,30] the following proposition.

Proposition 2.7: Let $\mathscr{A} = \bigotimes_{p \in \Pi} \mathscr{A}_p$, $U(G)$, and X_ξ be as above. Let $X^\Lambda(\xi) := |\Lambda| X_{\xi\Lambda}$, for $\xi \in \mathfrak{g}$. Then the subalgebras $\mathscr{C}^\Lambda := E_\mathfrak{g}(C(E, \mathscr{A}^\Lambda))$ are τ^Q-invariant. For any "sufficiently differentiable"[41] $\hat{f} \in C(E, \mathscr{A}^\Lambda)$, and for any $\Lambda \in \mathscr{F}_\Pi$, one has

$$\delta_Q(E_g(\hat f)) = \int \sum_{j=1}^{n} \Big(\{Q,F_j\}(F) \frac{\partial \hat f(F)}{\partial F_j} + i \frac{\partial Q(F)}{\partial F_j} \times [X^\wedge(\xi_j), \hat f(F)] \Big) E_g(dF), \qquad (2.13)$$

where the square bracket denotes the commutator in $\mathscr{A}^\wedge$. ∎

The time evolution of local perturbations of ω is unitarily implementable in the GNS representations $(\pi_\omega, \mathfrak{H}_\omega, \Omega_\omega)$ corresponding to the time invariant states ω, $\omega = \omega \circ \tau_t^Q$ $(t \in \mathbb{R})$.[1] This means that there is a self-adjoint operator Q_ω on $\mathfrak{H}_\omega$ such that $Q_\omega \Omega_\omega = 0$, and

$$\pi_\omega(\tau_t^Q(y)) = \exp(itQ_\omega)\pi_\omega(y)\exp(-itQ_\omega),$$

for all $t \in \mathbb{R}$, and for $y \in \mathscr{C}$. If ω is a ground state, then Q_ω can be interpreted as the energy operator in the corresponding "island of states" consisting of the local perturbations $\omega_y \in S(\mathscr{C})$ of ω, $\omega_y(x) := \omega(y^*xy)/\omega(y^*y)$ [for such $y \in \mathscr{C}$, for which $\omega(y^*y) \neq 0$]. There were extensive discussions in the past[15–19] on the form of Q_ω in equilibrium (hence time invariant[1]) states of mean-field interactions, and on the corresponding time evolution. To obtain some answers to this kind of questions in the framework of our formalism, we shall prove first the following lemma.

Lemma 2.8: Let $\omega \in S(\mathscr{C})$ be a factor state. Then the image $\pi_\omega(\mathscr{C})$ of the algebra $\mathscr{C} = \mathscr{A} \otimes \mathscr{N}$ at the GNS representation π_ω coincides with the image $\pi_\omega(\mathscr{A})$ of its subalgebra $\mathscr{A}$. Specifically,

$$\pi_\omega(E_g(\hat f)) = \pi_\omega(\hat f(F_\omega)),$$

where $\hat f(F) \in \mathscr{A}$ on the right-hand side is considered as an element of $P_g \mathscr{A}^{**}$. ∎

Proof: Let $F_\omega \in E$ be the point of the (generalized) classical phase space of the considered system to which the factor state ω is projected according to the Proposition 2.3. Then $\pi_\omega(E_g(\hat f_1)) = \pi_\omega(E_g(\hat f_2))$ iff $\hat f_1(F_\omega) = \hat f_2(F_\omega)$; this is because $\omega(E_g(B)) = 0$ for any Borel $B \subset E$ such that $F_\omega \notin B$, and due to (2.3), as well as due to the simplicity of $\mathscr{A}$. Hence each element $x := E_g(\hat f)$ with $\hat f(F) := yf(F)$ [$y \in \mathscr{A}$, $f \in C(E)$] is represented by $\pi_\omega(x) = f(F_\omega)\pi_\omega(y)$. The result is then obtained from the product structure of $\mathscr{C}_\bullet$. Q.E.D.

We shall derive now the form of the generator Q_ω of the mean-field theory in the GNS representation π_ω corresponding to a time invariant factor state ω. We shall use the following notation:

$$\xi_F^Q := \frac{d}{dt}\Big|_{t=0} g_Q(t,F) = \sum_{j=1}^{n} \frac{\partial Q(F)}{\partial F_j} \xi_j + \xi_F^0, \qquad (2.14)$$

where ξ_F^0 was introduced in Ref. 39.

Proposition 2.9:[42] Let $\omega \in S(\mathscr{C})$ be a τ^Q-invariant factor state. Let $F_\omega(\xi) := \omega(E_g(f_\xi))$. Then the generator Q_ω of τ^Q in the canonical cyclic (i.e., GNS) representation coincides with the (properly normalized) self-adjoint generator $X_\omega(\xi_{F_\omega}^Q)$ of the one-parameter group

$$t \mapsto \sigma_\omega(\exp(t\xi_{F_\omega}^Q))(\pi_\omega(x)) := \pi_\omega \circ \sigma(\exp(t\xi_{F_\omega}^Q))(x)$$

of *-automorphisms of $\pi_\omega(\mathscr{A}) = \pi_\omega(\mathscr{C})$. If all the one-parameter subgroups $t \mapsto \sigma(\exp(t\xi_j))$ are unitarily implementable with the self-adjoint generators $X_\omega(\xi_j)$, and if $\xi_F^0 \equiv 0$, the operator Q_ω can be written in the form

$$Q_\omega = \sum_{j=1}^{n} \frac{\partial Q(F_\omega)}{\partial F_j} X_\omega(\xi_j). \qquad (2.15)$$

If $\mathscr{A} := \otimes_{p \in \Pi} \mathscr{A}^p$ (hence it is quasilocal), and if τ^Q is defined as in (2.11), the restriction of the evolution $\pi_\omega \circ \tau^Q$ to any $\mathscr{C}^\wedge$ is generated by the element $Q_\omega^\wedge \in \pi_\omega(\mathscr{A})$ which can be expressed (for $\xi^0 \equiv 0$) in the form

$$Q_\omega^\wedge = \sum_{j=1}^{n} \frac{\partial Q(F_\omega)}{\partial F_j} \pi_\omega(X^\wedge(\xi_j)). \qquad (2.16)$$

Here $X^\wedge(\xi) := |\Lambda| X_{\xi\Lambda}$, as before. ∎

Proof: As ω is a factor state, we see from the proof of Proposition 2.3 that the central cover[28] $c(\pi_\omega)$ of the GNS representation π_ω is a minimal projection in $\mathscr{Z}$ majorized by any $E_g(B)$ with open $B \ni F_\omega$. The τ^Q-invariance of ω implies the time-invariance of $c(\pi_\omega)$ [with τ^Q canonically extended to an automorphism group of $P_g\mathscr{A}^{**} \supset c(\pi_\omega)\mathscr{C}$]. The representation π_ω is an isomorphism of the C^*-subalgebra $c(\pi_\omega)\mathscr{C}$ of $\mathscr{A}^{**}$. We have from the Lemma 2.8, for any $\hat f \in C(E,\mathscr{A})$,

$$\pi_\omega(E_g(\hat f)) = \pi_\omega(\hat f(F_\omega)).$$

The φ^Q-invariance of F_ω, as well as the formulas (2.10) and (2.11) give

$$\pi_\omega(\tau_t^Q(E_g(\hat f))) = \pi_\omega(\sigma(g_Q^{-1}(t,F_\omega))(\hat f(F_\omega))). \qquad (2.17)$$

The cocycle property[20] of g_Q implies the group property of $t \mapsto g_Q(t,F_\omega)$ for the φ^Q-invariant F_ω, i.e.,

$$g_Q(t_1 + t_2, F_\omega) = g_Q(t_1,F_\omega) g_Q(t_2,F_\omega), \quad \text{for all } t_1,t_2 \in \mathbb{R}.$$

Hence $g_Q(t,F_\omega) = \exp(\xi_{F_\omega}^Q)$ for some element $\xi_{F_\omega}^Q \in \mathfrak{g}$. One has then the invariance of ω with respect to the automorphism group $t \mapsto \sigma(\exp(-t\xi_{F_\omega}^Q))$ of $\mathscr{A}$, and the consequent unitary implementability[43] of this group in the representation π_ω, i.e.,

$$\pi_\omega(\sigma(\exp(-t\xi_F^Q))(\hat f(F_\omega))) = \exp(itX_\omega(\xi_{F_\omega}^Q))\pi_\omega(E_g(\hat f))\exp(-itX_\omega(\xi_{F_\omega}^Q)), \qquad (2.18)$$

for some self-adjoint operator $X_\omega(\xi_F^Q)$ on $\mathfrak{H}_\omega$ annihilating the cyclic vector Ω_ω. By comparison of (2.17) with (2.18) one obtains

$$\begin{aligned}\pi_\omega(\tau_t^Q(E_g(\hat f))) &= \exp(itQ_\omega)\pi_\omega(E_g(\hat f))\exp(-itQ_\omega) \\ &= \exp(itX_\omega(\xi_{F_\omega}^Q))\pi_\omega(E_g(\hat f)) \\ &\quad \times \exp(-itX_\omega(\xi_{F_\omega}^Q)), \quad \text{for all } t \in \mathbb{R}, \hat f \in C(E,\mathscr{A}).\end{aligned}$$

Due to the uniqueness[43] of the generator Q_ω, we have proved the first assertion of the proposition. The second assertion follows from (2.12) and the differential equation[39] for g_Q. The last assertion follows, with a help of previous considerations, from the $\sigma(G)$-invariance of local algebras $\mathscr{A}^\wedge$, and from the fact that the generators of the restrictions of $\sigma(G)$ to the local subalgebras themselves belong to these subalgebras.[42] Q.E.D.

