SU(2) classification of $N=2$ complex supersymmetric representations

E. H. Saidi[a)]
International Centre for Theoretical Physics, Trieste, Italy

(Received 15 January 1987; accepted for publication 18 May 1988)

The $N=2$ extended complex supersymmetry representations are constructed. The classification of $N=2$ multiplets in SU(2) representations is obtained. The relation with $N=2$ harmonic superspace is also discussed.

I. INTRODUCTION

The discovery of supersymmetry by Wess and Zumino[1] in 1974 has opened up a large area of research in particle physics. Indeed, even though there is no experimental evidence in its favor, almost all particle physicists believe that supersymmetry has played some role in the evolution of the early universe.[2]

From the theoretical point of view, supersymmetric theories possess many special features, e.g., Bose–Fermi symmetry and nonrenormalization theorems, and they have been viewed for some time as candidates for unification of all known forces. Supersymmetry is also required for consistent string theories[3] formulated in ten-dimensional space-time.

It is well known that a theory with $N=1$ supersymmetry in higher dimensions[4] will involve extended supersymmetry when reduced to four dimensions. However, an off-shell formulation of extended supersymmetry[5] for general N has not yet been achieved. Recently, Galperin *et al.*[6] have succeeded in obtaining an adequate off-shell formulation of $N=2$ Poincaré supersymmetry. Their formalism preserves the SU(2) symmetry combining the two supersymmetries.[5] The main idea of their approach is to project this SU(2) symmetry onto the $SU(2)/U^c(1)\simeq S^2$ sphere by introducing a new set of harmonic variables.[6,7] Among the obtained results, $N=2$ supersymmetry is realized linearly on a subset of the $N=2$ harmonic superspace (HS), called the analytic subspace. They also have derived the massless Lagrangians for the $N=2$ scalar and gauge multiplets. At the quantum level they prove the $N=2$ nonrenormalization theorem[8] and show that harmonic nonlocalities disappear if external legs of a diagram are put on shell.[9] The curved harmonic superspace and the fundamentals of the underlying $N=2$ supergravity theories have been also constructed.

Moreover, the harmonic superspace techniques give in a natural way the classification of the content of the $N=2$ on-shell supersymmetric multiplets in terms of the SU(2) representations as obtained for the $N=2$ on-shell scalar multiplet $(0^4,\frac{1}{2}^2)$ and the vector one $(0^2,\frac{1}{2}^2,1)$.[6]

In this paper, we give the full classification in terms of SU(2) representations of the $N=2$ on-shell (complex) supersymmetric multiplets. We find as particular cases the Fayet–Sohnius (FS) (Ref. 10) and the Howe–Stelle–Townsend (HST) (Ref. 11) hypermultiplets. The problem of massive supersymmetric representations with and without central charges and their connection with the SU(2) symmetry is examined in detail. We show in particular that the HST hypermultiplet can describe a massive representation up to a complexification.

The main idea of this study is to take advantage of the presence of the Cartan–Weyl $U^c(1)$ charge to build $N=2$ charge dependent sypersymmetric multiplets.

The presentation is as follows. In Sec. II, we derive the $N=2$ extended "complex" supersymmetry algebra, useful for the SU(2) classification.

In Sec. III, we construct the massive representations without central charges. First, we build the representations of the so-called $(-)$ and $(+)$ algebras. Then, we give those of the full $N=2$ extended complex algebra. The usual $N=2$ extended "real" representations are obtained through a reality condition. Cases of interest, such as the scalar, vector, and gravitational multiplets, are worked out explicitly.

In Sec. IV, central charges are introduced and we discuss in particular the relation between the Fayet hypersymmetry[12] and the Fayet–Sohnius (FS) representation.[10,12] The connection with the HST sector is also established.

In Sec. V, the massless representations are derived. We establish the relations between the massless $N=2$ scalar and vector multiplets and the analytic superfields given in Ref. 6.

Our results are summarized in the last section.

II. $N=2$ COMPLEX SUPERSYMMETRIC ALGEBRA IN THE HARMONIC SUPERSPACE

Consider the two beins $u_i^{\pm}$ $(u_i^- = \overline{u^{+i}})$ parametrizing the $SU(2)/U^c(1)\simeq S^2$ sphere introduced in Ref. 6. These parameters play an important role in the harmonization of the $N=2$ extended superspace. They satisfy the following constraints:

$$U^{\pm i}U_i^{\mp} = \pm 1\,, \quad U^{\pm i}U_i^{\pm} = 0\,, \tag{2.1a}$$

or equivalently

$$U^{Ii}U_i^J = \epsilon^{IJ}\,, \quad I,J = +,-\,, \quad \epsilon^{+-} = \epsilon_{-+} = 1\,. \tag{2.1b}$$

These parameters, also called harmonic variables, are used to project the SU(2) symmetry combining the two supersymmetries into $U^c(1)$ objects, namely,

$$U_i^I:\ A^i \to A^I = A^i U_i^I\,, \tag{2.2}$$

where A^i is an SU(2) isodoublet.

The inverse transformation $(U_i^I)^{-1}$ is well defined and may be constructed, using formally (2.1), as follows:

a) Permanent address: L.P.T., Faculté des Sciences, Av. Ibn Battouta, P.O. Box 1014, Rabat, Morocco.

 0022-2488/88/091949-09$02.50

$$(U_i^I)^{-1} \sim \epsilon_{IJ} U^{iJ}, \tag{2.3}$$

so that

$$(U_i^I)^{-1} : A^I \to (U_i^I)^{-1}(A^{\pm}) = c \int du\, \epsilon_{IJ} U^{iJ} A^I, \tag{2.4}$$

where c is a constant that can be dropped out. Indeed, integrating over the harmonic variables[6,13,14] and using (2.2), one finds

$$c \int du\, \epsilon_{IJ} A^I U^{iJ} = c A^i$$

and therefore $c = 1$ in our particular case. One can also check

$$\begin{aligned} U^{i,I}(A^+) &= U^{i-}(A^+) = -\tfrac{1}{2} A^i, \\ U^{i,I}(A^-) &= U^{i+}(A^-) = \tfrac{1}{2} A^i, \end{aligned} \tag{2.5a}$$

and combining together these relations, one has

$$\begin{aligned} U^{iI}(A^-) - U^{iI}(A^+) &= U^{i+}(A^-) - U^{i-}(A^+) \\ &= A^i. \end{aligned} \tag{2.5b}$$

From the relations (2.5), one sees that the projected (hidden) SU(2) symmetry (2.2) manifests itself by the conservation of the Cartan–Weyl charge. This is a direct consequence of relations (2.1) and (2.4). This procedure of projecting the SU(2) symmetry will permit us to explore further the $N = 2$ extended supersymmetry. It also gives a covariant formalism avoiding all the complexity of the SU(2) tensor calculus. Furthermore, the existence of the inverse transformation (2.4) ensures the restoration of the SU(2) symmetry at any step from the harmonic expression.[13,14]

We now want to exploit this fact at the level of the $N = 2$ extended supersymmetric algebra. For that purpose, define

$$\begin{aligned} Q_a^{\pm} &= Q_a^i U_i^{\pm} = Q_a^1 U_1^{\pm} + Q_a^2 U_2^{\pm}, \\ \overline{Q}_{\dot{a}}^{\pm} &= \overline{Q}_{\dot{a}}^i U_i^{\pm} = \overline{Q}_{\dot{a}}^1 U_1^{\pm} + \overline{Q}_{\dot{a}}^2 U_2^{\pm}, \end{aligned} \tag{2.6a}$$

or in a simple form

$$\begin{aligned} Q_a^I &= Q_a^i U_i^I, \\ \overline{Q}_{\dot{a}}^I &= \overline{Q}_{\dot{a}}^i U_i^I, \end{aligned} \quad I = +, - . \tag{2.6b}$$

The projections (2.6) generalize the usual complex representation,[12,15] namely,

$$\begin{aligned} Q_a^{\pm} &= (1/\sqrt{2})[Q_a^1 \pm i Q_a^2], \\ \overline{Q}_{\dot{a}}^{\mp} &= (1/\sqrt{2})[\overline{Q}_{\dot{a}}^1 \mp i \overline{Q}_{\dot{a}}^2]. \end{aligned} \tag{2.7}$$

Using now (2.5), the odd part of the $N = 2$ extended supersymmetric algebra[5,16] may be written as

$$\begin{aligned} \{Q_a^{\pm}, \overline{Q}_{\dot{a}}^{\mp}\} &= \pm 2\sigma_{a\dot{a}}^{\mu} P_{\mu}, \\ \{Q_a^{\pm}, Q_b^{\pm}\} &= \pm 2iZ\epsilon_{ab}, \\ \{\overline{Q}_{\dot{a}}^{\pm}, \overline{Q}_{\dot{b}}^{\mp}\} &= \pm 2i\overline{Z}\epsilon_{\dot{a}\dot{b}}, \\ \{Q_a^{\pm}, Q_b^{\pm}\} &= \{\overline{Q}_{\dot{a}}^{\pm}, \overline{Q}_{\dot{b}}^{\pm}\} = 0. \end{aligned} \quad (\pm) \tag{2.8}$$

Note, by the way, that the nonvanishing anticommutators in (2.8) correspond to a conserved Cartan–Weyl charge (the sum of charges is identically zero). This fact is, once more, related to relations (2.1) and (2.4).

Each of the sets $(\pm)$ in (2.8) constitutes an $N = 1$ complex supersymmetric algebra. These algebras are dual to each other, as one can see from their simple realization (2.7), and present a nontrivial generalization of the Fayet hypersymmetry. The duality of the $(+)$ and $(-)$ $N = 1$ complex superalgebras is precisely carried by the Cartan–Weyl charge. The passage from the $(+)$ to $(-)$ algebras and vice versa is ensured by the star conjugation (A_1) introduced in Ref. 6.

Furthermore, these two complex supersymmetric sets $(+)$ and $(-)$ are in fact two subalgebras of the following $N = 2$ extended complex superalgebra:

$$\begin{aligned} \{Q_a^I, \overline{Q}_{\dot{b}}^J\} &= 2\sigma_{a\dot{a}}^{\mu} P_{\mu} \epsilon^{IJ}, \\ \{Q_a^I, Q_b^J\} &= 2iz\epsilon_{ab}\epsilon^{IJ}, \\ \{\overline{Q}_{\dot{a}}^I, \overline{Q}_{\dot{b}}^J\} &= 2i\bar{z}\epsilon_{\dot{a}\dot{b}}\epsilon^{IJ}, \end{aligned} \quad (0) \tag{2.9}$$

where $I = +, -$ and $\epsilon^{+-} = \epsilon_{-+} = 1$.

The central charges z and $\bar{z}$ of the $N = 2$ real supersymmetry, which carry the "bridge" from the $(\pm)$ to $(\mp)$ subalgebras, are now connected with the Cartan–Weyl extension $(I = +, -)$. This shows that the information contained in the central charges are not affected by our procedure.

Using now the identity (A2)

$$\overline{Q}_{\dot{a}}^J = \epsilon^{JK} \overline{(Q_a^K)}, \tag{2.10}$$

the relations (2.9) become

$$\begin{aligned} \{Q_a^I, \overline{(Q_b^J)}\} &= -2\sigma_{a\dot{b}}^{\mu} P_{\mu} \delta_J^I, \\ \{Q_a^I, Q_b^J\} &= 2iz\epsilon_{ab}\epsilon^{IJ}, \\ \{\overline{(Q_a^I)}, \overline{(Q_b^J)}\} &= 2i\bar{z}\epsilon_{ab}\epsilon^{IJ}. \end{aligned} \quad (0) \tag{2.11}$$

In the absence of central charges $z = \bar{z} = 0$ (the bridge is broken), the algebra (0) splits into two independent (but dual) $N = 1$ complex supersymmetric algebras that turn out to be the generalization of the Fayet hypersymmetry as we shall see in more detail later.

We now come to the construction of the corresponding representations. Indeed, as we have mentioned earlier, the knowledge of the representations of the $(-)$, $(+)$, and (0) algebras leads to a deep understanding of the harmonic superspace. This also clarifies the massive hypermultiplet theories with and without central charges, and finally, it provides much information on the SU(2) symmetry combining the two supersymmetries.

III. MASSIVE REPRESENTATIONS WITHOUT CENTRAL CHARGES

The supersymmetric algebra (2.11) reduces to

$$\begin{aligned} \{Q_a^I, \overline{Q_b^J}\} &= -2\sigma_{a\dot{b}}^{\mu} P_{\mu} \delta_J^I, \\ &\text{all others are equal to zero,} \end{aligned} \tag{3.1}$$

which is, roughly speaking, isomorphic to $N = 1 \otimes N = 1$ complex ($N = 1$ quaternionic) supersymmetry or equivalently $N = 2 \otimes N = 2$ real supersymmetry. In order to construct the on-shell one-particle massive representations of the algebra (3.1), we shall adopt the following strategy. First, we construct those of the $(-)$ subalgebra. The representations of the $(+)$ subalgebra are easily deduced from the previous ones by duality. Finally, we discuss the case of the full algebra (3.1).

A. Representation of the (−) subalgebra

The method to follow is quite well known and was used by Salam and Strathdee in the early days of supersymmetry.[17] It is based on the Wigner method of induced representations. Indeed, when the Poincaré group is reduced to its little group O(3) $[p_\mu = (-m,0)]$, the supersymmetry algebra (3.1) with $I = J = -$ becomes

$$\{Q_a^-, \overline{Q_b^-}\} = 2m\delta_{ab} . \tag{3.2}$$

This relation defines a Clifford algebra. The operators Q_a^- and $\overline{Q_a^-}$ are, respectively, annihilation and creation operators of a spin-$\frac{1}{2}$ and a Cartan–Weyl charge (-1).