Notes 2.10: (i) If a τ^Q invariant state ω has a "sufficiently nice" decomposition[44] to time-invariant factor states $\varkappa \in S(\mathscr{C})$ projected onto mutually distinct classical states concentrated at $F_\varkappa \in E$,

$$\omega(y) = \int \varkappa(y)\mu^{\omega}(d\varkappa), \quad y\in\mathscr{C}, \tag{2.19}$$

with some probability measure μ^{ω} on $S(\mathscr{C})$,[45] then the corresponding generator Q_{ω} is equal to an integral of the generators of the form (2.15). This can be seen also from (2.15), where the first term containing the Poisson bracket vanishes due to the φ^{Q}-invariance of points $F = F_{\varkappa}$ lying in the integration domain.[42]

(ii) The generator (2.16) [resp. also (2.15)] of the time evolution in any time invariant factor state [or also the integral of such generators constructed in more general GNS representations according to the preceding note (i)] is called here (in an agreement with the traditional terminology)[14,15,46] the *Bogoliubov–Haag Hamiltonian.* ■

III. EQUILIBRIUM STATES

The finite quantal systems were described traditionally by operators, density matrices, and Wigner symmetries[30] in a separable Hilbert space $\mathfrak{H}$. It is supposed in the traditional version[47] of quantal statistical mechanics[48] that there is given a self-adjoint operator H corresponding to the energy observable in the described system of a (large but) finite system of "elementary constituents" with the property that the operator $\exp(-\beta H)$ is of trace class, $\exp(-\beta H)\in\mathscr{T}(\mathfrak{H})$, for all $\beta > 0$, where β^{-1} is the temperature of the system. If this is the case, then the equilibrium state ω_{β} of the system is described by the density matrix ρ_{β} in the usual way:

$$\omega_{\beta}(x) := \mathrm{Tr}(\rho_{\beta}x), \quad \text{with} \quad \rho_{\beta} := Z_{\beta}^{-1}\exp(-\beta H), \quad x\in\mathscr{L}(\mathfrak{H}). \tag{3.1}$$

The thermostatics[1c] of the system is described then by the partition function Z_{β},

$$Z_{\beta} := \mathrm{Tr}(\exp(-\beta H)). \tag{3.2}$$

This scheme cannot be used in the description of infinite systems, for which the operator of total energy H cannot be defined. Also, for the generator H_{π} of time evolution in a representation π the operator $\exp(-\beta H_{\pi})$ is usually not of trace class. One overcomes this difficulty[1,6] by enclosing a finite part of the system described by a local algebra of observables $\mathscr{A}^{\Lambda}$ into a finite box Λ, constructing *local Gibbs states* ω_{β}^{Λ} according to (3.1) with H replaced by a suitable H^{Λ}, and taking a thermodynamic limit $\Lambda\nearrow\Pi$. To the best of the present author's knowledge, the obtained limiting *Gibbs states* ω_{β} satisfy the *KMS condition* in all the cases in which it was possible to obtain a definite result. This condition is formulated for the total (infinite) system if its global time evolution is known.[49]

Definition 3.1: Let $\mathscr{C}$ be a unital C^*-algebra, and let τ be a strongly continuous one-parameter subgroup of the *-automorphism group of $\mathscr{C}$. Let $\mathscr{C}_{\tau}$ be the set of all entire analytic elements of $\mathscr{C}$, i.e., $y\in\mathscr{C}_{\tau}\Leftrightarrow$the function $z\mapsto\tau_z(y)$ is entire analytic. Let δ_{τ} be the generator (derivation)[6] of τ, and let $D(\delta_{\tau})\subset\mathscr{C}$ be its domain. Then the state $\omega\in S(\mathscr{C})$ is called (i) a *τ-KMS state at the inverse temperature β* (briefly a *(τ,β)-KMS state*) if there is a norm-dense τ-invariant *-subalgebra $\mathscr{C}_{\tau}^{0}$ of $\mathscr{C}_{\tau}$ such that for all $x,y\in\mathscr{C}_{\tau}^{0}$,

$$\omega(x\tau_{i\beta}(y)) = \omega(yx); \tag{3.3}$$

or (ii) a *τ-ground state* (or τ-β KMS state at $\beta = +\infty$), if

$$-i\omega(y^*\delta_{\tau}(y))\geqslant 0 \quad \text{for all} \quad y\in D(\delta_{\tau}). \qquad ■(3.4)$$

One can easily check that the relation (3.3) is fulfilled by the states ω_{β} of traditional quantum statistics defined in (3.1), if the time evolution is described by the automorphism group τ of $\mathscr{L}(\mathfrak{H})$ generated by H:

$$\tau_t(y) := \exp(itH)y\exp(-itH), \quad y\in\mathscr{L}(\mathfrak{H}), \quad t\in\mathbb{R}. \tag{3.5}$$

Let $\omega\in S(\mathscr{C})$ be a τ-invariant state for a strongly continuous one-parameter group $\tau\subset{}^*$-aut $\mathscr{C}$. Let H_{ω} be the canonically defined[43] generator of τ in the cyclic GNS representation $(\pi_{\omega},\mathfrak{H}_{\omega},\Omega_{\omega})$, i.e.,

$$\pi_{\omega}(\tau_t(x)) = \exp(itH_{\omega})\pi_{\omega}(x)\exp(-itH_{\omega}). \tag{3.6}$$

The following lemma is then valid.[50]

Lemma 3.2: In the conditions of the Definition 3.1, the state $\omega\in S(\mathscr{C})$ is a ground state iff ω is τ-invariant, and $H_{\omega}\geqslant 0$. For such a state we have $\exp(itH_{\omega})\in\pi_{\omega}(\mathscr{C})''$ = the strong operator closure[51–53] of the set $\pi_{\omega}(\mathscr{C})\subset\mathscr{L}(\mathfrak{H}_{\omega})$. ■

Some of the important properties of the general KMS states are collected[54] in the following proposition.

Proposition 3.3: Let $K_{\beta}\subset S(\mathscr{C})$ $(0<\beta\leqslant\infty)$ be the sets of (τ,β)-KMS states for a strongly continuous automorphism group $\tau\subset{}^*$-aut $\mathscr{C}$. Then

(i) Each state $\omega\in K_{\beta}$ is τ-invariant, i.e., $\omega\circ\tau_t = \omega\,(t\in\mathbb{R})$.

(ii) Each set K_{β} is a convex w^*-compact subset of $S(\mathscr{C})$.

(iiia) For $\beta\neq\infty$, K_{β} is a simplex[55] in $S(\mathscr{C})$.

(iiib) K_{∞} is a face[56] in $S(\mathscr{C})$.

(iva) The set $\mathscr{E}K_{\beta}$ of extremal points $\omega\in K_{\beta}$ $(\beta\neq\infty)$ consists of factor states: The centers of $\pi_{\omega}(\mathscr{C})''$ are trivial for $\omega\in\mathscr{E}K_{\beta}$.

(ivb) The extremal points $\omega\in\mathscr{E}K_{\infty}$ are pure states, i.e., $\pi_{\omega}(\mathscr{C})'' = \mathscr{L}(\mathfrak{H}_{\omega})$.

(v) For $\omega_j\in\mathscr{E}K_{\beta}$ $(\beta\neq\infty, j = 1,2)$ one has either $\omega_1 = \omega_2$, or $\omega_1 \,\delta\, \omega_2$, i.e., any two distinct extremal KMS states are disjoint. This means that the central covers of their GNS representations are mutually orthogonal.

(vi) The extremal decomposition of $\omega\in K_{\beta}$ $(\beta\neq\infty)$ coincides with its central decomposition.[57] ■

Now we shall study the specific properties of the (τ^{Q},β)-KMS states of our "C^*-dynamical system"[6] $(\mathscr{C},\tau^{Q})$. We shall consider from the beginning the specific quasilocal structure of the "physical" algebra of observables

$$\mathscr{C} = \mathscr{A}\otimes\mathscr{N}\sim C(E,\mathscr{A}), \quad \text{with} \quad \mathscr{A} = \bigotimes_{p\in\Pi}\mathscr{A}_p, \text{ and } \mathscr{N}$$

$\sim C(E)$. The time evolution τ^{Q} is determined according to (2.11). First we shall prove several auxilary assertions. It is assumed that all the traces in the forthcoming formulas exist. This happens, e.g., if the considered Hilbert space $\mathfrak{H}$ is finite dimensional.

Lemma 3.4: Let H and X be self-adjoint operators on the Hilbert space $\mathfrak{H}$. Let $[X,Y]$ denote the commutator of operators X and Y. Then, for any Borel function f:sp$(H)\to\mathbb{R}$, one has

$$\mathrm{Tr}(f(H)[H,X]) = 0. \qquad ■$$

Proof: From the commutativity of $f(H)$ with H, and from the invariance of the traces with respect to cyclic permutations of operators in products lying in the argument of a trace, we obtain

$$\mathrm{Tr}(f(H)HX) = \mathrm{Tr}(f(H)XH),$$

which gives the result. Q.E.D.

Lemma 3.5: Let X_ξ ($\xi \in \mathfrak{g}$) be the self-adjoint generators[58] of a strongly continuous representation $U(G)$ of the Lie group G in $\mathfrak{H}$. Let us choose an $\eta \in \mathfrak{g}$. Let f be a real-valued Borel function on the spectrum $\mathrm{sp}(X_\eta)$. Let $F^f_\eta \in \mathfrak{g}^*$ be defined by the relation

$$F^f_\eta(\xi) := \mathrm{Tr}(f(X_\eta)X_\xi), \quad \text{for all} \quad \xi \in \mathfrak{g}.$$

Then F^f_η is a stationary point of the one-parameter group $\{\mathrm{Ad}^*(\exp(t\eta)) : t \in \mathbb{R}\}$ of transformations of $\mathfrak{g}^*$. ■

Proof: We have to prove $F^f_\eta([\eta,\xi]) = 0$ $(\forall \xi \in \mathfrak{g})$.[25] Since $X_{[\eta,\xi]} = -i[X_\eta, X_\xi]$, we have to prove $\mathrm{Tr}(f(X_\eta)[X_\eta, X_\xi]) = 0$ $(\forall \xi \in \mathfrak{g})$. The last equation is valid due to Lemma 3.4. Q.E.D.

The following proposition shows that the Gibbs state ω_β from (3.1) determined by the Hamiltonian $H := Q_\omega$ (acting in a Hilbert space $\mathfrak{H}$) given by the formula (2.15) determines such expectation values $F_\omega(\xi)$ ($\xi \in \mathfrak{g}$) of observables X_ξ that are automatically stationary with respect to the corresponding classical Hamiltonian time evolution φ^Q of the "mean-field" $F \in \mathfrak{g}^*$.

Proposition 3.6: Let $\{\xi_j : j = 1,2,...,n\}$ be an arbitrary basis in $\mathfrak{g}$. Let $Q \in C^\infty(E,\mathbb{R})$. The point $\hat{F} \in \mathfrak{g}^*$ satisfying the "consistency condition"

$$\mathrm{Tr}\left[\exp\left(-\beta \sum_{j=1}^n \frac{\partial Q(\hat{F})}{\partial F_j} X_{\xi_j}\right) X_\xi\right] = \hat{F}(\xi)\mathrm{Tr}\left[\exp\left(-\beta \sum_{j=1}^n \frac{\partial Q(\hat{F})}{\partial F_j} X_{\xi_j}\right)\right],$$

for all $\xi \in \mathfrak{g}$, is a stationary point of the classical flow φ^Q on $\mathfrak{g}^*$, i.e., $\varphi^Q_t(\hat{F}) = \hat{F}$, for all $t \in \mathbb{R}$. ■

Proof: Let us set in the Lemma 3.5

$$\eta := \sum_{j=1}^n \frac{\partial Q(\hat{F})}{\partial F_j} \xi_j,$$

and $f(x) := \text{const} \exp(-\beta x)$ with const: $= [\mathrm{Tr}[\exp(-\beta X_\eta)]]^{-1}$. Then $F^f_\eta = \hat{F}$ is a stationary point of $t \mapsto \mathrm{Ad}^*(\exp(t\eta))$, i.e., $\hat{F}([\eta,\xi]) = 0$, for all $\xi \in \mathfrak{g}$. We have $\eta = d_{\hat{F}}Q$. The definition (2.1) of the Poisson structure on $\mathfrak{g}^*$ then implies[59]

$$\{Q, f_\xi\}(\hat{F}) = 0, \quad \text{for all} \quad \xi = \mathfrak{g},$$

hence also

$$\left.\frac{d}{dt}\right|_{t=0} f(\varphi^Q_t(\hat{F})) = 0, \quad \text{for any } f \in C^\infty(\mathfrak{g}^*).$$

This implies stationarity of $\hat{F}$, since the Hamiltonian equations are differential equations of the first order. Q.E.D.

The following lemma takes into account the quasilocal structure of $\mathscr{A}$ (resp. $\mathscr{C}$). We shall consider the set Π to be endowed with the structure of a commutative noncompact discrete group, e.g, Π could be the ν-dimensional lattice $\mathbb{Z}^\nu$. This is not any serious restriction of generality, but it enables us to take formulations of some assertions directly from the literature. Let the group multiplication in Π be denoted by "$+$", and let $0 \in \Pi$ be the unit element. Then we shall consider the algebra $\mathscr{A}^0 = \mathscr{L}(\mathscr{H}_0)$ as a subalgebra $\mathscr{A}^p$ (with $p = 0$) of $\mathscr{A} = \otimes_{p\in\Pi} \mathscr{A}^p$, and the mappings π_p from (1.1) will be considered as a translation automorphisms of $\mathscr{A}$ in a natural way: For $x \in \mathscr{A}^q$, we define $\pi_p(x) := \pi_{p+q} \circ \pi_q^{-1}(x)$, for $p,q \in \Pi$.

Lemma 3.7: Let $\mathscr{A}$ be the quasilocal algebra with the quasilocal structure generated by an action π_Π of countable discrete abelian group Π, as described above. Let $\omega \in S(\mathscr{A})$ be a π_Π-invariant factor state. Then ω is weakly π_Π-clustering,[1b] hence

$$\lim_{p\to\infty} \omega(\pi_p(x)y) = \omega(x)\omega(y), \quad \text{for all} \quad x,y \in \mathscr{A}.$$

Conversely, any π_Π-clustering state on $\mathscr{A}$ is a factorial state. ■

Proof: This is a corollary to the Proposition 2.3 of Ref. 33a. ■

A connection between expectations of mutually corresponding local and global quantities in equilibrium states gives the following almost trivial lemma.

Lemma 3.8: Let $\omega = S(\mathscr{A})$ be a π_Π-invariant product state,

$$\omega := \otimes_{p\in\Pi} \omega^p, \quad \text{with} \quad \omega^p := \omega^0 \circ \pi_p^{-1}, \quad \text{for} \quad p \in \Pi, \tag{3.7}$$

where $\omega^p \in S(\mathscr{A}^p)$. Let $\omega \in S(P_\mathfrak{g}\mathscr{A}^{**})$ denote also the normal extension of ω. Let $F_\omega(\xi) := \omega(E_\mathfrak{g}(f_\xi))$ for $\xi \in \mathfrak{g}$. Then $\omega^0(X_\xi) = F_\omega(\xi)$. ■

Proof: After insertion of (2.4) into the argument of ω, we obtain $\omega(X_{\xi\Lambda}) = \omega^0(X_\xi)$ for all $\Lambda \in \mathscr{F}_\Pi$. Since the central cover $c(\pi_\omega) \leqslant P_\mathfrak{g}$, due to permutation invariance of ω,[60] the result is obtained from (2.6). Q.E.D.

Now we are going to formulate our main theorem concerning the structure of the KMS states of the considered mean field theories. Due to the separability of the C^*-algebra $\mathscr{C}$, each KMS state is supported (and not only pseudosupported)[61] by extremal KMS states of the considered strongly continuous time evolution at the same temperature. Hence it is sufficient to consider extremal KMS states of our systems. Any other KMS state is expressible by a w^*-convergent integral of extremal KMS states determined by some regular Borel measure on $S(\mathscr{C})$ supported by the Borel set of the extremal KMS states.[62]

Theorem 3.9:[63] Let us consider the C^*-dynamical system $(\mathscr{C}, \tau_Q)$ with $\mathscr{C} = \mathscr{A} \otimes \mathscr{N}$, and the quasilocal structure defined by the action π_Π of the abelian group Π, as it was introduced above. Let $\omega \in S(\mathscr{C})$ be a state on $\mathscr{C}$ satisfying (2.5). Then the following two statements are equivalent:

(i) The state $\omega \in S(\mathscr{C})$ is an extremal KMS state at positive temperature $0 < \beta < \infty$.

(ii) The state ω is a π_Π-invariant product state

$$\omega = \otimes_{p\in\Pi} (\omega^0 \circ \pi_p^{-1}),$$

where ω^0 is the unique (σ_η, β)-KMS state on $\mathscr{A}^0 := \mathscr{L}(\mathscr{H}_0)$ corresponding to the one-parameter automorphism group $\sigma_\eta := \{\sigma_\eta(t) := \sigma(\exp(-t\eta)) : t \in \mathbb{R}\}$ of

$\mathscr{A}^0$, with $\eta := \xi^Q_{F_\omega}$, and this $F_\omega \in \mathfrak{g}^*$ satisfies the "consistency condition":

$$\omega(E_\mathfrak{g}(f_\xi)) = F_\omega(\xi), \quad \text{for} \quad \text{all} \quad \xi \in \mathfrak{g}. \qquad \blacksquare$$

Proof: (i) implies that ω is a τ^Q-invariant factor state. According to the Proposition 2.9, the time evolution of $\pi_\omega(\mathscr{C}) = \pi_\omega(\mathscr{A})$ induced by the group $\tau^Q \subset *$-aut $\mathscr{C}$ coincides with the evolution induced by the group $\sigma_\eta \subset \sigma(G)$ determined as in (ii). The "consistency condition" is satisfied here trivially—from the definition of F_ω entering in the determination of the generator $\eta \in \mathfrak{g}$. Hence ω is a (σ_η, β)-KMS state on $\mathscr{C}$. The restriction ω^0 of ω to the subalgebra $\mathscr{A}^0$ is also a (σ_η, β)-KMS state, since $\mathscr{A}^0 := \mathscr{L}(\mathscr{H}_0)$ is $\sigma(G)$-invariant. The cyclic vector Ω_ω is separating for $\pi_\omega(\mathscr{A})''$,[28] so that ω^0 is a faithful normal state on $\mathscr{A}^0$. Hence[61] ω^0 is the unique (σ_η, β)-KMS state on $\mathscr{L}(\mathscr{H}_0)$ determined by (3.1) with $H := X_\eta$, cf. Proposition 2.9 [with $Q_\omega := X_\omega(\eta) = X_\eta + \text{const}_\eta$, if the corresponding transformations are restricted to the subalgebra $\mathscr{A}^0$, and the additive scalar valued constant "const_η" might be inserted for a convenient renormalization]. The same considerations are applicable to the restriction ω^Λ of ω to the subalgebra $\mathscr{A}^\Lambda = \mathscr{L}(\otimes_{p \in \Lambda} \mathscr{H}_p)$ for any finite Λ: The state ω^Λ is the unique (σ_η, β)-KMS state on $\mathscr{A}^\Lambda$ determined by the Hamiltonian (2.16).

We have to prove that ω is π_Π-invariant product state on $\mathscr{A}$ (which has a unique "physically acceptable" extension to the "physical algebra" $\mathscr{C}$, coinciding with the given $\omega \in P_\mathfrak{g} \mathscr{A}^{**}$). Let $\tilde{\omega}^\Lambda := \otimes_{p \in \Lambda} \omega^0 \circ \pi_p^{-1} \in S(\mathscr{A}^\Lambda)$ be the product state determined by the (σ_η, β)-KMS state ω^0 on $\mathscr{A}^0$. Then $\tilde{\omega}^\Lambda$ also satisfies the (σ_η, β)-KMS property [cf. Eq. (3.3)], hence it coincides with the unique (σ_η, β)-KMS state ω^Λ on $\mathscr{A}^\Lambda$. The state $\omega \in S(\mathscr{A})$ is uniquely determined, however, by its restrictions to all local subalgebras $\mathscr{A}^\Lambda$. This proves the product property, as well as the π_Π-invariance of the extremal KMS state $\omega \in S(\mathscr{C})$. So we have proved the implication (i)$\Rightarrow$(ii).

Let ω be given by (ii). We have just proved that this product state ω satisfying the "consistency condition" (specifying also the values of $\eta := \xi^Q_{F_\omega}$) is a unique extremal (σ_η, β)-KMS state on $\mathscr{A}$. The Lemma 3.8 and Lemma 3.6 imply time invariance of $F_\omega = \varphi^Q_t(F_\omega) \in \mathfrak{g}^*$, hence also the equalities $\{Q, f\}(F_\omega) = 0$ [for all $f \in C^\infty(E)$]. Factoriality of $\omega \in S(\mathscr{A})$ follows from its product property,[33] cf. Lemma 3.7. The comparison of the expression (2.13) for the derivation δ_Q of the time evolution τ^Q taken in the representation π_ω with the generator of σ_η (with $\eta := \xi^Q_F$) expressed by a help of (2.14), the factoriality of ω, and the constancy $F_\omega = \varphi^Q_t F_\omega$ show that ω is also a (τ^Q, β)-KMS state on $\mathscr{C}$. The extremality follows now from the Proposition 3.3(vi). Q.E.D.

In proving the theorem we have also proved the following results.