For the construction of the spinorial representations of (3.2), one introduces the Clifford vacuum $|c\rangle$,

$$|c\rangle = |\{\delta\},q\rangle , \tag{3.3}$$

satisfying the Clifford condition

$$Q_a^-|c\rangle = 0 , \tag{3.4}$$

where $\{\delta\}$ represents a collection of quantum numbers. We shall restrict ourselves hereafter to the case $\{\delta\} = j$: the spin. The Clifford vacuum (3.3) is chosen charged q times ($q \in \mathbb{Z}$). The introduction of this charge, which is natural, will play a crucial role in our construction. This will permit us to classify the field contents of any $N = 2$ supersymmetric multiplet in SU(2) representations. Furthermore, it gives an insight on the possible theories in the harmonic superspace. Before going ahead, we assume, in the following, the nondegeneracy of the charge q, i.e., charges of type $q = (q+p) - p$, $p \in \mathbb{Z}$, are not taken into account. The degeneracy leads simply to adding auxiliary representations that are irrelevant on shell.

The full set of orthogonal states, denoted $R^-(j,q)$, is given by

$$R^-_{(j,q)} = \begin{cases} |j,q\rangle , \\ \overline{Q_a^-}|j,q\rangle = |j \pm \tfrac{1}{2}, q-1\rangle , \\ \overline{Q^{-2}}|j,q\rangle = |j,q-2\rangle . \end{cases} \tag{3.5}$$

This expression defines a unitary supersymmetric representation of the $(-)$ subalgebra. It is characterized by the mass m, the spin j, and the Cartan Weyl charge q.

Furthermore, if we define the $U^c(1)$ charge number of $R^-(j,q)$ as

$$\nu_{B,F} = \sum q_{n_{B,F}} \left(\sum n_{B,F} \right)^{-1} , \tag{3.6}$$

where n_B and n_F are, respectively, the number of bosonic, fermionic states, $q_{n_{B,F}}$ the charges of the corresponding states. One sees from the construction (3.5) that

$$\nu_B = \nu_F \tag{3.7}$$

for any supersymmetric representation.

Knowing also that a state of charge q carries a SU(2) symmetric representation[6] of dimension $(|q|+1)$, one can verify that bosonic and fermionic degrees of freedom match. Indeed, from (3.5), one finds

$$\begin{aligned} &(2j+1)[(q+1)+(q-1)] \\ &\quad = [2(j-\tfrac{1}{2})+1]q + [2(j+\tfrac{1}{2})+1]\,q . \end{aligned} \tag{3.8}$$

This means that there are $2(2j+1)q$ bosonic and $2(2j+1)q$ fermionic degrees of freedom. Therefore the dimension of the irreducible representation $R^-(j,q)$ is

$$\dim R^-(j,q) = 2^2|q|(2j+1) , \quad q > 1 , \tag{3.9a}$$

which reduces, for $j = 0$, to

$$\dim R^-(0,q) = 2^2|q| . \tag{3.9b}$$

As one sees, the relation (3.9a) is not valid for $q = 0$. However, in order to make it valid for $q = 1$, we have to double the dimension. These two cases will be considered later on in some detail.

The variation of the spin in the supersymmetric multiplet (3.5) is just as in the $N = 1$ real case: $\Delta J = 1$. However, the dimension (3.9) is larger than the usual $N = 1$ real one. To be more clear let us explore the properties of (3.5) through different examples.

Example 1: Fundamental multiplet, $j = 0$,

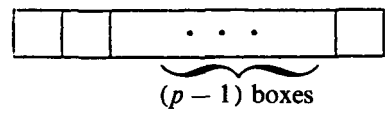
$$R^-(0,q) = \begin{cases} |0,q\rangle , \\ |\tfrac{1}{2}, q-1\rangle , \\ |0,q-2\rangle , \end{cases} \tag{3.10}$$

which can be represented for different values of q as seen in Table I. From Table I one learns

$$R^-_{\uparrow}(0,q) = \{0^{q+1}, \tfrac{1}{2}^q, 0^{q-1}\} , \quad q > 1 , \tag{3.11a}$$

$$R^-_{\downarrow}(0,q) = \{0^{-q+1}, \tfrac{1}{2}^{-q+2}, 0^{-q+3}\} , \quad q < 1 , \tag{3.11b}$$

where the power p of the spin j, j^p, in $R^-_{\uparrow}$ and $R^-_{\downarrow}$ represents the number of states of spin j belonging to the SU(2) symmetric representation $(p-1)$,



. (3.12)

We also note that the two representations (3.11a) and (3.11b) are symmetric with respect to the value $q = 1$, i.e.,

$$\begin{aligned} R^-_{\uparrow}(0,-q+2) &= R^-_{\downarrow}(0,q) , \\ R^-_{\downarrow}(0,-q-2) &= R^-_{\uparrow}(0,q) . \end{aligned} \tag{3.13}$$

Furthermore, one reads from (3.11)

$$R^-(0,1) = \{0^2, \tfrac{1}{2}D, 0^2\} , \tag{3.14}$$

where the symbol $\frac{1}{2}D$ means a Dirac spinor instead of Weyl spinor denoted as $\frac{1}{2}$. This can be justified by many arguments. The natural one is to demand the equality of bosonic and fermionic degrees of freedom. Another argument is just to use formally the result (3.9) which gives

TABLE I. The fundamental multiplet, $j = 0$.

q	$R^-(0,q)$			Multiplets
−3	0^4	$\frac{1}{2}^5$	0^6	$(0^{10},\frac{1}{2}^5)$
−2	0^3	$\frac{1}{2}^4$	0^5	$(0^8,\frac{1}{2}^4)$
−1	0^2	$\frac{1}{2}^3$	0^4	$(0^6,\frac{1}{2}^3)$
0	0^1	$\frac{1}{2}^2$	0^3	
1	0^2	$\frac{1}{2}D$	0^2	$(0^4,\frac{1}{2}^2)$
2	0^3	$\frac{1}{2}^2$	0^1	
3	0^4	$\frac{1}{2}^3$	0^2	$(0^6,\frac{1}{2}^3)$
4	0^5	$\frac{1}{2}^2$	0^3	$(0^8,\frac{1}{2}^4)$
5	0^6	$\frac{1}{2}^5$	0^4	$(0^{10},\frac{1}{2}^5)$

$$R^{-}(0,1)=\{0^2,\tfrac{1}{2}\}\,. \tag{3.15a}$$

However, in order to avoid zero norm states, one has to double the degrees of freedom (complex Clifford vacuum). This leads to the complex representation

$$R^{-}(0,1)=\{0^2+i0^2,\tfrac{1}{2}D\}\sim(0^4,\tfrac{1}{2}^2)\,, \tag{3.15b}$$

known in the literature as the FS representation. This multiplet is in fact the original one discovered by Fayet.[12] (This hypermultiplet had also been obtained by Salam and Strathdee.) We shall call it the Fayet hypermultiplet throughout this paper. It is a massive $N=2$ scalar multiplet without central charges ($z=\bar{z}=0$).

From Table I one also discovers that the $N=2$ massive scalar multiplet ($z=\bar{z}=0$) $(0^4,\tfrac{1}{2}^2)$ is described by other representations, namely,

$$\begin{aligned} R_{\uparrow}^{-}(0,0)&=\{0^1,\tfrac{1}{2}^2,0^3\}\,,\\ R_{\downarrow}^{-}(0,0)&=\{0^3,\tfrac{1}{2}^2,0^1\}\,. \end{aligned} \tag{3.16}$$

Even though the two representations are obtained from two different Clifford vacua (singlet or triplet), they describe the same multiplet. It is known as the HST hypermultiplet.[11] As in (3.15a), this hypermultiplet suffers since it has zero norm states (doublet of Weyl spinors). One can avoid this difficulty by complexifing (3.16) as we have done in (3.15b).

The remaining representations are associated with Clifford vacua in higher SU(2) symmetric representations.

Now before examining the next example, let us make the following remark. Write

$$\begin{aligned} R_{\uparrow}^{-}(0,2)&=(0^4,\tfrac{1}{2}^2)\,,\\ R_{\uparrow}^{-}(0,3)&=(0^6,\tfrac{1}{2}^3)\,,\\ R_{\uparrow}^{-}(0,5)&=(0^{10},\tfrac{1}{2}^5)\,. \end{aligned} \tag{3.17}$$

Then we have

$$R_{\uparrow}^{-}(0,5)=R_{\uparrow}^{-}(0,2)+R_{\uparrow}^{-}(0,3)\,, \tag{3.18}$$

or more generally

$$R_{\uparrow}^{-}(0,q)=R_{\uparrow}^{-}(0,p)+R_{\uparrow}^{-}(0,s)\,, \tag{3.19}$$

with $p+s=q$ $(p>0,s>0)$ and $p+s=q-2$ $(p<0,s<0)$.

However, the two representations $R_{\uparrow}^{-}(0,q)$ and $R_{\uparrow}^{-}(0,p)+R_{\uparrow}^{-}(0,s)$, although they describe the same multiplet, are different as long as the SU(2) symmetry is taken into account.

Example 2: The vector multiplet, $j=\tfrac{1}{2}$,

$$R^{-}(\tfrac{1}{2},q)=\begin{cases}|\tfrac{1}{2},q\rangle\,,\\ |0,q-1\rangle\,,\ |1,q-1\rangle\,,\\ |\tfrac{1}{2},q-2\rangle\,.\end{cases} \tag{3.20}$$

The variation of the equation in $R^{-}(\tfrac{1}{2},q)$ is now $\Delta j=1$ (see Table II). Similarly as for the fundamental multiplet, we have

$$\begin{aligned} &R_{\uparrow}^{-}(\tfrac{1}{2},q)=\{\tfrac{1}{2}^{q+1},0^q,1^q,\tfrac{1}{2}^{q-1}\}\,,\quad q>1\,,\\ &R_{\downarrow}^{-}(\tfrac{1}{2},-q)=\{\tfrac{1}{2}^{-q+1},0^{-q+2},1^{-q+2},\tfrac{1}{2}^{-q+3}\}\,,\quad q<1\,,\\ &R^{-}(\tfrac{1}{2},1)=\{\tfrac{1}{2}^2,0_D,1_D,\tfrac{1}{2}^2\}\,. \end{aligned} \tag{3.21}$$

One can check that

$$\begin{aligned} &\dim R_{\uparrow}^{-}(\tfrac{1}{2},q)=2^3|q|\,,\quad q\geqslant 2\,,\\ &\dim R^{-}(\tfrac{1}{2},1)=\dim R^{-}(\tfrac{1}{2},2)=2^4\,. \end{aligned} \tag{3.22}$$

TABLE II. The vector multiplet, $j=\tfrac{1}{2}$.

q		$R^{-}(\tfrac{1}{2},q)$			Multiplets
-1	$\tfrac{1}{2}^2$	0^3	1^3	$\tfrac{1}{2}^4$	$(0^3,\tfrac{1}{2}^6,1^3)$
0	$\tfrac{1}{2}^1$	0^2	1^2	$\tfrac{1}{2}^3$	
1	$\tfrac{1}{2}^2$	0_D	1_D	$\tfrac{1}{2}^2$	$(0^2,\tfrac{1}{2}^4,1^2)$
2	$\tfrac{1}{2}^3$	0^2	1^2	$\tfrac{1}{2}^1$	
3	$\tfrac{1}{2}^4$	0^3	1^3	$\tfrac{1}{2}^2$	$(0^3,\tfrac{1}{2}^6,1^3)$

Also, from the identity (3.13) that holds for any spin j, namely,

$$R_{\downarrow}^{-}(j,q)=R_{\uparrow}^{-}(j,-q+2)\,, \tag{3.23}$$

one finds

$$\dim R^{-}(\tfrac{1}{2},0)=\dim R^{-}(\tfrac{1}{2},1)=16\,. \tag{3.24}$$

This is in fact not surprising since all these three massive, $z=0$ vector multiplets (3.21) describe the same massive gauge multiplet. It corresponds, as it should, to an $N=1$ complex vector multiplet.[17] Furthermore, the previous discussion for the fundamental multiplet ($j=0$) still holds here. Indeed, the vector multiplet $(0^2,\tfrac{1}{2}^4,1^2)$ can be put either in a complex or in a real SU(2) representation. They are given in our formalism by the following.

(i) $R^{-}(\tfrac{1}{2},1)$ = complex representation that is the extension of the Fayet representation for the scalar hypermultiplet (3.14). Requiring the reality condition, the Fayet hypersymmetry reduces to the usual $N=1$ real supersymmetry.

(ii) $R_{\uparrow}^{-}(\tfrac{1}{2},2)\sim R_{\downarrow}^{-}(\tfrac{1}{2},0)$. This is the extension of the HST representation. This multiplet, however, suffers since it has zero norm states. The real bosonic states are put in SU(2) doublets. This problem is connected with the statistics of the Clifford vacuum of the representation $R^{-}(j,q)$. This difficulty could be removed by doubling the modes of $R^{-}(\tfrac{1}{2},0)$ as in (3.15b). This procedure might also be connected with $N=1$ quaternionic supersymmetry ($N=4$ real supersymmetry). Indeed, we have the following situation:

scalar multiplet	vector multiplet
$N=1$; $(0^2,\tfrac{1}{2})$	$N=2$; $(\tfrac{1}{2},0^2,1^2,\tfrac{1}{2}^3)$
$N=2$; $(0^2+0^2,\tfrac{1}{2}D)$	$N=4$; $(\tfrac{1}{2}D,0^2+0^2,1^2+1^2,\tfrac{1}{2}^3+\tfrac{1}{2}^3)$ $\sim(\tfrac{1}{2}_{\frac{1}{2}},0^4,1^4,\tfrac{1}{2}^6)$ of SO(4).