Proposition 3.10: (i) There is exactly one (σ_η, β)-KMS state of the system $(\mathscr{C}, \sigma(G))$, for each $\beta \in \mathbb{R}$, and for each $\eta \in \mathfrak{g}$. It is the product state of identical copies of the Gibbs state ω^p_β on subalgebras $\mathscr{A}^p$ $(p \in \Pi)$. The Gibbs state ω^0_β on $\mathscr{L}(\mathscr{H}_0)$ is determined by the Hamiltonian $H := X_\eta$ according to the formula (3.1). The points $F^\eta_\beta \in \mathfrak{g}^*$ determined by the "consistency condition" $F^\eta_\beta(\xi) := \omega^0_\beta(X_\xi)$ are stationary points of the Poisson (i.e., generalized Hamiltonian) flow φ^η on $\mathfrak{g}^*$ generated by the classical Hamiltonian $f_\eta \in C^\infty(E) : \varphi^\eta_t F := \text{Ad}^*(\exp(t\eta))F$; $F^\eta_\beta \equiv F^{\beta\eta}_1$.

(ii) For any stationary point $\hat{F} \in E$ of the classical flow φ^Q, there is at most one extremal (τ^Q, β)-KMS state of the system $(\mathscr{C}, \tau^Q)$. It exists iff the Gibbs state ω^0_β on $\mathscr{A}^0$ corresponding to the quantum-mechanical Hamiltonian X_η with

$$\eta := \sum_{j=1}^{n} \frac{\partial Q(\hat{F})}{\partial F_j} \xi_j \tag{3.8}$$

satisfies the "consistency condition":

$$\omega^0_\beta(X_\xi) = \hat{F}(\xi), \quad \text{for} \quad \text{all} \quad \xi \in \mathfrak{g}. \tag{3.9}$$

(iii) The set of all τ^Q-KMS states on $\mathscr{C}$ is independent of the choice of "physically acceptable" extensions $\omega \in S(\mathscr{C})$ of all states $\omega \in S(\mathscr{A})$ according to (2.5)—see also Proposition 2.9 of Ref. 20. $\blacksquare$

Now we can give an answer to the quenstion on existence of KMS states for any system of the considered class of mean-field theories. Let us introduce the mappings $\Psi_\beta : \mathfrak{g} \to \mathfrak{g}^*$ $(\beta \in \mathbb{R})$, $\Psi_\beta : \eta \mapsto \Psi_\beta(\eta) := F^\eta_\beta \in \mathfrak{g}^*$, such that for $X(\xi) := X_\xi$ one sets

$$\begin{aligned} F^\eta_\beta(\xi) &: \\ &= [\text{Tr} \exp(-\beta X(\eta))]^{-1} \text{Tr}[\exp(-\beta X(\eta)) X(\xi)]. \end{aligned} \tag{3.10}$$

The mapping Ψ_β is continuous, since the association $\eta \mapsto X(\eta)$ is bounded linear [we consider here finite dimensional representations $U(G)$ only]. The point $F^\eta_\beta \in \mathfrak{g}^*$ corresponds to the unique β-KMS state of the automorphism group $t \mapsto \sigma(\exp(t\eta)) \equiv \sigma_\eta(t)$ of $\mathscr{C}$. The ranges of the mappings Ψ_β are contained in the convex compact set $E := \text{conv}\, E_0 = \text{supp}\, E_\mathfrak{g}$, cf. Proposition 2.4.

The mapping $\xi^Q : F \mapsto \xi^Q_F \in \mathfrak{g} \equiv \mathfrak{g}^{**} \equiv T^*_F \mathfrak{g}^*$ is a continuous mapping from E into $\mathfrak{g}$, cf. (2.14). The existence of KMS states is then assured by the following proposition.

Proposition 3.11: Let $U(G)$ be a finite dimensional representation of any given compact connected Lie group G, $Q \in C^\infty(\mathfrak{g}^*, \mathbb{R})$, and let $(\mathscr{C}, \tau^Q)$ be a corresponding mean-field system (with supp $E_\mathfrak{g} = \text{conv}\, E_0$, as in Proposition 2.4). Then there is at least one (τ^Q, β)-KMS state of this system for any $\beta \in \mathbb{R}$. $\blacksquare$

Proof: The composed mapping $\Psi_\beta \circ \xi^Q : F \mapsto \Psi_\beta(\xi^Q_F) \in \mathfrak{g}^*$ is a continuous mapping of a convex compact subset E of $\mathfrak{g}^*$ into itself, hence it has a fixed point $\hat{F}$, due to the Schauder fixed point theorem (cf. Ref. 24c, p. 565). This means that $\hat{F} \in E$ satisfies the "consistency equation"

$$\hat{F} = \Psi_\beta(\xi^Q_{\hat{F}}). \tag{3.11}$$

We can construct now a permutation invariant product state ω according to (3.7), with

$$\omega^0(y) := [\text{Tr} \exp(-\beta X(\xi^Q_{\hat{F}}))]^{-1} \text{Tr}[\exp(-\beta X(\xi^Q_{\hat{F}})) y],$$

for $y \in \mathscr{A}^0 \equiv \mathscr{L}(\mathscr{H}_0)$. Then one can use Lemma 3.8 to see from Theorem 3.9 that ω is the desired (τ^Q, β)-KMS state. Q.E.D.

From the proof of this proposition one can see the validity of the following assertion.

Corollary 3.12: Let G be a connected compact Lie group, and $U(G)$ its finite dimensional unitary representation. Let $E:=\operatorname{conv} E_0$, as in Proposition 2.4. Assume that E is endowed with the canonical Poisson structure of the space $\mathfrak{g}^*$. Then the (generalized) Hamiltonian flow φ^Q on E corresponding to an arbitrary $Q\in C^\infty(E,\mathbb{R})$ has a fixed point $\hat{F}\in E$, $\varphi^Q_t\hat{F}=\hat{F}(t\in\mathbb{R})$, with $\hat{F}$ satisfying the "consistency condition" (3.11). ■

We cannot obtain such a complete and general information on the τ^Q-ground states (i.e., the KMS states at $\beta=\infty$), since there is no uniqueness theorem on ground states for general Hamiltonians for "local" subsystems. We can, however, formulate the following assertion.

Proposition 3.13: Let the C^*-dynamical system $(\mathscr{C}:=\mathscr{A}\otimes\mathscr{N},\tau^Q;\sigma(G),\pi_\Pi)$ be given as above. Let $\omega\in S(\mathscr{C})$ be determined from the product state $\omega:=\otimes_{p\in\Pi}\omega^p\in S(\mathscr{A})$ by the natural extension (2.5). Let the F_ω given by (2.5) be φ^Q-invariant, and let all the ω^p $(p\in\Pi)$ be ground states on $\mathscr{A}^p$ with respect to the restrictions of σ_η $(\eta:=\xi^Q_{F_\omega})$ to $\mathscr{A}^p$. Then ω is a factorial ground state on $\mathscr{C}$ for the evolution τ^Q. If all the $\omega^p(p\in\Pi)$ are pure then ω is an extremal τ^Q-ground state. ■

Proof: The factoriality of ω is a consequence of the cluster properties according to Lemma 3.7. The condition (3.4) is fulfilled for $\tau:=\sigma_\eta$ due to the factoriality of ω, due to the $\mathrm{Ad}^*(\exp(t\eta))$-invariance of F_ω, and due to the trivial fulfillment of (3.4) by local elements $y\in\mathscr{A}^\Lambda$ $(\Lambda\in\mathscr{F}_\Pi)$. An application of the Proposition 2.9 proves the fullfillment of (3.4) also for $\tau:=\tau^Q$. The remaining assertion is valid, since the restriction of the product of pure local states ω^p to any local subalgebra $\mathscr{A}^\Lambda$ is a pure state ω^Λ (hence it cannot be a convex combination of two mutually different states). Q.E.D.

Now we shall consider the question on existence of ground states. Let $\Omega_\xi\in\mathscr{H}_0$ be an eigenvector of $X(\xi)$ corresponding to the minimal eigenvalue. Then the permutation invariant product state $\omega_\xi\in S(\mathscr{C})$ constructed from ω^0_ξ according to (3.7), with

$$\omega^0_\xi(y):=(\Omega_\xi,y\Omega_\xi),\quad y\in\mathscr{L}(\mathscr{H}_0),$$

is a ground state for the system $(\mathscr{C},\sigma_\xi)$. Let us define the mapping

$$\Psi_\infty:\mathfrak{g}\to\mathfrak{g}^*,\quad \xi\mapsto F^\xi_\infty:=\Psi_\infty(\xi),$$
$$F^\xi_\infty(\eta):=(\Omega_\xi,X(\eta)\Omega_\xi),\quad \eta\in\mathfrak{g}.$$

Let the mapping $\xi^Q:\mathfrak{g}^*\to\mathfrak{g}$, $F\mapsto\xi^Q_F$ be defined according to (2.14).

Proposition 3.14: Let the above defined mapping Ψ_∞ can be chosen continuous for the considered representation $U(G)$ (with $E:=\operatorname{supp} E_\mathfrak{g}$ convex compact). Then the system $(\mathscr{C},\tau^Q)$ has a ground state. ■

Proof: The composed mapping $\Psi_\infty\circ\xi^Q$ of the two continuous mappings Ψ_∞ and ξ^Q is a continuous mapping of the convex compact set E into itself, cf. Proposition 2.4. Then an application of the Schauder fixed point theorem, and the use of Proposition 3.13 gives the result (cf. also the proof of Proposition 3.11). Q.E.D.

Let us close this section with some easily derivable results on symmetries of the considered systems.[64]

Remarks 3.15: (i) $\gamma\in{}^*$-aut $\mathscr{C}$ iff there is a homeomorphism $\varphi_\gamma:E\to E$, and a function $\hat{\gamma}:E\to{}^*$-aut $\mathscr{A}$, $F\mapsto\hat{\gamma}_F$, with $\hat{\gamma}$ strongly continuous, such that for all $\hat{f}\in C(E,\mathscr{A})$ one has

$$\gamma(\hat{f})(F)=\hat{\gamma}_F(\hat{f}(\varphi_\gamma F)),\quad\text{for all}\quad F\in E.\tag{3.12}$$

(ii) Let $\mathscr{C}:=C(E,\mathscr{A})$, $E_\mathfrak{g}(\mathscr{C})\subset P_\mathfrak{g}\mathscr{A}^{**}$, and let $E_\mathfrak{g}\circ\gamma\circ E_\mathfrak{g}^{-1}$ be extendable to an *-aut $P_\mathfrak{g}\mathscr{A}^{**}$. Let $\hat{\gamma}$ and φ_γ be defined according to (3.12). Then φ_γ is uniquely determined by $\hat{\gamma}$. This connection of $\hat{\gamma}$ and φ_γ is expressed by the formula:

$$E_\mathfrak{g}(f_\xi\circ\varphi_\gamma)=s^*-\lim_\Lambda\int\hat{\gamma}_F(X_{\xi\Lambda})E_\mathfrak{g}(dF),\forall\xi\in\mathfrak{g}.\tag{3.13}$$

(iii) For $\gamma:=\sigma_g$ (some fixed $g\in G$) one has $\varphi_\gamma=\mathrm{Ad}^*(g^{-1})$.