(3.25)

Example 3: The gravitino multiplet, $j=1$ (see Table III). This hypermultiplet is analogous to the fundamental hypermultiplet. The Fayet and HST representations are well adapted for the gravitino multiplet (Table III). More gener-

TABLE III. The gravitino multiplet, $j=1$.

q		$R^{-}(1,q)$			Multiplet
0	1	$\tfrac{1}{2}^2$	$\tfrac{3}{2}^2$	1^3	
1	1^2	$\tfrac{1}{2}D$	$\tfrac{3}{2}D$	1^2	$(\tfrac{1}{2}^2,1^4,\tfrac{3}{2}^2)$
	1^3	$\tfrac{1}{2}^2$	$\tfrac{3}{2}^2$	1	

ally, this is valid for all bosonic Clifford vacua. For the fermionic Clifford vacua, one deals with a situation analogous to the vector multiplet (3.21). In summary, the irreducible massive representations of the (−) subalgebra without central charges are given by

$$R_{\uparrow}^{-}(j,q)=\{j^{q+1};\ (j\pm\tfrac{1}{2})^{q};\ j^{q-1}\},\quad q>1,$$
$$R_{\downarrow}^{-}(j,q)=\{j^{-q+1};\ (j+\tfrac{1}{2})^{-q+2};\ j^{-q+3}\},\quad q<1,$$
$$R^{-}(j,1)=\{j^{2};\ (j+\tfrac{1}{2})D;\ j^{2}\},\qquad (3.26)$$

satisfying the following properties:

$$R_{\uparrow}^{-}(j,-q+2)=R_{\uparrow}^{-}(j,q),$$
$$\dim R_{\uparrow}^{-}(j,q)=2^{2}q(2j+1),\qquad (3.27)$$
$$\dim R^{-}(j,1)=\dim R^{-}(j,2).$$

Furthermore, for each representation, the charge number ν is conserved. For the interesting cases, $q=1$ and $q=0$ ($q=2$), we get the extension of the Fayet and HST hypermultiplets for any spin j, namely,

$$\begin{aligned}&(1)\quad R^{-}(j,1)=\text{generalized Fayet representation},\\ &\left.\begin{aligned}&(2a)\ R^{-}(j,0)\\ &(2b)\ R^{-}(j,2)\end{aligned}\right\}=\text{generalized HST representations}.\end{aligned}\qquad (3.28)$$

Finally, we note that in order to avoid zero norm states, these representations should be complexified.

B. Case of the (+) subalgebra

The Clifford algebra is written in this case as

$$\{Q_{a}^{+},\overline{Q_{b}^{+}}\}=2,m\delta_{ab}\,.\qquad (3.29)$$

This is obtained from (3.2) by conjugating the Cartan–Weyl charge (A2). Therefore the corresponding representations may be deduced from those of the (−) subalgebra (3.22). They are given by

$$R^{+}(j,q)=[R^{-}(j,q)]^{*},\qquad (3.30)$$

or equivalently, using (A1) and (3.27),

$$R_{\uparrow}^{+}(j,q)=R_{\downarrow}^{-}(j,-q),\quad -q<1,$$
$$R_{\downarrow}^{+}(j,q)=R_{\uparrow}^{-}(j,-q),\quad -q>1.\qquad (3.31)$$

One can easily check this result by giving an example. Indeed, using the previous technique, one finds for the fundamental multiplet the information contained in Table IV, which turns out to be identical to (3.11). This result is predicted by the $U^{c}(1)$ conjugates (3.31). Therefore, the Cartan–Weyl conjugation (*) introduced in Ref. 6 connects the representations of the (−) and (+) subalgebras. This can also be seen in (A3).

TABLE IV. The fundamental multiplet in the case of the (+) subalgebra.

q	$R^{+}(0,q)$			Multiplets
−3	0^{4}	$\tfrac{1}{2}^{3}$	0^{2}	$(0^{6},\tfrac{1}{2}^{3})$
−2	0^{3}	$\tfrac{1}{2}^{2}$	0^{1}	
−1	0^{2}	$\tfrac{1}{2}D$	0^{2}	$(0^{4},\tfrac{1}{2}^{2})$
0	0^{1}	$\tfrac{1}{2}^{2}$	0^{3}	
1	0^{2}	$\tfrac{1}{2}^{3}$	0^{4}	$(0^{6},\tfrac{1}{2}^{3})$

C. Representations of the full algebra

In the rest frame, the (0) algebra (3.1) becomes

$$\{Q_{a}^{I},\overline{(Q_{b}^{J})}\}=2m\delta_{ab}\delta_{J}^{I},$$
$$\text{all others are equal to zero.}\qquad (3.32)$$

This algebra is an $N=2$ extended complex supersymmetric algebra. Its representations are larger than the previous ones. They are obtained by following the same technique presented previously.

Define the Clifford vacuum $|c\rangle=|j,q\rangle$. The Clifford condition on $|C\rangle$ requires

$$Q^{I}|C\rangle=0,$$

which collects both Clifford conditions for the (−) and (+) subalgebras. The set of orthogonal states may be written as

$$\left\{\begin{array}{|c|}\hline |j\cdot q\rangle\\ |j-\tfrac{1}{2},q+1\rangle\\ |j+\tfrac{1}{2},q+1\rangle\\ |j,q+2\rangle\\ \hline\end{array}\quad\begin{array}{|c|}\hline |j-\tfrac{1}{2},q-1\rangle\\ |j-1,q\rangle\\ |j,q\rangle\\ |j-\tfrac{1}{2},q+1\rangle\\ \hline\end{array}\quad\begin{array}{|c|}\hline |j+\tfrac{1}{2},q-1\rangle\\ |j,q\rangle\\ |j+1,q\rangle\\ |j+\tfrac{1}{2},q+1\rangle\\ \hline\end{array}\quad\begin{array}{|c|}\hline |j,q-2\rangle\\ |j-\tfrac{1}{2},q-1\rangle\\ |j+\tfrac{1}{2},q-1\rangle\\ |j,q\rangle\\ \hline\end{array}\right\}.\qquad (3.33)$$

This consists essentially of four $N=1$ complex representations of type $R^{\pm}(j,q)$. The charge number ν defined earlier is still conserved. The variation of the spin in $R(j,q)$ is $\Delta j=2$ for $j\geqslant 1$ and $\Delta j=1$ for $j=0$. The dimension of (2.35) is given by the sum of its four subrepresentations.

Let us now illustrate some general features of these representations with examples.

Example 1: $j=0$.

The representation $R(j,q)$ reduces to a 3×3 table ($\Delta j=1$),

$$R(0,0)=\left\{\begin{array}{|c|}\hline 0^{1}\\ \tfrac{1}{2}^{2}\\ 0^{3}\\ \hline\end{array}_{R}\quad\begin{array}{|c|}\hline \tfrac{1}{2}^{2}\\ 0_{D},1_{D}\\ \tfrac{1}{2}^{2}\\ \hline\end{array}_{F}\quad\begin{array}{|c|}\hline 0^{3}\\ \tfrac{1}{2}^{2}\\ 0^{1}\\ \hline\end{array}_{R}\right\}.\qquad (3.34)$$

This complex multiplet $(0^{10},\tfrac{1}{2}^{8},1^{2})$ describes two $N=2$ massive real multiplets $(0^{5},\tfrac{1}{2}^{4},1)$. The subscripts R and F mean HST and Fayet-type $N=1$ multiplets discussed previously:

$$\dim R(0,0) = 8 + 16 + 8 = 32 = 2^5\,, \quad R(0,1) = \left\{ \begin{array}{|c|}\hline 0^2 \\ \tfrac{1}{2}^3 \\ 0^4 \\ \hline \end{array} \quad \begin{array}{|c|}\hline \tfrac{1}{2}^1 \\ 0^2, 1^2 \\ \tfrac{1}{2}^3 \\ \hline \end{array}_R \quad \begin{array}{|c|}\hline 0^2 \\ \tfrac{1}{2}D \\ 0^2 \\ \hline \end{array}_F \right\}. \tag{3.35}$$

It is also a complex multiplet $(0^{12},\tfrac{1}{2}^9,1^2)$ that has to be doubled ($N=1$ quaternionic) in order to avoid zero norm states,

$$R(0,-1) = \left\{ \begin{array}{|c|}\hline 0^2 \\ \tfrac{1}{2}D \\ 0^2 \\ \hline \end{array}_F \quad \begin{array}{|c|}\hline \tfrac{1}{2}^3 \\ 0^2, 1^2 \\ \tfrac{1}{2}^1 \\ \hline \end{array}_R \quad \begin{array}{|c|}\hline 0^4 \\ \tfrac{1}{2}^3 \\ 0^2 \\ \hline \end{array} \right\} = R(0,1)\,. \tag{3.36}$$

Example 2: $j = \frac{1}{2}$.

$$R(\tfrac{1}{2},0) = \left\{ \begin{array}{|c|}\hline \tfrac{1}{2} \\ 0^2 \\ 1^2 \\ \tfrac{1}{2}^3 \\ \hline \end{array}_R \quad \begin{array}{|c|}\hline 0^2 \\ \tfrac{1}{2}D \\ \\ 0^2 \\ \hline \end{array}_F \quad \begin{array}{|c|}\hline 1^2 \\ \tfrac{1}{2}D \\ \tfrac{3}{2}D \\ 1^2 \\ \hline \end{array}_F \quad \begin{array}{|c|}\hline \tfrac{1}{2}^3 \\ 0^2 \\ 1^2 \\ \tfrac{1}{2} \\ \hline \end{array}_R \right\} \sim (0^8,\tfrac{1}{2}^2,1^8,\tfrac{3}{2}^2)\,, \quad \dim R(\tfrac{1}{2},0) = 2^6\,. \tag{3.37}$$

One also finds for $q = 1$

$$R(\tfrac{1}{2},1) = \left\{ \begin{array}{|c|}\hline \tfrac{1}{2}^2 \\ 0^3 \\ 1^3 \\ \tfrac{1}{2}^4 \\ \hline \end{array} \quad \begin{array}{|c|}\hline 0 \\ \tfrac{1}{2}^2 \\ \\ 0^3 \\ \hline \end{array}_R \quad \begin{array}{|c|}\hline 1 \\ \tfrac{1}{2}^2 \\ \tfrac{3}{2}^2 \\ 1^3 \\ \hline \end{array}_R \quad \begin{array}{|c|}\hline \tfrac{1}{2}^2 \\ 0_D \\ 1_D \\ \tfrac{1}{2}^2 \\ \hline \end{array}_F \right\} = R(\tfrac{1}{2},-1) \sim (0^9,\tfrac{1}{2}^{14},1^9,\tfrac{3}{2}^2)\,, \quad \dim R(\tfrac{1}{2},1) = 36 + 36 = 72\,. \tag{3.38}$$

More generally, we have

$$R_{\substack{1 \\ q>1}}(j,q) = \left\{ \begin{array}{|c|}\hline j^{q+1} \\ (j-\frac{1}{2})^{q+2} \\ (j+\frac{1}{2})^{q+2} \\ j^{q+3} \\ \hline \end{array} \quad \begin{array}{|c|}\hline (j-\frac{1}{2})^q \\ (j-1)^{q+1} \\ j^q \\ (j-\frac{1}{2})^{q+2} \\ \hline \end{array} \quad \begin{array}{|c|}\hline (j+\frac{1}{2})^q \\ j^{q+1} \\ (j+1)^{q+1} \\ (j+\frac{1}{2})^{q+2} \\ \hline \end{array} \quad \begin{array}{|c|}\hline j^{q-1} \\ (j-\frac{1}{2})^q \\ (j+\frac{1}{2})^q \\ j^{q+1} \\ \hline \end{array} \right\}, \tag{3.39}$$

$$R_{\substack{1 \\ q<-1}}(j,q) = \left\{ \begin{array}{|c|}\hline j^{-q+1} \\ (j-\frac{1}{2})^{-q} \\ (j+\frac{1}{2})^{-q} \\ j^{-q-1} \\ \hline \end{array} \quad \begin{array}{|c|}\hline (j-\frac{1}{2})^{-q+2} \\ (j-1)^{-q+1} \\ j^{-q+1} \\ (j-\frac{1}{2})^{-q} \\ \hline \end{array} \quad \begin{array}{|c|}\hline (j+\frac{1}{2})^{-q+2} \\ j^{-q+1} \\ (j+1)^{-q+1} \\ (j+\frac{1}{2})^{-q} \\ \hline \end{array} \quad \begin{array}{|c|}\hline j^{-q+3} \\ (j-\frac{1}{2})^{-q+2} \\ (j+\frac{1}{2})^{-q+2} \\ j^{-q+1} \\ \hline \end{array} \right\} = R_{\substack{1 \\ }}(j,-q) \tag{3.40}$$

and

$$R(j,0) = \left\{ \begin{array}{|c|}\hline j \\ (j-\frac{1}{2})^2 \\ (j+\frac{1}{2})^2 \\ j^3 \\ \hline \end{array}_R \quad \begin{array}{|c|}\hline (j-\frac{1}{2})^2 \\ (j-1)_D \\ j_D \\ (j-\frac{1}{2})^2 \\ \hline \end{array}_F \quad \begin{array}{|c|}\hline (j+\frac{1}{2})^2 \\ j_D \\ (j+1)_D \\ (j+\frac{1}{2})^2 \\ \hline \end{array}_F \quad \begin{array}{|c|}\hline j^3 \\ (j-\frac{1}{2})^2 \\ (j+\frac{1}{2})^2 \\ j \\ \hline \end{array}_R \right\}, \tag{3.41}$$

$$R(j,1) = \left\{ \begin{array}{|c|}\hline j^2 \\ (j-\frac{1}{2})^3 \\ (j+\frac{1}{2})^3 \\ j^3 \\ \hline \end{array} \quad \begin{array}{|c|}\hline (j-\frac{1}{2})_D \\ (j-1)^2 \\ j^2 \\ (j-\frac{1}{2})_D \\ \hline \end{array}_F \quad \begin{array}{|c|}\hline (j+\frac{1}{2})_D \\ j^2 \\ (j+1)^2 \\ (j+\frac{1}{2})_D \\ \hline \end{array}_F \quad \begin{array}{|c|}\hline j^2 \\ (j-\frac{1}{2})_D \\ (j+\frac{1}{2})_D \\ j^2 \\ \hline \end{array}_F \right\} = R(j,-1)\,. \tag{3.42}$$

Among the four subrepresentations in (3.41), there are two Fayet-type multiplets and two other types (HST). However, in (3.42) there are three FS-type multiplets; the fourth is a general one.