(iv) If $\gamma\circ\tau^Q_t\equiv\tau^Q_t\circ\gamma$, then $\varphi_\gamma\circ\varphi^Q_t\equiv\varphi^Q_t\circ\varphi_\gamma$. If φ_γ is, moreover, a Poisson automorphisms[27] of E, then $\varphi^*_\gamma Q=Q+Q^0$, where Q^0 is constant on the Ad^* orbits of G in E.

(v) If $Q\circ\mathrm{Ad}^*(g)\equiv Q$, and $\hat{F}$ satisfies the consistency condition (3.9), then (3.9) is satisfied also by $\mathrm{Ad}^*(g)\hat{F}$ inserted for $\hat{F}$. ■

IV. SOME HINTS FOR APPLICATIONS

Let us consider the system determined by the quasilocal algebra $\mathscr{A}$ defined by its local subalgebras $\mathscr{A}^\Lambda$ as in (1.2), and by the local Hamiltonians Q^Λ from (1.5). Suppose that $\mathscr{A}^0:=\mathscr{L}(\mathscr{H}_0)$ is finite dimensional. We have an unambiguously defined global dynamics of the infinite system,[20] and we have a mathematically clear description (except of a detailed knowledge of the stability properties of states) of the corresponding equilibrium thermodynamics of such a system (cf. Sec. III). At the beginning of this section we shall describe briefly a general algorithm of *how to calculate* interesting quantities connected with a system from the considered class of mean-field theories. Afterwards we shall sketch applications of our methods to two simple nontrivial examples modeling the superconductivity.

(i) First, since $\mathscr{H}_0$ is finite dimensional, we can embed the matrices $X_j\in\mathscr{L}(\mathscr{H}_0)$ into a finite dimensional Lie algebra $X_\mathfrak{g}$ of matrices X_ξ $(\xi\in\mathfrak{g})$, where $\mathfrak{g}$ is the Lie algebra of a connected compact Lie group G. The group G can be taken, e.g., to be the unitary group $U(N)$ of $\mathscr{H}_0$, with $N:=\dim\mathscr{H}_0$, but it would be better to take G "as small as possible"—to simplify the forthcoming calculations. To obtain the minimal Lie algebra of matrices in $\mathscr{L}(\mathscr{H}_0)$ containing the X_j occurring in (1.5), we have to add those commutators $Y_{jk}:=i[X_j,X_m]\neq c^l_{jm}X_l$ which are linearly independent of the original set X_j $(j=1,2,...,k)$ to this set, and subsequently we have to add to this new set of matrices similar commutators between the matrices from the set $\{X_j,Y_{jm};j,m=1,2,...,k\}$, etc. The finite dimensionality ensures that we shall obtain a Lie algebra $X_\mathfrak{g}$ generated by $\{X_j,Y_{jk},Z_{jkl},...\}$ in a finite number of steps. ■

(ii) The next step in the investigation of the dynamics of our lattice model is the formulation and investigation of the classical equation corresponding to the group G and to the function $Q\in C^{\infty}(E)$. The Lie algebra $\mathfrak{g}$ of the group G is represented in $\mathscr{H}_0$ by the set $X_{\mathfrak{g}}$ of self-adjoint operators. The set E could be investigated subsequently, after recognizing some facts on the structure of the Ad* (G) orbits. The function Q is determined from the local Hamiltonians Q^{Λ} by the correspondence of (1.3) with (1.5). To obtain Q, we can choose $\Lambda:=\{0\}:=$ "the one point set containing a chosen point $0\in\Pi$" in (1.5), and insert into (1.5) of the coordinates $F_j:=F(\xi_j)$ for the operators X_j. The Poisson brackets are determined by the commutators between X_j's. For an arbitrary normalized $\psi\in\mathscr{H}_0$, let $F_\psi(\xi):=(\psi,X_\xi\psi)$. According to (2.1), as well as according to the relation

$$[X_\xi,X_\eta]=iX_{[\xi,\eta]},\quad \xi,\eta\in\mathfrak{g}, \tag{4.1}$$

we have the expression for Poisson brackets of the functions $f_\xi(F):=F(\xi)$,

$$\{f_\xi,f_\eta\}(F_\psi)=-F_\psi([\xi,\eta])=i(\psi,[X_\xi,X_\eta]\psi). \tag{4.2}$$

We can introduce the corresponding structure coefficients c^m_{jk} in the chosen basis X_j ($j=1,2,\ldots,n$) of the Lie algebra $X_{\mathfrak{g}}$,

$$[X_j,X_k]=ic^m_{jk}X_m, \tag{4.3}$$

and the corresponding Poisson brackets for the coordinate functions $F_j(F):=f_{\xi_j}(F)$ are obtained immediately:

$$\{F_j,F_k\}(F)=-c^m_{jk}F_m(F). \tag{4.4}$$

The next step is the writing down the Hamilton equations of motion[20] for the coordinate functions F_j:

$$\dot F_j=\{Q,F_j\},\quad j=1,2,..,n, \tag{4.5}$$

where the right-hand side can be calculated from (4.4). Equations (4.5) can be nonlinear even for quadratic Q, as it will be seen in the forthcoming examples. The fact of the existence of the flow φ^Q solving (4.5) is known,[24] but detailed investigation of properties of φ^Q might be very complicated for a general Q. However, for an analysis of equilibrium thermodynamics, it might be sufficient to know the stationary points $F\in\mathfrak{g}^*$ of (4.5), i.e., the points F, where $\{Q,F_j\}(F)=0$ for all j. ■

(iii) Now we can investigate the set $E:=\operatorname{supp}E_{\mathfrak{g}}$. Let us denote here $X_j:=X_{\xi_j}$ ($j=1,2,\ldots n$), where $\{\xi_j: j=1,2,\ldots,n\}$ is a basis of the Lie algebra $\mathfrak{g}$, and the self-adjoint X_ξ ($\xi\in\mathfrak{g}$) form the minimal matrix Lie algebra in $\mathscr{L}(\mathscr{H}_0)$ containing the originally given X_j's. Then, according to the Proposition 2.4, the set $\operatorname{conv}\{F_\psi\in\mathfrak{g}^*:F_\psi(\xi_j):=(\psi,X_j\psi),\ \|\psi\|=1,\ \psi\in\mathscr{H}_0,\ j=1,2,\ldots,n\}=\{F\in\mathfrak{g}^*:F(\xi)\in\operatorname{conv}\operatorname{sp}(X_\xi)$ for all $\xi\in\mathfrak{g}\}\subset\mathfrak{g}^*$ coincides with $E:=\operatorname{supp}E_{\mathfrak{g}}$.[30] An effective help in the construction of the set $E\subset\mathfrak{g}^*$, as it will be seen in our analysis of specific examples, is the Ad*(G) invariance of E. The set E is the set of all possible values of the "mean field" (i.e., of the macroscopic intensive quantities) of our system; all the values $F\in E$ are really achieved in some (even in factor) states. ■

(iv) With the solution φ^Q known, we can calculate all the details of the quantum microscopic evolution of arbitrarily local quantities $y\in\mathscr{A}$ by solving the equation for the unitary operator-valued function $U_Q:=U\circ g_Q$.[39] The time evolution of an arbitrary element $\hat f\in C(E,\mathscr{A})\sim\mathscr{C}=\mathscr{A}\otimes\mathscr{N}$ is described, according to (2.10) and (2.11), by the evolution φ^Q in $\mathscr{N}\sim C(E)$ as well as by the evolution of elements $x\in\mathscr{A}$ considered as constant functions on E. Hence

$$\tau^Q_t(x)=\int\sigma(g_Q^{-1}(t,F))(x)E_{\mathfrak{g}}(dF),\quad\text{for all}\quad x\in\mathscr{A}. \tag{4.6}$$

Elements from $\mathscr{A}$ are constructed by norm-continuous algebraic operations from elements $x\in\mathscr{A}^p$ ($p\in\Pi$), each $\mathscr{A}^p$ is invariant with respect to the action of $\sigma(G)$ described by a copy of the unitary action $\sigma(g)(x):=U(g)xU(g^{-1})$ ($x\in\mathscr{A}^0$, $g\in G$), and any C^*-morphism is norm continuous.[53] So, to get the knowledge of the action of τ^Q on $\mathscr{A}$, it is sufficient to know its action onto the elements of $\mathscr{A}^0$. For the cyclic representation π_ω corresponding to a factor state $\omega\in S(\mathscr{C})$ we have from Lemma (2.8)

$$\pi_\omega(\tau^Q_t(x))=\pi_\omega(U_Q^{-1}(t,F_\omega)xU_Q(t,F_\omega)),\quad\forall x\in\mathscr{A}^0,\quad\forall t\in\mathbb{R}. \tag{4.7}$$

We see that the time evolution of local elements from $\mathscr{A}$ depends on the "classical projection"[65] $F_\omega\in\mathfrak{g}$ of the considered factor state ω. Since each state has a canonical (central) decomposition into an integral of factor states, one can see that the knowledge of the evolution in (4.7) (in addition to the evolution φ^Q of classical observables) determines $\tau^Q\in{}^*$-aut $\mathscr{C}$ completely. Let us stress here several facts concerning this point (iv) of our exposition. ■

Remarks 4.1: (a) The time evolution (4.7) of local elements of $\mathscr{A}$ depends in the considered mean-field type theories on the chosen initial value F_ω determined by the (factor) state ω of the whole infinite system. This value F_ω is not changed by local perturbations of the state ω, hence it does not depend on the initial state of a local subsystem $\mathscr{A}^{\Lambda}$ ($\Lambda\in\mathscr{F}_{\Pi}$), i.e., on the restriction of ω to $\mathscr{A}^{\Lambda}$. In this way, the time evolution of local quantities is one-sidedly influenced (i.e., without an occurrence of any "locally observable feedback") by the classical macroscopic quantities of the large quantal system.