Finally, before examining the effect of central charge, one concludes from this study inspired from the HS techniques that massive Fayet and HST hypermultiplets without central charges (3.14) and (3.16) are well defined modulo a complexification. Therefore the original massless hypermultiplet actions obtained in Ref. 6 might be extended to complex massive ones without introducing central charges. This result is valid as long as the Cartan–Weyl charge is nondegenerate, i.e., the harmonic expansions are put on shell.[13,14]

IV. REPRESENTATIONS WITH CENTRAL CHARGES

In this section, we want to explore the representations of the full algebra with central charges (2.11) taking into account the Cartan–Weyl charges. The effect of central charges is well known. They reduce the variation of spin, Δj,[18] and are present only in massive theories. But let us first of all rewrite the algebra (2.11) in the rest frame,

$$\begin{aligned}
&\{Q^I_a, \overline{(Q^J_b)}\} = 2m\delta_{ab}\delta^I_J\,,\\
&\{Q^I_a, Q^J_b\} = 2iz\epsilon_{ab}\epsilon^{IJ}\,,\\
&\{\overline{Q}^I_{\dot a}, \overline{Q}^J_{\dot b}\} = 2i\bar z\epsilon_{\dot a\dot b}\epsilon^{IJ}\,.
\end{aligned} \tag{4.1}$$

Now, defining $z=\rho e^{2i\varphi}$ ($\rho>0$, $0\leqslant\varphi<\pi$) and making the phase transformations on the Q's,

$$Q^I_a = e^{i\varphi}\tilde Q^I_a\,,\quad \overline{Q^I_b} = e^{-i\varphi}\,\overline{\tilde Q^I_b}\,, \tag{4.2}$$

we get an analogous algebra to (4.1) where z reduces to $\rho=|z|$.

Now following Refs. 18 and 19 we introduce the following appropriate variable change:

$$\begin{aligned}
S_{a(r)} &= (1/\sqrt2)(\tilde Q^+_a - (-)^r_i\epsilon_{ab}\,\overline{\tilde Q^-_b})\\
&= (1/\sqrt2)(\tilde Q^+_a - (-)^r_i\epsilon_{ab}\tilde{\tilde{\overline{Q}}}{}^+_b)\sim S^+_{a(r)}\,,\\
\overline{S}_{a(r)} &= (1/\sqrt2)((-)^r_i\epsilon_{ab}\tilde Q^-_b + \overline{\tilde Q^+_a})\\
&= (1/\sqrt2)((-)^r_i\epsilon_{ab}\tilde Q^-_b - \tilde{\tilde{\overline{Q}}}{}^-_a)\sim \overline{S^+_{a(r)}}\,.
\end{aligned} \tag{4.3}$$

Then, the algebra (4.1) becomes diagonal:

$$\begin{aligned}
&\{S_{a(r)}, \overline{S}_{b(s)}\} = 2(m-(-)^r\rho)\delta_{rs}\delta_{ab}\,,\\
&\{S_{a(r)}, S_{b(s)}\} = 0\,,\quad \{\overline{S}_{a(r)}, \overline{S}_{b(s)}\} = 0\,.
\end{aligned} \tag{4.4}$$

The last equations of (4.4) are obvious since they do not preserve the Cartan–Weyl charge (2.9). The first equation defines a Clifford algebra and gives a lower limit on the mass of the representations

$$\rho\leqslant m<?\,. \tag{4.5}$$

The construction of the representations of this algebra is achieved in an analogous way as before. One postulates a Clifford vacuum of mass m, spin j, and $U^c(1)$ charge q:

$$|c\rangle = |j,q\rangle\,,\quad S_{a(r)}|c\rangle = 0\,. \tag{4.6}$$

Then one works out all the sets of orthogonal states. One finds

$$R(\rho\neq m,j,q) = \left\{ \boxed{\begin{matrix}|j,q\rangle\\ |j-\tfrac12,q+1\rangle\\ |j+\tfrac12,q+1\rangle\\ |j,q+2\rangle\end{matrix}} \quad \boxed{\begin{matrix}|j-\tfrac12,q+1\rangle\\ |j-1,q+2\rangle\\ |j,q+2\rangle\\ |j-\tfrac12,q+3\rangle\end{matrix}} \quad \boxed{\begin{matrix}|j+\tfrac12,q+1\rangle\\ |j,q+2\rangle\\ |j+1,q+2\rangle\\ |j+\tfrac12,q+3\rangle\end{matrix}} \quad \boxed{\begin{matrix}|j,q+2\rangle\\ |j-\tfrac12,q+3\rangle\\ |j+\tfrac12,q+3\rangle\\ |j,q+4\rangle\end{matrix}} \right\}, \tag{4.7}$$

and for $\rho=m$

$$R(\rho\neq m,j,q) = \begin{cases} |j,q\rangle\\ |j-\tfrac12,q+1\rangle\,,\\ |j+\tfrac12,q+1\rangle\,,\\ |j,q+3\rangle\,. \end{cases} \tag{4.8}$$

These representations share many properties with the expressions (3.35), (3.5), (3.27), and (3.31). Indeed, if one goes to the limit $\rho=0$, one recovers all the previous results since

$$R(\rho\to0,j,q) = R(j,q+2)\quad \text{in (3.33)}\,. \tag{4.9}$$

Furthermore, in the limit ρ goes to m, the Clifford algebra (4.4) reduces to an $N=1$ complex supersymmetry (Fayet hypersymmetry). This coincides with the $(+)$ subalgebra (2.8). The representations (4.8) are therefore the same as (3.27). For the particular case $j=0$, the multiplet (4.8) is written as

$$\begin{aligned}
&R_1(\rho=m,j,q) = \{0^{q+1},\tfrac12^{q+2},0^{q+3}\}\,,\quad q>-1\,,\\
&R_1(\rho=m,j,q) = \{0^{-q+1},\tfrac12^{-q+2},0^{-q+3}\}\,,\quad q<-1\,,
\end{aligned} \tag{4.10}$$

$$R(\rho=m,0,-1) = \{0^2,\tfrac12 D,0^2\}\,. \tag{4.11}$$

The representation (4.11) that corresponds to the Fayet hypermultiplet in presence of central charges was found by Sohnius.[10] It is known in the literature as the FS hypermultiplet. For this multiplet, the central charge ρ satisfies the constraint

$$P^2=\rho^2=z\bar z=m^2\,. \tag{4.12}$$

Furthermore, one learns from (4.10) that an analogous representation for the HST multiplet satisfying (4.12) is also present up to a complexification,

$$R_1(\rho=m,0,0) = \{0^1,\tfrac12{}^2,0^3\} = R_1(\rho=m,0,-2)\,, \tag{4.13}$$

which corresponds to the extension of the $N=2$ scalar multiplet $(0^4,\tfrac12{}^2)$ satisfying (4.12) to the HST sector.

One also learns from (4.7) and (4.8) that these representations (4.11) and (4.12) are generalizable for higher spin j; for instance, the vector multiplet is given by

$$R(\rho=m,\tfrac12,-1) = \{\tfrac12{}^2,0_D,1_D,\tfrac12{}^2\}\,, \tag{4.14}$$

which turns out to coincide with (3.21).

In summary, the representations of $N=2$ complex supersymmetric algebra (2.11) satisfying the constraint $P^2=\rho^2$ ($\rho>0$) are identical to those of ($+$), and subsequently ($-$), subalgebras. This fact generalizes the Sohnius result to any supermultiplet of spin j.

V. MASSLESS REPRESENTATIONS

The constraint (4.5) shows that the central charge is absent in massless theories. Therefore, the superalgebra (2.11) reduces to

$$\{Q_a^I, \overline{Q_b^J}\} = \begin{pmatrix} 1 & 0 \\ 0 & 0 \end{pmatrix}_{ab} \delta_J^I ,$$

$$\text{all others are equal to zero} , \tag{5.1}$$

where we have chosen $P=(-\frac{1}{4},0,0,\frac{1}{4})$. Equation (5.1) implies $Q_2^I=\overline{Q}_2^I=0$. Thus the resulting Clifford algebra reads as

$$\{Q^I,Q^J\}=\delta_J^I , \quad Q_1^I \equiv Q^I . \tag{5.2}$$

Similarly as before, we introduce a Clifford vacuum of helicity λ and a Cartan-Weyl charge q,

$$|c\rangle = |\lambda,q\rangle , \quad Q^I|c\rangle = 0 . \tag{5.3}$$

Acting successively with creation operators $\overline{(Q^I)}$ on $|c\rangle$, we get the following set of massless charged orthogonal states:

$$R(\lambda,q) = \begin{cases} |\lambda,q\rangle , & |\lambda+\frac{1}{2},q-1\rangle , \\ |\lambda+\frac{1}{2},q+1\rangle , & |\lambda+1,q\rangle , \end{cases} \tag{5.4}$$

where the doubling of helicities, required by CPT symmetry, is understood. The charge number ν is still conserved,

$$\nu_B = \nu_F . \tag{5.5}$$

Let us now examine some general features of (5.4) through examples. Indeed, adopting the same terminology as before, one has the information contained in Table V for the massless vector multiplet ($\lambda=0$). From Table V, one sees that the three representations describe the same object, namely,

$$(0^4,\tfrac{1}{2}^4,1^2) \sim 2\times(0^2,\tfrac{1}{2}^2,1) , \tag{5.6}$$

which turns out to be the $N=2$ complex gauge multiplet.

The two generalized HST representations $R(0,1)=R(0,-1)$ are not interesting in our particular case since they have zero norm states (real representations). However, the generalized Fayet representation $R(0,0)$ supplemented by the reality condition reproduces exactly the vector superfield $V^{++}=\overline{V}^{++}$ (Ref. 6). In the HS, this is given, in the Wess–Zumino gauge, by

$$\begin{aligned} V^{++} = {} & (i/\sqrt{2})\theta^{+2}C - (i/\sqrt{2})\bar{\theta}^{+2}\overline{C} + i\theta^{+}\sigma^{\mu}\bar{\theta}^{+}v_{\mu} \\ & + 2\theta^{+2}\bar{\theta}^{+}\bar{\lambda}^{-} + 2\bar{\theta}^{+2}\theta^{+}\lambda^{-} \\ & + \theta^{+2}\bar{\theta}^{+2}D^{--} , \end{aligned} \tag{5.7}$$

where D^{--} is an auxiliary field. The higher representations $R(0,q)$, $q>1$, introduce many gauge fields and are comparable to (3.27). Now, before examining other cases, let us remark that in the case $\lambda=-1$, one gets the same representations as (5.4) and (5.6). Indeed, one finds

$$R(-1,0) = \{-1_D, -\tfrac{1}{2}^2, -\tfrac{1}{2},0_D\} , \tag{5.8}$$

which is the CPT symmetric multiplet of $R(0,0)$. This is in fact more general since one can check from (5.4) that

$$\text{CPT}\, R(\lambda,q) = R(-\lambda,\tfrac{1}{2},q) . \tag{5.9}$$

The scalar multiplet, $\lambda=-\frac{1}{2}$, recovers the massless $N=2$ scalar hypermultiplet. The associated table is given by Table VI. The first observation one makes from Table VI is that these representations are CPT invariant. This can also be seen from (5.10) since

$$\text{CPT}\, R(-\tfrac{1}{2},q) = R(-\tfrac{1}{2},q) . \tag{5.10}$$

This is the only CPT self-invariant multiplet. Furthermore, one also reads from (5.11) that the Fayet and HST representations are still present for the massless hypermultiplets. These are exactly the analytic superfields discussed in Ref. 6. In the harmonic superspace, these hypermultiplets are given by

$$\begin{aligned} R(-\tfrac{1}{2},0) \equiv \phi^{+} = {} & \varphi^{+} + \sqrt{2}\theta^{+}\psi + \sqrt{2}\bar{\theta}^{+}\bar{\chi} \\ & + \theta^{+2}F^{-} + \bar{\theta}^{+2}G^{-} \\ & + \theta^{+}\sigma^{\mu}\bar{\theta}^{+}A_{\mu} + \cdots , \end{aligned} \tag{5.11}$$

$$\begin{aligned} R(-\tfrac{1}{2},1) \equiv \Omega = \overset{*}{\overline{\Omega}} = {} & \omega + \sqrt{2}\theta^{+}\zeta \\ & + \sqrt{2}\bar{\theta}^{+}\bar{\zeta}^{-} + \theta^{+2}H^{--} \\ & + \bar{\theta}^{+2}\overline{H}^{--} + i\theta^{+}\sigma^{\mu}\bar{\theta}^{+}B_{\mu}^{--} + \cdots . \end{aligned} \tag{5.12}$$

However, the additional terms can be disregarded if the harmonic expansions are put on shell.[13,14] This corresponds, as we have already mentioned, to a nondegenerate Cartan–Weyl charge.

We now come to our last example concerning the gravitational multiplet, $\lambda=1$. (See Table VII.) The $N=2$ supergravity multiplet shares many features with the fundamental and vector multiplets. It might be described, in the HS, by a generalized Fayet representation satisfying the reality condition. A formulation of this hypermultiplet in terms of analytic superfields has been proposed in Ref. 6. However, this

TABLE V. The massless vector multiplet, $\lambda=0$.

q	$R(0,q)$				Multiplet
-1	0^2+0^2,	$\frac{1}{2}^3$,	$\frac{1}{2}$,	1^2	$(0^4,\frac{1}{2}^4,1^2)$
0	0_D+0_D,	$\frac{1}{2}^2$,	$\frac{1}{2}^2$,	1_D	
1	0^2+0^2,	$\frac{1}{2}$,	$\frac{1}{2}^3$,	1^2	

TABLE VI. The scalar multiplet, $\lambda=-\frac{1}{2}$.

q	$R(-\frac{1}{2},q)$				Multiplet
-1	$\frac{1}{2}^2$	0^3	0	$\frac{1}{2}^2$	$(0^4,\frac{1}{2}^2)$
0	$\frac{1}{2}D$	0^2	0^2	$\frac{1}{2}D$	
1	$\frac{1}{2}^2$	0	0^3	$\frac{1}{2}^2$	

TABLE VII. The gravitational multiplet, $\lambda = 1$.

q	$R(1,q)$				Multiplet
-1	1^2	$\frac{1}{2}^3$	$\frac{1}{2}$	2^2	$(1^2,\frac{1}{2}^4,2^2)$
0	1_d	$\frac{1}{2}^2$	$\frac{1}{2}^2$	2_D	
1	2^2	$\frac{1}{2}$	$\frac{1}{2}^3$	1^2	

description, related to the central charges, is not recovered by our study. This is because the $D = 4$ multiplet (5.15) is massless and therefore has no central charge.

More generally, the massless representations for a given λ and q are summarized as follows:

$$
\begin{aligned}
R_{\uparrow}(\lambda,q) &= \{\lambda^{q+1},(\lambda+\tfrac{1}{2})^q,(\lambda+\tfrac{1}{2})^{q+2},(\lambda+1)^{q+1}\}, \\
&\qquad q>0, \\
R_{\downarrow}(\lambda,q) &= \{\lambda^{-q+1},(\lambda+\tfrac{1}{2})^{-q+2}, \\
&\qquad (\lambda+\tfrac{1}{2})^{-q},(\lambda+1)^{-q+1}\}, \quad q<0 \\
&= R_{\uparrow}(\lambda,-q), \\
R(\lambda,0) &= \{\lambda_D,(\lambda+\tfrac{1}{2})^2+(\lambda+\tfrac{1}{2})^2,(\lambda+1)_D\},
\end{aligned}
\tag{5.13}
$$

where the helicity $\lambda = 0$ has to be doubled as a consequence of the CPT symmetry.

VI. CONCLUSION

In this paper, we have classified the content of the $N = 2$ complex supersymmetric representations in SU(2) multiplets. We have noted first that the SU(2) symmetry combining the two supersymmetries manifests itself as a conservation of the Cartan–Weyl charge. We have taken advantage of the presence of this charge to construct the representations of $N = 2$ complex supersymmetry. The procedure is based on the Wigner method of induced representations used by Salam and Strathdee in the early days of supersymmetry. The main difference is the presence of the Cartan–Weyl charge, which can be created and annihilated. As a result, we succeed in classifying all the content of $N = 1$ and $N = 2$ complex supersymmetry multiplets in SU(2) representations. From this classification, one learns that the so-called FS and complex HST representations are present for massive (massless) cases with and without central charges. Furthermore, these representations are also present in the case of higher spin j. For the generalized Fayet representations, they turn out to play an important role in the $N = 2$ HS. They describe the $N = 2$ (complex) matter multiplet, the $N = 2$ real gauge multiplet, and probably the $N = 2$ supergravity prepotential.

In this discussion, we have considered only nondegenerate Cartan–Weyl charges that correspond to an "on-shell" description of the harmonic expansions. A generalization to degenerate charges is possible. Finally, we expect that this procedure is also applicable to the $N = 3$ case and might give other information on the $N = 3$ off-shell procedure of Ref. 6.

ACKNOWLEDGMENTS

The author would like to thank Professor J. Strathdee, Professor M. Sohnius, and Professor J. A. Helayel-Neto for helpful discussions and Professor M. Sohnius and Professor M. Awada for reading the manuscript.

He would also like to thank Professor Abdus Salam, the International Atomic Energy Agency, and UNESCO for hospitality at the International Centre for Theoretical Physics, Trieste.

APPENDIX

1. Star and complex conjugation operations

We have

$$
\begin{aligned}
&(u_i^{\pm})^* = \pm u_i^{\mp}, \quad \overline{u_i^{\pm}} = \mp u^{\mp i}, \\
&\overline{u_i^{\pm}} = u^{\pm i}, \quad u_i^+ u_j^- - u_j^+ u_i^- = \epsilon_{ij},
\end{aligned}
\tag{A1}
$$

and using (A1) one can check

$$
\begin{aligned}
&(Q_a^I)^* = \epsilon^{IJ} Q_a^J, \quad (\bar{Q}_{\dot{a}}^I)^* = \epsilon^{IJ} \bar{Q}_{\dot{a}}^J, \\
&\overline{(Q_{\dot{a}}^I)} = \epsilon_{IJ} \bar{Q}_{\dot{a}}^J, \quad \overline{(\bar{Q}_{\dot{a}}^I)} = -\epsilon_{IJ} Q_a^J, \\
&\overline{(Q_a^I)} = \epsilon^{IJ}\epsilon_{JK} \bar{Q}_{\dot{a}}^K = \bar{Q}_{\dot{a}}^I, \\
&\overline{(\bar{Q}_{\dot{a}}^I)} = -\epsilon^{IJ}\epsilon_{JK} Q_a^K = -Q_a^I, \\
&\bar{Q}_{\dot{a}}^J = \epsilon^{JI}\,\overline{(Q_a^I)},
\end{aligned}
\tag{A2}
$$

with $I = +, -$ and $\epsilon^{+-} = \epsilon_{-+} = 1$.

2. Star and Hermitian conjugations of the $(\pm)$ subalgebras

We have

$$
\begin{aligned}
&(-)^* = (+), \quad (+)^* = -(-), \\
&(-)^+ = (-), \quad (+)^+ = (+), \\
&(-)^{*+} = (+), \quad (+)^{*+} = -(-).
\end{aligned}
\tag{A3}
$$

[1] J. Wess and B. Zumino, Nucl. Phys. B **70**, 89 (1974).
[2] J. Ellis, "Phenomenology of GUTS," LAPP TH48-TH8174; S. Diomopoulos and H. Georgi, Nucl. Phys. B **139**, 150 (1981).
[3] J. H. Schwarz, Caltech preprint 85/35, 1985.
[4] D. V. Nieuwenhuizen, in *Supergravity '81, Trieste* (Cambridge U.P., Cambridge, 1982).
[5] J. R. Haag, J. T. Kopuszanski, and M. Sohnius, Nucl. Phys. B **88**, 157 (1975).
[6] A. Galperin, E. Ivanov, S. Kalitzin, V. Ogievetsky, and E. Sokatchev, Class. Quantum Gravit. **1**, 469 (1984); ICTP, Trieste, preprint IC/84/43, 1984.
[7] A. Salam and J. Strathdee, Ann. Phys. (NY) **141**, 316 (1982).
[8] A. Galperin, E. Ivanov, V. Ogievetsky, and E. Sokatchev, Class. Quantum. Gravit. **2**, 601 (1985).
[9] A. Galperin, E. Ivanov, V. Ogievetsky, and E. Sokatchev, Class. Quantum. Gravit. **2**, 617 (1985).
[10] M. Sohnius, Nucl. Phys. B **138**, 109 (1978); Phys. Rep. **128** (1985).
[11] P. S. Howe, K. S. Stelle, and P. K. Townsend, Nucl. Phys. B **214**, 519 (1983).
[12] P. Fayet, Nucl. Phys. B **113**, 135 (1975); A. Salam and J. Strathdee, Nucl. Phys. B **97**, 293 (1975).
[13] A. E. Hassouni, T. Lhallabi, E. G. Oudgrhiri, and E. H. Saidi, Class. Quantum. Gravit. **5**, 287 (1988).
[14] T. Lhallabi and E. H. Saidi, ICTP, Trieste, preprint IC/86/58, 1986.
[15] A. Salam and J. Strathdee, Nucl. Phys. B **80**, 499 (1974).
[16] E. H. Saidi, Int. J. Mod. Phys. A **2**, 625 (1987).
[17] A. Salam and J. Strathdee, Nucl. Phys. B **84**, 127 (1974).
[18] J. Wess and J. Bagger, *Supersymmetry and Supergravity* (Princeton U.P., Princeton, 1983).
[19] S. Ferrara and C. Savoy, in *Supergravity '81, Trieste* (Cambridge U.P., Cambridge, 1982).

On the symmetric representations of SU(5). Matrix elements of the generators in the subgroup bases SU(2)×SU(2)×U(1) and SU(2)×SU(2)

J. Vanthournout, H. De Meyer,[a] and G. Vanden Berghe
Seminarie voor Wiskundige Natuurkunde, Rijksuniversiteit Gent, Krijgslaan 281-S9, B9000 Gent, Belgium

(Received 18 November 1987; accepted for publication 18 May 1988)

Matrix elements of the group generators for the totally symmetric irreducible representations of SU(5) are obtained in closed form employing the decomposition chain $SU(5)\supset SU(4)\times U(1)\supset SO(4)\times U(1)$. The $SU(4)\approx SU(2)\times SU(2)$ subgroup herein also occurs at the tail of the inclusion chain $SU(5)\supset SO(5)\supset SO(4)$. Therefore, closed form expressions for the matrix elements of the SU(5) generators in the latter basis are established, too.

I. INTRODUCTION

One of the very popular models for describing even–even nuclei is the interacting boson model[1] (IBM1) in which the physical states are associated with the representation states of a totally symmetric irreducible representation (irrep) of the unitary group SU(6). On physical grounds it is required that angular momentum and its projection are good quantum numbers, hence that each of the states is also a pure SO(3) representation state. There exist three group inclusion chains by which SU(6) is decomposable into the physical SO(3), namely, $SU(6)\supset SU(3)\supset SO(3)$, $SU(6)\supset SO(6)\supset SO(5)\supset SO(3)$, and $SU(6)\supset SU(5)\supset SO(5)\supset SO(3)$. In the present paper and in a forthcoming one we shall fix our attention to the last chain and in particular to the totally symmetric representations of SU(5).

In the context of IBM1 our ultimate interest consists in the construction of independent SO(3) scalars in the enveloping algebras associated to any one of the so-called dynamical symmetry algebras SU(3), SO(6), or SU(5). Such scalars can serve as symmetry preserving perturbation terms which, when added to the model Hamiltonian, lead to a closer reproduction of various observable nuclear data. Hence, it is of great importance to establish first closed form expressions for the matrix elements in the physical SO(3) basis of the dynamical symmetry group generators under consideration. This program has so far been successfully realized in the case of the SU(3) limit[2] and that of the SO(6) limit.[3]

Here we are concerned with the remaining case of SU(5). Several years ago this problem had been treated by Chacon and Moshinsky.[4,5] These authors had first constructed an SO(3) state basis in which the states are explicitly realized as functions of five collective coordinates (three being the Euler angles and the two remaining ones denoted as usual by β and γ). Next, they have demonstrated that the problem of finding the matrix elements of the SU(5) generators becomes equivalent to the calculation of certain reduced Wigner coefficients in the chain $SO(5)\supset SO(3)$. In the present paper, the same problem is reconsidered and subsequently solved using the method that has been elaborated in our study of the SO(6) limit[3] of IBM1. Indeed, the appearance of the SO(5) subgroup in the reduction chain certifies that it is efficient to establish first the matrix elements of the SU(5) generators in an $SU(2)\times SU(2)\approx SO(4)$ basis according to the chain $SU(5)\supset SO(5)\supset SO(4)$, hence an SO(4) basis in which the states carry one SO(5) label. Then, in a further stadium, the physical SO(3) states will be projected out of a restricted subset of so-called intrinsic $SU(2)\times SU(2)$ states[6,7] by means of the Hill–Wheeler technique. In order to calculate in closed form the matrix elements of the SU(5) generators in that SO(4) basis, we introduce another decomposition chain, namely, $SU(5)\supset SU(4)\times U(1)\supset SO(4)\times U(1)\supset SO(4)$, and we establish a second SO(4) state basis, whereby the additional label is provided by the U(1) group. In the latter basis the matrix elements of the SU(5) generators are easily constructed on account of a complete set of recursion relations. Then, the return to the former SO(4) basis is realized by diagonalizing explicitly the SO(5) Casimir operator.