(b) With the classical evolution φ^Q known (it is a solution, as a rule, of a finite set of nonlinear coupled ordinary autonomous differential equations), the equation[39] for U_Q is a finite dimensional linear ordinary differential equation with time-dependent coefficients depending, moreover, on the parameter $F_\omega\in\mathfrak{g}^*$. The calculation of the dynamics of our infinite quantum system is reduced in this way to the solution of two finite dimensional ordinary differential equations (of the dimension at most $n:=\dim G$), only one of which is nonlinear. The dimensions in the examples of the strong coupling BCS model, resp. of the Josephson junction modeled by an interaction of two such BCS models are $n=3$, resp. $n=6$, as we shall see later in this section. ■

(v) The extremal KMS states of our system are constructed now according to Theorem 3.9. The essential tool of this construction is solution (for the unknown elements $F_\omega\in\mathfrak{g}^*$) of the "consistency condition"

$$\omega_\beta^0(X_\xi)=F_\omega(\xi). \tag{4.8}$$

Here the Gibbs state ω_β^0 is constructed according to (3.1) from the Hamiltonian X_η, $\eta:=\xi_{F_\omega}^Q$, where the Lie algebra element $\xi_{F_\omega}^Q$ is defined in (2.14). Equation (4.8) is a finite dimensional transcendental equation for F_ω as an implicitly given function of β and $c_{j_1j_2\cdots j_m}^{(m)}$ from the polynomial Q in (1.3).

The solutions F_ω of (4.8) need not be invariant with respect to a transformation $\varphi:E\to E$, which is a symmetry transformation of the Hamiltonian Q (cf. Remarks 3.15, and Ref. 64), i.e.,

$$Q\circ\varphi=Q,\ \text{and}\quad \text{simultaneously}\ \varphi(F_\omega)\neq F_\omega. \tag{4.9}$$

This is a kind of "symmetry breaking" which often occurs in the presence of phase transitions[1,6] at low temperatures. A further investigation of symmetries of the considered class of systems as well as of their "breaking" is postponed to a separate paper.[64] ■

We shall now illustrate the described procedure on two nontrivial, although technically relatively simple models.

A. The strong-coupling BCS model

The local systems are described in this model by the $\frac{1}{2}$-spin quantum variables, i.e., $\mathscr{H}_0:=\mathbb{C}^2$, and elements of $\mathscr{A}^0$ are all the 2×2 complex matrices. The local Hamiltonians Q^Λ of the considered version of the BCS model are described (up to an additive scalar "renormalization constant")[15] in terms of the Pauli matrices[24c] σ_j, $j=1,2,3$; $\sigma_{jp}:=\pi_p(\sigma_j)$ $(p\in\Pi)$, resp. of their combinations $\sigma_\pm:=\sigma_1\pm i\sigma_2$, as follows:

$$Q^\Lambda:=-\epsilon\sum_{p\in\Lambda}\sigma_{3p}-\frac{\lambda}{4|\Lambda|}\sum_{p\in\Lambda}\sigma_{+p}\sum_{q\in\Lambda}\sigma_{-q}, \tag{4.10}$$

where ϵ and λ are some positive numbers. Since the σ matrices are proportional to the generators of the two-dimensional unitary irreducible representation of the group SU(2), we shall choose $G:=\mathrm{SU}(2)$ in our considerations. The Pauli matrices satisfy the commutation relations (with the imaginary unit $i\in\mathbb{C}$)

$$[\sigma_j,\sigma_k]=2i\epsilon_{jkl}\sigma_l.$$

(We use here the standard summation convention; ϵ_{jkl} are components of the totally antisymmetric unit tensor of the third order in $\mathbb{R}^3$, $\epsilon_{123}:=1$). These relations correspond to the commutation relations for the generators $X_{\xi_j}:=\frac{1}{2}\sigma_j$ $(j=1,2,3)$ of the representation of SU(2) associated with the basis $\{\xi_j;j=1,2,3\}$ of the Lie algebra $\mathfrak{g}:=\mathfrak{su}(2)$ for which we have

$$[\xi_j,\xi_k]=\epsilon_{jkl}\xi_l,\quad j,k,(l)=1,2,3. \tag{4.11}$$

The classical Hamiltonian Q is now of the form

$$Q(F)=-2\epsilon F_3-\lambda(F_1^2+F_2^2)\equiv-2\epsilon F_3-\lambda F_+F_-, \tag{4.12}$$

where $F_\pm:=F_1\pm iF_2$, and $F_j(F):=F(\xi_j)$, $(F\in\mathfrak{g}^*)$. The Poisson structure on $\mathfrak{g}^*=\mathfrak{su}(2)^*$ is determined by the Poisson brackets

$$\{F_j,F_k\}=-\epsilon_{jkl}F_l,\quad j,k,(l)=1,2,3. \tag{4.13}$$

By the substitution into Eq. (4.5) of (4.12) for Q, and by taking the Poisson brackets from the relations (4.13), we obtain the classical equations of motion for our model, i.e.,

$$\dot F_1=2(\epsilon-\lambda F_3)F_2, \tag{4.14a}$$

$$\dot F_2=-2(\epsilon-\lambda F_3)F_1, \tag{4.14b}$$

$$\dot F_3=0. \tag{4.14c}$$

The solution is elementary:

$$F_+(t)=F_+(0)\exp(-2i(\epsilon-\lambda F_3)t), \tag{4.15}$$

with $F_3:=F_3(t)=F_3(0)$, for all $t\in\mathbb{R}$.

Since the time evolution φ^Q is nontrivial, the (φ^Q-invariant) Ad*-orbits have nonzero (even) dimension. Being submanifolds of the three-dimensional space $\mathfrak{su}(2)^*$, the orbits are two dimensional (except of a zero-dimensional orbit consisting of the point $F=0$). The compactness of the group SU(2) implies compactness of the orbits. Symplectic manifolds are orientable. The orbits are homeomorphic to a two-dimensional sphere S^2: The existence of the invariant $\mathbf{F}^2:=F_1^2+F_2^2+F_3^2$ of the Poisson structure, $\{\mathbf{F}^2,F_j\}=0$ $(j=1,2,3)$, shows that the Ad*(SU(2)) orbits are spheres $S_r^2:=\{F\in\mathfrak{su}(2):\mathbf{F}^2=r^2\}$, $r\geqslant0$. Let us consider now the structure of supp $E_\mathfrak{g}$.

Lemma 4.2: $E:=\mathrm{supp}\,E_\mathfrak{g}=\{F\in\mathfrak{su}(2)^*:\mathbf{F}^2\leqslant\frac{1}{4}\}$. ■

Proof: The spectra of the generators X_{ξ_j} (j = 1,2,3) are the two-point sets $\{1/2,-1/2\}$. One can use the Ad* invariance of E, the spheric form of the Ad* orbits, as well as the convexity of E,[66] to obtain the result. Q.E.D.

The quantum evolution is determined by the elements

$$\xi_F^Q:=d_FQ=-2\epsilon\xi_3-2\lambda(F_1\xi_1+F_2\xi_2),\quad F\in\mathfrak{g}^*, \tag{4.16}$$

of the Lie algebra $\mathfrak{g}$. Writing $X(\xi):=X_\xi$, the corresponding (one-spin) "Bogoliubov–Haag" generator in the representation U is

$$X(\xi_F^Q)=-\epsilon\sigma_3-\lambda(F_1\sigma_1+F_2\sigma_2). \tag{4.17}$$

The evolutiuon of arbitrary quantal observables (or states) can be calculated now in the way described in the point (iv) above, i.e., by solving a nonautonomous linear ordinary differential equation.

Let us write down the "consistency condition" for values $F_\omega\in\mathfrak{su}(2)^*$ of classical macroscopic quantities in extremal KMS states of the considered BCS model. Let $n(F)$ be the three-dimensional unit vector with coordinates

$$n_1:=\frac{\lambda F_1}{a(F)},\quad n_2:=\frac{\lambda F_2}{a(F)},\quad n_3:=\frac{\epsilon}{a(F)}, \tag{4.18}$$

with

$$a(F):=\sqrt{\epsilon^2+\lambda^2F_+F_-}. \tag{4.19}$$

If the components of the three-dimensional matrix-valued vector σ are the Pauli matrices, we can write $X(\xi_F^Q)=-a(F)n(F)\cdot\sigma$. The "consistency condition" is satisfied by the Gibbs state ω_β^0 on $\mathscr{L}(\mathbb{C})^2$ corresponding to the Hamiltonian $H:=X(\xi_F^Q)$ according to (3.1) iff the relation

$$\omega_\beta^0(X_\xi)=F(\xi) \tag{4.20}$$

is fulfilled. This is the condition for $F=F_\omega$. By taking $\xi:=\xi_j$ $(j=1,2,3)$ to be the elements of the chosen basis of

$\mathfrak{su}(2)$, the condition (4.20) can be rewritten in the form of the equations

$$n_j(F_\omega)\,\mathrm{th}(\beta a(F_\omega)) = 2F_\omega(\xi_j), \quad j=1,2,3. \tag{4.21}$$

These conditions are satisfied by the following two sets S_n and S_s of pure classical macroscopic states $F:=F_\omega$:

normal: $S_n := \{F\in\mathfrak{su}(2)^*: F_+ = 0, F_3 = \frac{1}{2}\mathrm{th}(\beta\epsilon)\}$ for any $0<\beta<\infty$, $\epsilon>0$, $\lambda>0$. ■

superconducting: $S_s := \{F\in\mathfrak{su}(2)^*: F_3 = \epsilon/\lambda,$

$2a(F) = \lambda\,\mathrm{th}(\beta a(F))\}$,

for $0<2\epsilon<\lambda$,

if the positive temperature is $\beta^{-1} < T_c := \epsilon[\mathrm{th}^{-1}(2\epsilon/\lambda)]^{-1}$; the set S_s is a circle lying in the plain $F_3 = \epsilon/\lambda$ in $\mathfrak{su}(2)^*$ with the positive radius $|F_+|$, and the center at $F_1 = F_2 = 0$. ■

The set S_n corresponds to the "normal conducting phase," and the set S_s describes equilibrium values F_ω for the "superconducting phase."[17] The local quantal Hamiltonians (4.10) as well as the classical Hamiltonian (4.12) are all invariant with respect to arbitrary rotations around the third axis [in the chosen basis of $\mathbb{R}^3 = \mathfrak{su}(2)$]. The same is true for the points $F\in S_n$. But the points $F\in S_s$ corresponding to the subcritical values of the temperature are not invariant with respect to the arbitrary rotations $\varphi\in S^1$ around the 3rd axis: we have $F_+\neq 0$ for these points. Hence, the symmetry is broken in the superconducting phase. Let us note, that the thermodynamic limit ω of local Gibbs states for subcritical β^{-1} is given by the integral (2.19) of the factor-KMS states $\varkappa$ over the circle S_s ($\ni F_\varkappa$) with the (φ invariant, $\varphi\in S^1$) Lebesgue measure μ^ω.[16,17,30]