Our method to arrive at the matrix elements in the SO(3) basis over an intermediate $SO(4)\approx SU(2)\times SU(2)$ basis (a task that will be completed in a forthcoming paper) has the advantage that it is purely algebraical in the sense that there is no reference to a particular realization of the states. Furthermore, we obtain in this paper intermediate results that could have their importance outside the context of the interacting boson model.

The paper is outlined as follows. In Secs. II and III we present the tensorial decomposition of the SU(5) generators in the SO(4) bases respectively associated to the chains $SU(5)\supset SO(5)\supset SO(4)$ and $SU(5)\supset SU(4)\times U(1)\supset SO(4)$. The Lie algebra commutators are derived and the branching rules for symmetric irreducible SU(5) representations are discussed. In Sec. IV the matrix elements of the SU(5) generators in the $SO(4)\times U(1)$ basis are calculated. In the final section the transformation to the SO(5) labeled SO(4) basis is carried out.

II. THE CHAIN OF GROUPS SU(5)⊃SO(5)⊃SO(4)

It is well known that according to the reduction chain $SU(5)\supset SO(5)\supset SO(3)$, the SU(5) Lie algebra is generated by the operators l_μ $(-1\leqslant\mu\leqslant 1)$ that constitute a basis of the principal SO(3) subalgebra, together with, respectively, the components G^2_μ $(-2\leqslant\mu\leqslant 2)$ of a five-dimensional SO(3) tensor representation, the components G^3_μ $(-3\leqslant\mu\leqslant 3)$ of seven-dimensional SO(3) tensor representation, and the components G^4_μ $(-4\leqslant\mu\leqslant 4)$ of a nine-di-

[a] Research Director at the National Fund for Scientific Research (N.F.W.O.), Belgium.

 0022-2488/88/091958-06$02.50

mensional SO(3) tensor representation.[8,9] By making the identification $l_\mu = \sqrt{10}G^1_\mu$, the complete set of SU(5) Lie algebra commutators is obtained from the single formula[10]

$$\begin{aligned}\left[G^{k_1}_{\kappa_1}, G^{k_2}_{\kappa_2}\right] = \sum_{k_3,\kappa_3} &[(2k_1+1)(2k_2+1)(2k_3+1)]^{1/2} \\ &\times \begin{pmatrix} k_1 & k_2 & k_3 \\ \kappa_1 & \kappa_2 & -\kappa_3 \end{pmatrix} \\ &\times (-1)^{-\kappa_3}[(-1)^{k_1+k_2+k_3}-1] \\ &\times \begin{Bmatrix} k_1 & k_2 & k_3 \\ 2 & 2 & 2 \end{Bmatrix} G^{k_3}_{\kappa_3}. \end{aligned} \tag{2.1}$$

The subset of the ten tensor components G^1_μ, G^3_ν generates the SO(5) subalgebra.

On the other hand, according to the reduction scheme $SU(5)\supset SO(5)\supset SU(2)\times SU(2)\approx SO(4)$, the adjoint irrep of SU(5) decomposes into the $SU(2)\times SU(2)$ reps (1,0), (0,1), (1,1), (0,0), and twice $(\frac{1}{2},\frac{1}{2})$. Hence denoting by s_μ $(-1\leqslant\mu\leqslant 1)$ and t_μ $(-1\leqslant\mu\leqslant 1)$ the generators of both SU(2) groups which satisfy, respectively,

$$\begin{aligned} &[s_0, s_{\pm 1}] = \pm s_{\pm 1}, \quad [s_{-1}, s_{+1}] = s_0, \\ &[t_0, t_{\pm 1}] = \pm t_{\pm 1}, \\ &[t_{-1}, t_{+1}] = t_0, \quad [s_\mu, t_\nu] = 0 \quad (-1\leqslant\mu,\nu\leqslant 1), \end{aligned} \tag{2.2}$$

the remaining SU(5) generators should be represented as the components $V^{1,1}_{\mu,\nu}$ $(-1\leqslant\mu,\nu\leqslant 1)$, $Z^{0,0}_{0,0}$, $T^{1/2,1/2}_{\mu,\nu}$ $(\mu,\nu\in\{-\frac{1}{2},+\frac{1}{2}\})$, and $U^{1/2,1/2}_{\mu,\nu}$ $(\mu,\nu\in\{-\frac{1}{2},+\frac{1}{2}\})$ of four distinctive $SU(2)\times SU(2)$ tensor representations. Clearly, the following nonzero coupled commutator relations must hold:

$$\begin{aligned} &[\mathbf{s},T]^{1/2\,1/2} = [\mathbf{t},T]^{1/2\,1/2} = -(\sqrt{3}/2)T, \\ &[\mathbf{s},U]^{1/2\,1/2} = [\mathbf{t},U]^{1/2\,1/2} = -(\sqrt{3}/2)U, \\ &[\mathbf{s},V]^{1\,1} = [\mathbf{t},V]^{1\,1} = -\sqrt{2}V. \end{aligned} \tag{2.3}$$

Let us remember that for any SO(3) or SU(2) tensors X^{k_1} and Y^{k_2} of respective ranks k_1 and k_2 by definition[11]

$$(X^{k_1}Y^{k_2})^k_\mu = \sum_{\mu_1,\mu_2} \langle k_1\mu_1 k_2\mu_2|k\mu\rangle X^{k_1}_{\mu_1} Y^{k_2}_{\mu_2} \tag{2.4}$$

and

$$\begin{aligned}[X^{k_1}, Y^{k_2}]^k_\mu &= \sum_{\mu_1,\mu_2} \langle k_1\mu_1 k_2\mu_2|k\mu\rangle [X^{k_1}_{\mu_1}, Y^{k_2}_{\mu_2}] \\ &= (X^{k_1}Y^{k_2})^k_\mu - (-1)^{-k_1-k_2+k}(Y^{k_2}X^{k_1})^k_\mu . \end{aligned} \tag{2.5}$$

Notice that in order to simplify notations we omit the tensor superscripts. The commutator properties given so far provide sufficient information for establishing the exact relationship between the two SU(5) realizations. After straightforward calculations one arrives at

$$\begin{aligned} &s_0 = \tfrac{3}{10}l_0 - (1/\sqrt{10})G^3_0, && t_0 = \tfrac{1}{10}l_0 + (3/\sqrt{10})G^3_0, \\ &s_{\pm 1} = G^3_{\pm 3}, && t_{\pm 1} = \tfrac{1}{5}l_{\pm 1} + \sqrt{(3/5)}G^3_{\pm 1}, \\ &T_{\pm 1/2,\pm 1/2} = \pm\sqrt{2}G^3_{\pm 2}, && T_{\pm 1/2,\mp 1/2} = \mp((\sqrt{3}/5)l_{\pm 1} - (2/\sqrt{5})G^3_{\pm 1}), \\ &U_{\pm 1/2,\pm 1/2} = \sqrt{(2/7)}(2G^2_{\pm 2} + \sqrt{3}G^4_{\pm 2}), && U_{\pm 1/2,\mp 1/2} = -\sqrt{(2/7)}(G^2_{\pm 1} - \sqrt{6}G^4_{\pm 1}), \\ &Z_{0,0} = \sqrt{(2/7)}(G^2_0 - (3/\sqrt{5})G^4_0), && V_{0,0} = (1/\sqrt{14})(3G^2_0 + \sqrt{5}G^4_0), \\ &V_{\pm 1,0} = G^4_{\pm 3}, && V_{0,\pm 1} = (1/\sqrt{7})(\sqrt{6}G^2_{\pm 1} + G^4_{\pm 1}), \\ &V_{\pm 1,\pm 1} = G^4_{\pm 4}, && V_{\pm 1,\mp 1} = -(1/\sqrt{7})(\sqrt{3}G^2_{\pm 2} - 2G^4_{\pm 2}). \end{aligned} \tag{2.6}$$

The distinction which has been made between T and U is such that they are the T components which close together with the generators of $SO(4)\approx SU(2)\times SU(2)$ into the SO(5) subalgebra. It is now also straightforward to derive from (2.6) with the help of (2.1) the remaining nontrivial commutators of the $SU(2)\times SU(2)$ tensor components, which again in coupled form can be summarized as follows:

$$\begin{aligned} &[T,T]^{1\,0} = -[U,U]^{1\,0} = -2\mathbf{s}, \\ &[T,T]^{0\,1} = -[U,U]^{0\,1} = -2\mathbf{t}, \\ &[T,U]^{0\,0} = 5Z, \quad [T,U]^{1\,1} = 2V, \\ &[T,V]^{1/2\,1/2} = -\tfrac{3}{2}U, \quad [U,V]^{1/2\,1/2} = -\tfrac{3}{2}T, \\ &[V,V]^{1\,0} = \sqrt{6}\mathbf{s}, \quad [V,V]^{0\,1} = \sqrt{6}\mathbf{t}, \\ &[Z,T]^{1/2\,1/2} = U, \quad [Z,U]^{1/2\,1/2} = T. \end{aligned} \tag{2.7}$$

In the SO(3) basis Hermiticity requirements imply that $(G^k_\mu)^\dagger = (-1)^\mu G^k_{-\mu}$ $(1\leqslant k\leqslant 4;\ -k\leqslant\mu\leqslant k)$. Correspondingly, one deduces in the SO(4) basis the properties

$$\begin{aligned} &T^\dagger_{\mu,\nu} = (-1)^{\mu+\nu}T_{-\mu,-\nu}, \\ &U^\dagger_{\mu,\nu} = (-1)^{\mu+\nu+1}U_{-\mu,-\nu}, \\ &V^\dagger_{\mu,\nu} = (-1)^{\mu+\nu}V_{-\mu,-\nu}, \quad Z^\dagger_{0,0} = Z_{0,0}. \end{aligned} \tag{2.8}$$

In the present work we are ultimately concerned with SU(5) as one of the dynamical symmetry groups of the SU(6) interacting boson model. Since any symmetric SU(6) irrep decomposes into totally symmetric SU(5) irreps, i.e.,

$$\begin{aligned} &SU(6)\to SU(5): \ [N,0,0,0,0] \to \sum_n [n,0,0,0], \\ &\qquad\qquad \text{with } n = N, N-1, \ldots, 0, \end{aligned} \tag{2.9}$$

we may confine ourselves to the problem of reducing the symmetric SU(5) irreps only. The decomposition into SO(5) irreps proceeds without degeneracies, namely,

$$\begin{aligned} &SU(5)\to SO(5): \ [n,0,0,0] \to \sum_\tau [\tau,0], \\ &\qquad\qquad \text{with } \tau = n, n-2, \ldots, 1, \text{ or } 0. \end{aligned} \tag{2.10}$$

Next, the further reduction into SO(4) irreps is prescribed by the rule[12]

$$\mathrm{SO}(5)\to\mathrm{SO}(4):\ [\tau,0]\to\sum_{s=t}(s,t),$$
$$\text{with } s=t=\tau/2,\tau/2-\tfrac{1}{2},\tau/2-1,\ldots,0. \tag{2.11}$$

Since no labels are missing the decomposition of symmetric SU(5) [or SU(6)] irreps is complete. Hence the orthonormal SU(5) basis states that make the SO(5) and SO(4) subalgebras apparent can be denoted as

$$|n,\tau,s,m_s,s,m_t\rangle \quad (m_s,m_t\in\{-s,-s+1,\ldots,s\}). \tag{2.12}$$

The matrix elements of the SO(5) generators expressed in the orthonormal basis (2.10) have been derived by Kemmer *et al.*[12] For the sake of completeness we list them here in the form of their SU(2)×SU(2) reduced equivalents:

$$(n,\tau',s',s'\|\mathbf{s}\|n,\tau,s,s)=(n,\tau',s',s'\|\mathbf{t}\|n,\tau,s,s)$$
$$=\delta_{\tau',\tau}\delta_{s',s}[s(s+1)]^{1/2}(2s+1), \tag{2.13}$$

$$\begin{aligned}(n,\tau',s',s'\|T\|n,\tau,s,s)&=\delta_{\tau',\tau}\{\delta_{s',s+1/2}[(\tau-2s)\\&\times(\tau+2s+3)(s+1)(2s+1)]^{1/2}\\&-\delta_{s',s-1/2}[(\tau-2s+1)\\&\times(\tau+2s+2)s(2s+1)]^{1/2}\}.\end{aligned} \tag{2.14}$$

It is one of the principal aims of the present paper to find, also in closed form, the corresponding reduced matrix elements of the tensors Z, U, and V. We can already predict that their action upon states of the type (2.10) will either shift the SO(5) label τ by ± 2 or will leave that label unchanged. This follows from the fact that the Z, U, and V components together constitute the symmetric SO(5) irrep [2,0]. Then, projection of the symmetric part out of the Kronecker product reduction[8]

$$\begin{aligned}[2,0]\times[\tau,0]&=[\tau-2,0]+[\tau,0]+[\tau+2,0]\\&+[\tau-2,2]+[\tau-2,4]+[\tau,2]\end{aligned} \tag{2.15}$$

proves the assertion.