B. A model of Josephson junction

This model is constructed by combining two copies of the models described in Sec. II A. The behavior of its macroscopic quantities for the system interacting with a thermal reservoir was partly examined already in the classical paper[18] by Hepp and Lieb.[67] The group G is now the direct product $SU(2)_a\times SU(2)_b$ (here we distinguish the two copies of the group and of other related objects by an index $c\in\{a;b\}$), and it is represented in $\mathscr{L}(\mathbb{C}_a^2\otimes\mathbb{C}_b^2)$ by the tensor product $U(SU(2)_a)\otimes U(SU(2)_b)$ of the two two-dimensional representations $U(SU(2)_a)$ and $U(SU(2)_b)$. The group is now six-dimensional, and the elements $\xi_{cj} \in\mathfrak{g} := \mathfrak{su}(2)_a\oplus\mathfrak{su}(2)_b$ ($c=a,b; j=1,2,3$) of a basis of the Lie algebra $\mathfrak{g}$ satisfy the commutation relations

$$[\xi_{cj},\xi_{dk}] = \delta_{cd}\epsilon_{jkl}\xi_{cl} \quad (\text{no summation over } c),$$
$$c,d\in\{a;b\}, \text{ and } j,k,(l) = 1,2,3. \tag{4.22}$$

For the corresponding coordinate functions $F_{cj} := F(\xi_{cj})$ of the classical (generalized) phase space $\mathfrak{g}^*$, and for the corresponding complex functions $F_{c\pm} := F_{c1}\pm iF_{c2}$ we obtain the Poisson brackets

$$\{F_{a+},F_{a-}\} = 2iF_{a3}, \quad \{F_{a3},F_{a\pm}\} = \pm iF_{a\pm}, \tag{4.23}$$

and similarly for $a\to b$. Clearly, $\{F_{aj},F_{bk}\}=0$. The dynamics is given by the classical Hamiltonian

$$Q(F) := -2\epsilon_a F_{a3} - \lambda_a F_{a+}F_{a-} - 2\epsilon_b F_{b3} - \lambda_b F_{b+}F_{b-} + \kappa(F_{a+}F_{b-} + F_{a-}F_{b+}). \tag{4.24}$$

Here ϵ_c, λ_c, and κ are some real constants. The corresponding equations of motion (4.5) have the form:

$$\dot F_{a+} = 2i[(\lambda_a F_{a3} - \epsilon_a)F_{a+} - \kappa F_{a3}F_{b+}], \tag{4.25a}$$
$$\dot F_{b+} = 2i[(\lambda_b F_{b3} - \epsilon_b)F_{b+} - \kappa F_{b3}F_{a+}], \tag{4.25b}$$
$$\dot F_{a3} = -i\kappa(F_{a+}F_{b-} - F_{a-}F_{b+}), \tag{4.25c}$$
$$\dot F_{b3} = i\kappa(F_{a+}F_{b-} - F_{a-}F_{b+}). \tag{4.25d}$$

It might be instructive to compare the derivation and the form of Eqs. (4.25) with the derivation and the form of the corresponding equations in Ref. 19. Let us formulate some facts on the classical motion described by these equations.

Proposition 4.3: The classical equations (4.25) on the manifold $\mathfrak{g}^*$ (with $\mathfrak{g} := \mathfrak{su}(2)\oplus\mathfrak{su}(2)$) [corresponding to the Poisson structure (4.23) and to the dynamics given by the Hamiltonian Q from (4.24)] are completely integrable[24] for almost all (with respect to the Lebesgue measure) values of parameters, e.g., for $\kappa\neq 0$. They have the following independent (global) integrals of motion:

$$Q, \quad F_3 := F_{a3} + F_{b3},$$
$$r_a^2 := F_{a1}^2 + F_{a2}^2 + F_{a3}^2, \quad r_b^2 := F_{b1}^2 + F_{b2}^2 + F_{b3}^2. \ ■$$

Proof: By direct calculations from Eqs. (4.25). Q.E.D.

The integrals $r_{a,b}$ are of "kinematical character": they determine the Ad* orbit on which the motion is realized, i.e., the Cartesian product of two spheres S^2 with radii $r_{a,b}$ forming the symplectic manifold for the actual motion of the system. These integrals are present for an arbitrary Poisson flow on our $\mathfrak{g}^*$, hence they are independent of the Hamiltonian $Q\in C^\infty(\mathfrak{g}^*)$. The integrals Q and F_3 are the "dynamical" ones: their existence leads to the integrability of the system on the four-dimensional symplectic manifold $S^2_{r_a}\times S^2_{r_b}$. One can use a Sard's theorem[24] to prove integrability for almost all values of the integrals of motion.

The "consistency condition" (4.8) is formulated with a help of the Bogoliubov–Haag Hamiltonian $X(\xi_F^Q)$ of the present model,

$$\begin{aligned} X(\xi_F^Q) &= -\epsilon_a\sigma_3^a + (\kappa F_{b1} - \lambda_a F_{a1})\sigma_1^a \\ &\quad + (\kappa F_{b2} - \lambda_a F_{a2})\sigma_2^a - \epsilon_b\sigma_3^b \\ &\quad + (\kappa F_{a1} - \lambda_b F_{b1})\sigma_1^b + (\kappa F_{a2} - \lambda_b F_{b2})\sigma_2^b \\ &= -a(F)n^a\cdot\sigma^a - b(F)n^b\cdot\sigma^b, \end{aligned} \tag{4.26}$$

where

$$a(F)^2 := \epsilon_a^2 + (\kappa F_{b1} - \lambda_a F_{a1})^2 + (\kappa F_{b2} - \lambda_a F_{a2})^2,$$
$$b(F)^2 := \epsilon_b^2 + (\kappa F_{a1} - \lambda_b F_{b1})^2 + (\kappa F_{a2} - \lambda_b F_{b2})^2,$$

and the three-vectors n^a and n^b have components

$$n_j^a := \frac{1}{a(F)}(\lambda_a F_{aj} - \kappa F_{bj}), \quad j=1,2; \quad n_3^a := \frac{\epsilon_a}{a(F)};$$
$$n_j^b := \frac{1}{b(F)}(\lambda_b F_{bj} - \kappa F_{aj}), \quad j=1,2; \quad n_3^b := \frac{\epsilon_b}{b(F)}.$$

Using the commutativity $[\sigma^a,\sigma^b]=0$, the standard calculations with the σ-matrices lead to the following form of the "consistency condition":

$$2F_{aj} = n_j^a \,\mathrm{th}(\beta a(F)),$$
$$2F_{bj} = n_j^b \,\mathrm{th}(\beta b(F)), \quad j=1,2,3\,. \qquad (4.27)$$

The solutions $\{F_a;F_b\}\in\mathfrak{su}(2)^* \oplus \mathfrak{su}(2)^*$ of these equations give, according to Proposition 3.10, classification of extremal KMS states at positive temperatures β^{-1}. We postpone the detailed analysis of Eqs. (4.25) and (4.27) to another paper.

Note 4.4: Let us finish with some additional historical notes: In our considerations and calculations, we have extensively used the algebraic description of classical macroscopic quantities of the investigated infinite quantal dynamical systems. Such a description of classical "observables at infinity" was exploited in modeling the measurement process in quantum theory,[68] by trials of obtaining some answer to the old interpretation problem of the theory—the "measurement problem" in quantum physics. The mathematical background for these investigations was formulated[69] soon after the formulation of basic principles of quantum mechanics. In the present time the physically most interesting connections of the described phenomena and of the used mathematical formalism might be found, perhaps, in quantum field theory and its applications in particle and solid-state physics.[70]

ACKNOWLEDGMENTS

The author is indebted to Professor W. Thirring for his kind and stimulating hospitality at the Institute of Theoretical Physics of the University of Vienna, and to Professor A. Rieckers for his encouraging interest in the recent author's work. The author is grateful also to Professor K. Hepp, to Professor F. Strocchi, and to Professor A. Verbeure for sending him copies of their interesting papers.

[1](a) D. Ruelle, *Statistical Mechanics: Rigorous Results* (Benjamin, New York, 1969); (b) G. G. Emch, *Algebraic Methods in Statistical Mechanics and Quantum Field Theory* (Wiley, New York, 1972); (c) W. Thirring, *Lehrbuch der Mathematischen Physik* (Springer, New York, 1980), Band 4.

[2](a) C. Domb and M. S. Green (Eds.), *Phase Transitions and Critical Phenomena* (Academic, New York, 1972), Vol. 1 and Vol. 2; (b) H. E. Stanley, *Introduction to Phase Transitions and Critical Phenomena* (Claredon, Oxford, 1971).

[3]L. Onsager, Phys. Rev. **65**, 117 (1944).

[4]R. Haag, Dan. Mat. Fys. Medd. **29**, No. 12 (1955).

[5]R. Haag and D. Kastler, J. Math. Phys. **5**, 848 (1964).

[6]O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics* (Springer, New York, 1979 and 1981), Vol. I and II.

[7]Here we give some (accidentally chosen) examples of works dealing in the last few years with the problem of phase transitions by taking the thermodynamic limit of local equilibria: M. van den Berg, J. T. Lewis, and J. V. Pulè, Helv. Phys. Acta **59**, 1271 (1986); R. Kotecký, Commun. Math. Phys. **82**, 391 (1981); R. Kotecký and S. Miracle-Sole, J. Stat. Phys. **47**, 773 (1987); J. T. Lewis, "Why do Bosons Condense?", T. Kennedy and E. H. Lieb, "A Model for Crystallization: A Variation on the Hubbard Model," both in *Proceedings of the Conference on Statistical Mechanics and Field Theory, Groningen*, August 1985, in Lecture Notes in Physics (Springer, New York, 1986); E. H. Lieb and H.-T. Yau, Commun. Math. Phys. **112** (1987).

[8]D. W. Robinson, *The Thermodynamic Pressure in Quantum Statistical Mechanics*, in Lecture Notes in Physics 9 (Springer, New York, 1971); G. Gallavotti and A. Martin-Löf, Commun. Math. Phys. **25**, 87 (1972); H. Narnhofer, Acta Phys. Austriaca **54**, 221 (1982).

[9]H. Araki and G. L. Sewell, Commun. Math. Phys. **52**, 103 (1977).

[10]J. Quaegebeur and A. Verbeure, Ann. Inst. Henri Poincarè **XXXII**, 343 (1980).