For future use we mention the expressions of the quadratic SO(5) and SU(5) Casimir operators in terms of both the SO(3) and the SU(2)×SU(2) generator basis,[13]

$$\begin{aligned}C_{2,\mathrm{SO}(5)}&=-\sum_{k=1,3}\sum_{\mu=-k}^{k}(-1)^{\mu}G^{k}_{\mu}G^{k}_{-\mu}\\&=T_{1/2,1/2}T_{-1/2,-1/2}-T_{1/2,-1/2}T_{-1/2,1/2}\\&+s_0-\mathbf{s}^2-\mathbf{t}^2,\end{aligned} \tag{2.16}$$

$$\begin{aligned}C_{2,\mathrm{SU}(5)}&=-\sum_{k=1}^{4}\sum_{\mu=-k}^{k}(-1)^{\mu}G^{k}_{\mu}G^{k}_{-\mu}\\&=T_{1/2,1/2}T_{-1/2,-1/2}-T_{1/2,-1/2}T_{-1/2,1/2}\\&-U_{1/2,1/2}U_{-1/2,-1/2}+U_{1/2,-1/2}U_{-1/2,1/2}\\&-\sum_{\mu,\nu=-1}^{1}(-1)^{\mu+\nu}V_{\mu,\nu}V_{-\mu,-\nu}-\frac{5}{4}(Z_{0,0})^2\\&+2s_0-\mathbf{s}^2-\mathbf{t}^2,\end{aligned} \tag{2.17}$$

whereby for any vector $\mathbf{u}$ the notation $\mathbf{u}^2$ means $-2u_{+1}u_{-1}+u_0^2-u_0$. The Casimirs $C_{2,\mathrm{SO}(5)}$ and $C_{2,\mathrm{SU}(5)}$ are clearly also SU(2)×SU(2) invariants. Their eigenvalues are known to be $-\frac{1}{2}\tau(\tau+3)$ and $-\frac{4}{5}n(n+5)$, respectively.[13–15] Hence in the state basis (2.10) we immediately obtain their reduced matrix elements, i.e.,

$$(n,\tau',s',s'\|C_{2,\mathrm{SO}(5)}\|n,\tau,s,s)=-\tfrac{1}{2}\delta_{\tau',\tau}\delta_{s',s}\tau(\tau+3)(2s+1), \tag{2.18}$$

$$(n,\tau',s',s'\|C_{2,\mathrm{SU}(5)}\|n,\tau,s,s)=-\tfrac{4}{5}\delta_{\tau',\tau}\delta_{s',s}n(n+5)(2s+1). \tag{2.19}$$

The direct calculation of the matrix elements of the tensors U, V, and Z in the SU(2)×SU(2) basis seems to be a problem of insurmountable complexity. Instead, we shall proceed by first considering yet another basis that is associated to the inclusion SU(5)⊃SU(4)×U(1)⊃SO(4)×U(1)⊃SO(4).

III. THE CHAIN SU(5)⊃SU(4)×U(1)⊃SO(4)×U(1)⊃SO(4)

The operator $Z_{0,0}$ is an SO(4)≈SU(2)×SU(2) scalar that commutes with the two sets $\{s_\mu\}$, $\{t_\mu\}$ of SU(2) generators. Hence the set $\{s_\mu,t_\mu,Z_{0,0}\}$ is a generator basis of an SO(4)×U(1) subgroup of SU(5). Let us analyze the tensorial contents of the remaining SU(5) generators $T_{\mu,\nu}$, $U_{\mu,\nu}$, and $V_{\mu,\nu}$ with respect to this basis. First, we notice that from the commutator $[Z_{0,0},V_{\mu,\nu}]=0$, it follows that the SU(2)×SU(2) tensor V behaves as a U(1) scalar. [Clearly, with respect to SO(4)×U(1) we should write $V^{1/2,1/2,0}$.] Also, the set of U(1) scalar operators $\{s_\mu,t_\mu,V_{\mu,\nu}\}$ generates the SU(4) subgroup in the chain SU(5)⊃SU(4)×U(1)⊃SO(4)×U(1). The remaining generators $T_{\mu,\nu}$ and $U_{\mu,\nu}$ that are the components of two analogous bispinors under SU(2)×SU(2) can be linearly combined to form U(1) representations. We define

$$A_{\mu,\nu}=T_{\mu,\nu}+U_{\mu,\nu},\quad B_{\mu,\nu}=T_{\mu,\nu}-U_{\mu,\nu},\quad \forall\mu,\nu\in\{-\tfrac{1}{2},\tfrac{1}{2}\}, \tag{3.1}$$

and from (2.7) it follows that

$$[Z,A]^{1/2\ 1/2}=A,\quad [Z,B]^{1/2\ 1/2}=-B. \tag{3.2}$$

This proves that the components of the newly introduced SU(2)×SU(2) bispinors A and B behave as U(1) ladder operators when acting upon eigenstates of the U(1) generator $Z_{0,0}$. (In other terms, a completer notation should have been $A^{1/2,1/2,1}$ and $B^{1/2,1/2,-1}$). Since in the SU(5) algebra there are no tensorial components present that shift the additive U(1) label by more than unity, it follows immediately that the A components (B components) intercommute, which can also directly be verified by combining (3.1) with certain commutators in (2.7). One can easily derive all the commutators of the type $[A,B]$, $[A,V]$, and $[B,V]$, but it is irrelevant to tabulate here all these results explicitly. On the other hand, we shall make use further on of the Hermiticity properties

$$A^{\dagger}_{\mu,\nu}=(-1)^{\mu+\nu}B_{-\mu,-\nu},\quad \forall\mu,\nu\in\{-\tfrac{1}{2},\tfrac{1}{2}\}. \tag{3.3}$$

Since the five operators $Z_{0,0}$, $\mathbf{s}^2$, s_0, $\mathbf{t}^2$, and t_0 mutually commute, the SU(5) states can also be unambiguously labeled by their respective eigenvalues and we can establish an

SU(5) state basis of orthonormalized eigenstates of these operators. Let us denote such orthonormal states by

$$|n,z,s,m_s,s,m_t\} \quad m_s,m_t\in\{-s,-s+1,...,s\}. \tag{3.4}$$

As before, the fact that only symmetric SU(5) irreps are considered implies that $s=t$, where as it follows from (2.8) and (2.9) that s can take on the values $n/2,n/2-\frac{1}{2},n/2-1,...,\frac{1}{2},0$. There remains to find the range of the quantum number z defined by

$$Z_{0,0}\,|n,z,s,m_s,s,m_t\} = z|n,z,s,m_s,s,m_t\}. \tag{3.5}$$

This range easily follows from the branching of the symmetric irreps of SU(5) into irreps of SU(4)×U(1) and one finds the following:

$$\mathrm{SU}(5)\to\mathrm{SO}(4)\times\mathrm{U}(1):\ [n,0,0,0]\to\sum_{s,z}(s,s,z)$$
$$\text{with } z\in\{n/5,n/5-1,...,-4n/5+1,-4n/5\}$$
$$\text{and } s\in\{\tfrac{1}{2}(z+4n/5),\tfrac{1}{2}(z+4n/5)-1,...,\tfrac{1}{2}\text{ or }0\}. \tag{3.6}$$

IV. MATRIX ELEMENTS IN THE SU(2)×SU(2)×U(1) BASIS

We want to calculate the matrix elements of all the SU(5) generators in the state basis (3.4) where the s and z values are given by (3.6). For certain generators the result is immediate. Indeed, restricting once more to the reduced matrix elements one has

$$\{n,z',s',s'\|Z\|n,z,s,s\} = \delta_{z',z}\delta_{s',s}z(2s+1), \tag{4.1}$$

and

$$\{n,z',s',s'\|\mathbf{s}\|n,z,s,s\} = \{n,z',s',s'\|\mathbf{t}\|n,z,s,s\} = \delta_{z',z}\delta_{s',s}(2s+1)[s(s+1)]^{1/2}. \tag{4.2}$$

In order to calculate the reduced matrix elements of the tensors A, B, and V let us consider their nonzero commutators that follow from (3.1) and (2.7) and that are written in coupled form as

$$\begin{aligned}
&[A,B]^{0\,0} = -10Z, && [A,B]^{1\,1} = -4V,\\
&[A,B]^{1\,0} = -4\mathbf{s}, && [A,B]^{0\,1} = -4\mathbf{t},\\
&[A,V]^{1/2\,1/2} = -\tfrac{3}{2}A, && [B,V]^{1/2;1/2} = \tfrac{3}{2}B,\\
&[V,V]^{1\,0} = \sqrt{6}\mathbf{s}, && [V,V]^{0\,1} = \sqrt{6}\mathbf{t}.
\end{aligned} \tag{4.3}$$

Next, we shall frequently make use of the formula[16,17]

$$\langle\alpha' l'\|(X^{k_1}Y^{k_2})^k\|\alpha l\rangle = (-1)^{l+l'+k}(2k+1)^{1/2}\sum_{l'',\alpha''}\begin{Bmatrix}k_1 & k_2 & k\\ l & l' & l''\end{Bmatrix} \times\langle\alpha' l'\|X^{k_1}\|\alpha'' l''\rangle\langle\alpha'' l''\|Y^{k_2}\|\alpha l\rangle, \tag{4.4}$$

which expresses the reduced matrix elements of the vector coupling of two tensors X^{k_1} and Y^{k_2} in terms of the reduced matrix elements of these tensors. Also l represents an SO(3) or SU(2) representation label, whereas α denotes any set of labels that distinguish between states with the same l value.

It should be noticed that the property (3.3) is easily carried over into a relationship between reduced matrix elements, namely,

$$\{n,z+1,s',s'\|A\|n,z,s,s\}^* = -\{n,z,s,s\|B\|n,z+1,s',s'\} \quad (s'\in\{s+\tfrac{1}{2},s-\tfrac{1}{2}\}). \tag{4.5}$$

Similarly, it follows from (2.8) that

$$\{n,z,s',s'\|V\|n,z,s,s\}^* = \{n,z,s,s\|V\|n,z,s',s'\} \quad (s'\in\{s-1,s,s+1\}). \tag{4.6}$$

Let us for simplicity introduce the notations

$$\begin{aligned}
F(z,s) &= \{n,z-1,s+\tfrac{1}{2},s+\tfrac{1}{2}\|B\|n,z,s,s\},\\
G(z,s) &= \{n,z-1,s-\tfrac{1}{2},s-\tfrac{1}{2}\|B\|n,z,s,s\}.
\end{aligned} \tag{4.7}$$

On account of the branching rule (3.6) these functions are subject to certain boundary conditions, namely,

$$G(z,0) = F(z,\tfrac{1}{2}(z+4n/5)) = F^*(n/5+1,s) = G^*(n/5+1,s) = F^*(z,-\tfrac{1}{2}) = 0. \tag{4.8}$$

On the other hand, formula (4.4) enables us to establish recursion relations between the functions $F(z,s)$, $G(z,s)$, and the reduced matrix elements of the tensor V. As an example, let us take $X^{k_1}=A^{1/2,1/2}$ and $Y^{k_2}=A^{1/2,1/2}$, then it follows from (2.5) that $[A,A]^{1,0}=0=2(AA)^{1,0}$ and the application of (4.4) with $s'=s$ yields

$$-sG(z-1,s+\tfrac{1}{2})F(z,s) + (s+1)F(z-1,s-\tfrac{1}{2})G(z,s) = 0. \tag{4.9}$$

Similarly, the coupled commutator $[A,A]^{0,1}$ gives rise to the complex conjugate of (4.9). Two other relations of F and G alone are obtained from $[A,B]^{0,0}$ and $[A,B]^{1,0}$ (or equivalently $[A,B]^{0,1}$) each time with $s'=s$, namely,

$$|F(z,s)|^2 + |G(z,s)|^2 - |F(z+1,s-\tfrac{1}{2})|^2 - |G(z+1,s+\tfrac{1}{2})|^2 = 20z(2s+1)^2, \tag{4.10}$$

$$-s|F(z,s)|^2 + (s+1)|G(z,s)|^2 + (s+1)|F(z+1,s-\tfrac{1}{2})|^2 - s|G(z+1,s+\tfrac{1}{2})|^2 = 8s(s+1)(2s+1)^2. \tag{4.11}$$

In the supposition that the functions $F(z,s)$ and $G(z,s)$ are known already, the relations that follow $[A,B]^{1,1}$, chosing $s'=s$ and $s'=s+1$ (or $s'=s-1$), respectively,

$$-s^2|F(z,s)|^2 + s^2|G(z+1,s+\tfrac{1}{2})|^2 - (s+1)^2|G(z,s)|^2 + (s+1)^2|F(z+1,s-\tfrac{1}{2})|^2 = -8s(s+1)(2s+1)\{n,z,s,s\|V\|n,z,s,s\} \tag{4.12}$$

and

$$G^*(z,s+1)F(z,s) - F(z+1,s+\tfrac{1}{2})G^*(z+1,s+\tfrac{1}{2}) = 8(s+1)\{n,z,s+1,s+1\|V\|n,z,s,s\}, \tag{4.13}$$

can be used as defining equations for the reduced matrix elements of the tensor V. There are many other recursion relations following from the coupled commutators of the

type $[A,V]$, $[B,V]$, and $[V,V]$, but it is not relevant to list them here. Let us simply remark that the system of all the possible relations possesses up to some arbitrary phase factors a unique solution that satisfies the boundary conditions (4.8). After straightforward calculations one arrives at the following intermediate results:

$$\begin{aligned} F(z,s) &= 2[(s+1)(2s+1)(z+4n/5-2s) \\ &\quad \times (n/5+1-z)]^{1/2}, \\ G(z,s) &= 2[s(2s+1)(z+4n/5+2s+2) \\ &\quad \times (n/5+1-z)]^{1/2}. \end{aligned} \tag{4.14}$$