[11]D. Ruelle, *Thermodynamic Formalism* (Addison-Wesley, Reading, MA, 1978).

[12]R. Haag, N. M. Hugenholtz, and M. Winnink, Commun. Math. Phys. **5**, 215 (1967); cf. Definition 3.1 of the present paper.

[13]D. W. Robinson, Commun. Math. Phys. **7**, 337 (1968).

[14]R. Haag, Nuovo Cimento **25**, 287 (1962).

[15]W. Thirring and A. Wehrl, Commun. Math. Phys. **4**, 303 (1967); W. Thirring, *ibid.* **7**, 181 (1968).

[16]D. A. Dubin and G. L. Sewell, J. Math. Phys. **11**, 2990 (1970); G. G. Emch and H. J. F. Knops, *ibid.* **11**, 3008 (1970); M. Fannes, H. Spohn, and A. Verbeure, *ibid.* **21**, 355 (1980).

[17]F. Jelinek, Commun. Math. Phys. **9**, 169 (1968).

[18]K. Hepp and E. H. Lieb, Helv. Phys. Acta **46**, 573 (1973).

[19]E. Duffner, Z. Phys. B **63**, 37 (1986).

[20]P. Bóna, J. Math. Phys. **29**, 2223 (1988).

[21]Also the following papers are relevant in this direction: E. Duffner and A. Rieckers, Z. Naturforsch. **43a**, 321 (1988); T. Unnerstall, "Phase-spaces and dynamical descriptions of infinite mean-field quantum systems," preprint of the University of Tübingen 1988, to appear in J. Math. Phys.

[22]S. Sakai, *C*-Algebras and W*-Algebras* (Springer, Berlin, 1971).

[23]The polynomial Q in (1.3) can be considered as a function of coordinates $F_j := F(\xi_j)$ of elements $F\in\mathfrak{g}^*$ in the dual basis to a chosen basis $\{\xi_j\}$ of $\mathfrak{g}$.

[24](a) R. Abraham and J. E. Marsden, *Foundations of Mechanics*, 2nd ed. (Benjamin/Cummings, Reding, MA, 1978); (b) W. Thirring, *Lehrbuch der Mathematischen Physik* (Springer, New York, 1977), Band 1; (c) Y. Choquet-Bruhat, C. DeWitt-Morette, and M. Dillard-Bleick, *Analysis, Manifolds, and Physics*, revised edition (North-Holland, Amsterdam, 1982); (d) L. Markus and K. R. Meyer, *Generic Hamiltonian Dynamical Systems are neither Integrable nor Ergodic*, Memoirs of AMS 144 (AMS, Providence, RI, 1974).

[25]A. A. Kirillov, *Elementy teorii predstavlenií* (Nauka, Moscow, 1978).

[26]M. Takesaki, *Theory of Operator Algebras* (Springer, New York, 1979).

[27]C. M. Marle, in *Bifurcation Theory, Mechanics and Physics*, edited by C. P. Bruter, A. Aragnol, and A. Lichnerowicz (Reidel, Dordrecht, 1983).

[28]G. K. Pedersen, *C*-Algebras and their Automorphism Groups* (Academic, New York, 1979).

[29]Remember that any weakly continuous action of a one-parameter group on commutative von Neumann algebra is trivial, cf., e.g., Proposition II.5 in P. Bóna, Acta Phys. Slov. **25**, 3 (1975).

[30](a) P. Bóna, *Classical Projections and Macroscopic Limits of Quantum Mechanical Systems* (preprint 1984, a revised version will be published by Comenius University Press, Bratislava, 1990); (b) P. Bóna, Czech. J. Phys. B **37**, 482 (1987).

[31]Cf. Sec. II C, especially Proposition 2.9, of Ref. 20.

[32]Cf. Proposition 2.6 of Ref. 20.

[33]O. E. Lanford and D. Ruelle, Commun. Math. Phys. **13**, 194 (1969); cf. also Theorem 2.6.10 of Ref. 6.

[34]G. Choquet, *Lectures on Analysis* (Benjamin, Reading, MA, 1969), Vol. I, Proposition 19.9.

[35]Cf. Ref. 6, Example 4.1.31 and the next following remarks.

[36]M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1972), Vol. I, Theorem VIII.33, and Theorem VIII.24.

[37]For details see Lemma 6.2.17 of Ref. 30.

[38]Cf. Ref. 20, Sec. II C.

[39]Cf. Ref. 20, Eq. (2.32).

[40]Cf. Ref. 20, Corollary 4.4.

[41]Cf. Ref. 20, Remark 3.8.

[42]The natural generalization of this result [based on the formula (3.27) of Ref. 20] to permutation invariant τ^Q-invariant states that could be decomposed to τ^Q-noninvariant factor states will appear in *Proceedings of International Conference on Selected Topics in Quantum Field Theory and Mathematical Physics*, Liblice Castle, 25–30 June 1989 (World Scientific, Singapore, 1990); in the contribution by cf. also T. Unnerstall, cf. also Tübingen preprint, 1989.

[43]I. E. Segal, Duke Math. J. **18**, 221 (1951); cf. Ref. 1, Theorem 6.2.2.

[44]Cf. Ref. 6, Chap. 4.4.

[45]For the correct definition of the appropriate integral decomposition of ω cf. Ref. 30, Theorem 5.1.38(v), and also in Ref. 30, the proof of Proposition 6.3.6.

[46]N. N. Bogoliubov, Sov. Phys. JETP **34**, 58, 73 (1958).

[47]G. E. Uhlenbeck and G. W. Ford, *Lectures in Statistical Mechanics* (AMS, Providence, RI, 1963).
[48]L. D. Landau and E. M. Lifshitz, *Statisticheskaya Fizika* (Nauka, Moscow, 1964).
[49]Cf. Ref. 6, Definitions 5.3.1, 5.3.18, and 5.3.21.
[50]Reference 6, Proposition 5.3.19.
[51]M. A. Naĭmark, *Normirovannye Kol'tsa*, 2nd rev ed. (Nauka, Moscow, 1968), Sec. 34.2; cf. also the von Neumann "bicommutant theorem" with its relevant connections in any of the Refs. 6, 22, 26, 28, 52, and 53.
[52]W. Thirring, *Lehrbuch der Mathematischen Physik* (Springer, New York, 1979), Band 3.
[53]J. Dixmier, *Les algèbres d'opérateurs dans l'espace Hilbertien*, 2nd ed. (Gauthier-Villars, Paris, 1969).
[54]Cf. Chap. 5 of Ref. 6.
[55]Cf. Ref. 34, Definition 28. 1, or Ref. 1(a), paragraph D.5.3, or Ref. 6, Vol. I, p. 328.
[56]A face $\mathscr{F}$ of a compact convex set $\mathscr{S}$ is here any subset $\mathscr{F} \subset \mathscr{S}$ such, that $\omega \in \mathscr{F}$ and $\omega = \Sigma_{j=1}^{n} \lambda_j \omega_j$ (with $\lambda_j > 0$, $\Sigma_{j=1}^{n} \lambda_j = 1$) implies $\omega_j \in \mathscr{F}$. Cf. Ref. 6, Vol. I, footnote on p. 313.
[57]Cf. Ref. 6, Chap. 4 and Theorem 5.3.30.
[58]A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (PWN, Warsaw, 1977).
[59]Cf. also Ref. 20, Sec. II B.
[60]L. van Hemmen, Fortschr. Phys. **26**, 397 (1978).
[61]Cf. Chap. 5.3 of Ref. 6.
[62]Cf. Chap. 4.2 of Ref. 6.
[63]Cf. also G. G. Emch in Ref. 2(a), Vol. 1 for a heuristic formulation of some of the assertions of this theorem in the case of the models with quadratic Q.
[64]P. Bóna, preprint 1989 (to be published).
[65]Cf. our Proposition 2.3.
[66]Cf. Proposition 2.4 of this paper.
[67]The model was formulated and investigated by another method also in the paper: K. Hepp, "On the Classical Limit in Quantum Mechanics," Lecture at the International School in Mathematical Physics, Camerino, Italy, 30 Sept.–12 Oct., 1974.
[68]K. Hepp, Helv. Phys. Acta **45**, 237 (1972); H. Primas, preprint ETH-Zürich (1972); P. Bóna, Acta Phys. Slov. **23**, 149 (1973); **27**, 101 (1977); and ACTA F.R.N. Univ. Comen.-PHYSICA XX, 65 (1980).
[69]J. von Neumann, Compos. Math. **6**, 1 (1938).
[70]Cf., e.g., F. A. Berezin, Commun. Math. Phys. **63**, 131 (1978); G. Morchio and F. Strocchi, J. Math. Phys. **28**, 622 (1987); Commun. Math. Phys. **111**, 593 (1987).

Erratum: Some results from a Mellin transform expansion for the heat kernel [J. Math. Phys. 30, 1226 (1989)]

A. P. C. Malbouisson[a)] and F. R. A. Simao
Centro Brasileiro de Pesquisas Físicas, CBPF/CNPq, Rua Dr. Xavier Siguad, 150, 22290 Rio de Janeiro, RJ, Brazil

A. F. de Camargo Filho
Instituto de Físicas Teórica, Rua Pamplona, 145, 01405 São Paulo, SP, Brazil

(Received 19 May 1989; accepted for publication 28 July 1989)

After the proofs of our above mentioned paper were corrected, we noted that even if the list of singularities of Seeley's kernel[1] $K(s;x,x)$ at $s=(j-D)/m$, $j=1,2,\ldots$ includes the point $s=0$, corresponding to $j=D$ (D integer), the function $K(s;x,x)$ is effectively not singular at $s=0$, as a result of theorem 4 of Seeley.[1] This allows us to write the asymptotic expansion (1,2) in the form

$$F(t;x,x) = -\sum_{l=0}^{\infty} a_l(x)t^l - \sum_j \Gamma\left(\frac{D-j}{m}\right) R_j(x)t^{(j-D)/m}, \tag{1}$$

where $a_l(x) = -K(l;x,x)$. The sum over j is restricted to the condition $(j-D)/m \neq 1,2,\ldots$.

Then the formula for the anomaly (4.2) in arbitrary (integer) dimension D may be written in the simpler form,

$$A = q\,\mathrm{Tr}\{(A+B)[a_0(x)+P_0(x,x)]\}. \tag{2}$$

[a)] Present address: Centre de Physique Theorique, Ecole Polytechnique, 91.128, Palaiseau, France.

[1] R. T. Seeley, Am. Math. Soc. Proc. Symp. Pure Math. 10, 2885 (1967).

 0022-2488/89/123008-01$02.50