Taking into account (4.5), (4.7), (4.12), and (4.13), the reduced matrix elements in the $SO(4)\times U(1)$ basis of the tensors A, B, and V are brought into the following final forms:

$$\begin{aligned} &\{n,z',s',s'\|A\|n,z,s,s\} \\ &\quad = -2[(2s+1)(n/5-z)]^{1/2}\delta_{z',z+1} \\ &\qquad \times\{[(s+1)(z+4n/5+2s+4)]^{1/2}\delta_{s',s+1/2} \\ &\qquad + [s(z+4n/5-2s+2)]^{1/2}\delta_{s',s-1/2}\}, \end{aligned} \tag{4.15}$$

$$\begin{aligned} &\{n,z',s',s'\|B\|n,z,s,s\} \\ &\quad = 2[(2s+1)(n/5+1-z)]^{1/2}\delta_{z',z-1} \\ &\qquad \times\{[(s+1)(z+4n/5-2s)]^{1/2}\delta_{s',s+1/2} \\ &\qquad + [s(z+4n/5+2s+2)]^{1/2}\delta_{s',s-1/2}\}, \end{aligned} \tag{4.16}$$

$$\begin{aligned} &\{n,z',s',s'\|V\|n,z,s,s\} \\ &\quad = \tfrac{1}{2}\delta_{z',z}\{(2s+1)(z+4n/5+2)\delta_{s',s} \\ &\qquad + [(2s+1)(2s+3)(z+4n/5-2s) \\ &\qquad \times (z+4n/5+2s+4)]^{1/2}\delta_{s',s+1} \\ &\qquad + [(2s-1)(2s+1)(z+4n/5-2s+2) \\ &\qquad \times (z+4n/5+2s+2)]^{1/2}\delta_{s',s-1}\}. \end{aligned} \tag{4.17}$$

As a verification of these formulas we calculate the reduced matrix elements of the Casimir operator $C_{2,SU(5)}$. Expressing (2.17) in terms of the $SU(2)\times SU(2)\times U(1)$ tensor components and using (2.4) in reversed order it is found that

$$\begin{aligned} C_{2,SU(5)} &= (AB)^{0,0}_{0,0} - \tfrac{5}{4}(Z_{0,0})^2 \\ &\quad + 5Z_{0,0} - 3(VV)^{0,0}_{0,0} - \mathbf{s}^2 - \mathbf{t}^2. \end{aligned} \tag{4.18}$$

Next, we apply Eq. (4.4) again and substitute the matrix elements (4.15)–(4.17). Finally, it follows that

$$\{n,z',s',s'\|C_{2,SU(5)}\|n,z,s,s\} = -\tfrac{4}{5}\delta_{z',z}\delta_{s',s}n(n+5)(2s+1), \tag{4.19}$$

which is in accordance with (2.19) since $C_{2,SU(5)}$ is diagonal in any basis. For our purposes it is, however, more important to derive the matrix elements of the Casimir operators $C_{2,SO(5)}$ that is clearly not diagonal in the $SO(4)\times U(1)$ basis. Proceeding in the same way as for $C_{2,SU(5)}$ and by taking (2.16) as the starting point, we consecutively obtain

$$\begin{aligned} C_{2,SO(5)} &= \tfrac{1}{4}[(AA)^{0,0}_{0,0} + (BB)^{0,0}_{0,0} \\ &\quad + (AB)^{0,0}_{0,0} + (BA)^{0,0}_{0,0}] - \mathbf{s}^2 - \mathbf{t}^2 \end{aligned} \tag{4.20}$$

$$\begin{aligned} &\{n,z',s',s'\|C_{2,SO(5)}\|n,z,s,s\} \\ &\quad = \tfrac{1}{2}(2s+1)\delta_{s',s}\{\delta_{z',z+2}[(n/5-z)(n/5-z-1)(z+4n/5-2s+2)(z+4n/5+2s+4)]^{1/2} \\ &\qquad - \delta_{z',z}[(n/5-z)(z+4n/5+4) + (n/5-z+1)(z+4n/5) + 4s(s+1)] \\ &\qquad + \delta_{z',z-2}[(n/5-z+2)(n/5-z+1)(z+4n/5-2s)(z+4n/5+2s+2)]^{1/2}\}. \end{aligned} \tag{4.21}$$

V. MATRIX ELEMENTS IN THE SU(2)×SU(2) BASIS

We are now concerned with the transformation of the orthonormal $SO(4)\times U(1)$ basis (3.4) into the orthonormal $SO(4)$ basis (2.12), which makes the subgroup $SO(5)$ apparent. Let us notice that $Z_{0,0}$ is diagonal in the former basis and that $C_{2,SO(5)}$ is diagonal in the latter. But we dispose already of the matrix elements of $C_{2,SO(5)}$ in the basis (3.4). Hence by diagonalizing explicitly the operator $C_{2,SO(5)}$ we obtain the transformation matrix that carries one basis into the other. Since the same technique has been developed in our study of $SO(6)$, the reader is referred to Ref. 3 for more details. Following this method one arrives, after lengthy but straightforward calculations, at the matrix elements of $Z_{0,0}$, namely,

$$\begin{aligned} (n,\tau',s',s'\|Z\|n,\tau,s,s) &= (2s+1)\delta_{s',s}\{[(n-\tau)(n+\tau+5)(\tau-2s+1)(\tau-2s+2)(\tau+2s+3)(\tau+2s+4)/(2\tau+7) \\ &\quad \times (2\tau+5)^2(2\tau+3)]^{1/2}\delta_{\tau',\tau+2} + [(n-\tau+2)(n+\tau+3)(\tau-2s-1)(\tau-2s)(\tau+2s+1) \\ &\quad \times (\tau+2s+2)/(2\tau+3)(2\tau+1)^2(2\tau-1)]^{1/2}\delta_{\tau',\tau-2} \\ &\quad + (2n+5)[20s(s+1) - 3\tau(\tau+3)]/5(2\tau+1)(2\tau+5)\delta_{\tau',\tau}\}. \end{aligned} \tag{5.1}$$

From the coupled commutator $U = [Z,T]^{1/2\ 1/2}$, by means of formula (4.4) and by substituting the results (2.14) and (5.1), the matrix elements of the tensor U are established:

$$\begin{aligned} &(n,\tau',s',s'\|U\|n,\tau,s,s) \\ &\quad = [(s+1)(2s+1)]^{1/2}\delta_{s',s+1/2}\{-2[(n-\tau)(n+\tau+5)(\tau-2s+1)(\tau+2s+3) \\ &\qquad \times (\tau+2s+4)(\tau+2s+5)/(2\tau+7)(2\tau+5)^2(2\tau+3)]^{1/2}\delta_{\tau',\tau+2} \\ &\qquad + (2n+5)(4s+3)[(\tau-2s)(\tau+2s+3)]^{1/2}/(2\tau+1)(2\tau+5)\delta_{\tau',\tau} \end{aligned}$$

$$+2[(n-\tau+2)(n+\tau+3)(\tau-2s)(\tau-2s-1)(\tau-2s-2)(\tau+2s+2)/(2\tau+3)$$
$$\times(2\tau+1)^2(2\tau-1)]^{1/2}\delta_{\tau',\tau-2}\}+[s(2s+1)]^{1/2}\delta_{s',s-1/2}\{2[(n-\tau)(n+\tau+5)(\tau-2s+1)(\tau-2s+2)$$
$$\times(\tau-2s+3)(\tau+2s+3)/(2\tau+7)(2\tau+5)^2(2\tau+3)]^{1/2}\delta_{\tau',\tau+2}+(2n+5)(4s+1)[(\tau-2s+1)$$
$$\times(\tau+2s+2)]^{1/2}/(2\tau+1)(2\tau+5)\delta_{\tau',\tau}-2[(n-\tau+2)(n+\tau+3)(\tau-2s)(\tau+2s)(\tau+2s+1)$$
$$\times(\tau+2s+2)/(2\tau+3)(2\tau+1)^2(2\tau-1)]^{1/2}\delta_{\tau',\tau-2}\}. \tag{5.2}$$

Similarly, we derive from the coupled commutator $[T,U]^{1\,1}=2V$ that

$$(n,\tau',s',s'\|V\|n,\tau,s,s)$$
$$=-\tfrac{1}{2}[(2s+1)(2s+3)]^{1/2}\delta_{s',s+1}\{[(n\quad\tau)(n+\tau+5)(\tau+2s+3)(\tau+2s+4)(\tau+2s+5)$$
$$\times(\tau+2s+6)/(2\tau+7)(2\tau+5)^2(2\tau+3)]^{1/2}\delta_{\tau',\tau+2}$$
$$+(2n+5)[(\tau-2s)(\tau-2s-1)(\tau+2s+3)(\tau+2s+4)]^{1/2}/(2\tau+1)(2\tau+5)\delta_{\tau',\tau}$$
$$+[(n-\tau+2)(n+\tau+3)(\tau-2s)(\tau-2s-1)(\tau-2s-2)(\tau-2s-3)/(2\tau+3)$$
$$\times(2\tau+1)^2(2\tau-1)]^{1/2}\delta_{\tau',\tau-2}\}+\tfrac{1}{2}(2s+1)\delta_{s',s}\{[(n-\tau)(n+\tau+5)(\tau+2s+3)(\tau+2s+4)(\tau-2s+1)$$
$$\times(\tau-2s+2)/(2\tau+7)(2\tau+5)^2(2\tau+3)]^{1/2}\delta_{\tau',\tau+2}+(2n+5)[4s(s+1)+(\tau+1)(\tau+2)]/(2\tau+1)$$
$$\times(2\tau+5)\delta_{\tau',\tau}+[(n-\tau+2)(n+\tau+3)(\tau+2s+1)(\tau+2s+2)(\tau-2s-1)(\tau-2s)/(2\tau+3)$$
$$\times(2\tau+1)^2(2\tau-1)]^{1/2}\delta_{\tau',\tau-2}\}-\tfrac{1}{2}[(2s-1)(2s+1)]^{1/2}\delta_{s',s-1}\{[(n-\tau)(n+\tau+5)$$
$$\times(\tau-2s+1)(\tau-2s+2)(\tau-2s+3)(\tau-2s+4)/(2\tau+7)(2\tau+5)^2(2\tau+3)]^{1/2}\delta_{\tau',\tau+2}$$
$$+(2n+5)[(\tau-2s+2)(\tau-2s+1)(\tau+2s+1)(\tau+2s+2)]^{1/2}/(2\tau+1)(2\tau+5)\delta_{\tau',\tau}$$
$$+[(n-\tau+2)(n+\tau+3)(\tau+2s-1)(\tau+2s)(\tau+2s+1)(\tau+2s+2)/(2\tau+3)$$
$$\times(2\tau+1)^2(2\tau-1)]^{1/2}\delta_{\tau',\tau-2}\}. \tag{5.3}$$

It is clear that these results can be verified by considering other coupled commutators in (4.3). In this way we have done a lot of consistency tests.

ACKNOWLEDGMENTS

We want to thank Dr. J. Van der Jeugt (University of Southampton, U.K.) and Professor R. T. Sharp (McGill University, Montréal) for valuable discussions.

The National Fund for Scientific Research (Belgium) is acknowledged for their financial support which made it possible for Professor Sharp to stay one month at the University of Ghent.

[1]A. Arima and F. Iachello, Ann. Phys. (NY) **99**, 253 (1976); **111**, 201 (1978); **123**, 468 (1979).
[2]H. De Meyer, G. Vanden Berghe, and J. Van der Jeugt, J. Math. Phys. **26**, 3109 (1985).
[3]J. Vanthournout, J. Van der Jeugt, H. De Meyer, and G. Vanden Berghe, J. Math. Phys. **28**, 2529 (1987).
[4]E. Chacón and M. Moshinsky, J. Math. Phys. **17**, 668 (1976).
[5]E. Chacón and M. Moshinsky, J. Math. Phys. **18**, 870 (1977).
[6]D. L. Hill and J. A. Wheeler, Phys. Rev. **89**, 1102 (1953).
[7]S. A. Williams and D. L. Pursey, J. Math. Phys. **9**, 1230 (1968).
[8]B. G. Wybourne, *Symmetry Principles and Atomic Spectroscopy* (Wiley–Interscience, New York, 1970).
[9]W. G. McKay and J. Patera, *Table of Dimensions, Indices and Branching Rules for Representations of Simple Lie Algebras* (Dekker, New York, 1981).
[10]B. R. Judd, *Operator Techniques in Atomic Spectroscopy* (McGraw-Hill, New York, 1963).
[11]A. R. Edmonds, *Angular Momentum in Quantum Mechanics* (Princeton U.P., Princeton, 1957).
[12]N. Kemmer, D. L. Pursey, and S. A. Williams, J. Math. Phys. **9**, 1224 (1968).
[13]B. G. Wybourne, *Classical Groups for Physicists* (Wiley–Interscience, New York, 1974).
[14]Sun Hong Zhou, M. Moshinsky, A. Frank, and P. Van Isacker, Kinam **5**, 135 (1983).
[15]V. K. B. Kota, J. Van der Jeugt, H. De Meyer, and G. Vanden Berghe, J. Math. Phys. **28**, 1644 (1987).
[16]U. Fano and G. Racah, *Irreducible Tensorial Sets* (Academic, New York, 1959).
[17]A. de-Shalit and I. Talmi, *Nuclear Shell Theory* (Academic, New York, 1963).

Soliton and antisoliton interactions in the "good" Boussinesq equation

V. S. Manoranjan
Department of Chemical and Process Engineering, University of Surrey, Guilford, Surrey GU2 5HX, England

T. Ortega and J. M. Sanz-Serna
Departamento de Matemática Aplicada y Computación, Facultad de Ciencias, Universidad de Valladolid, Valladolid, Spain

(Received 5 November 1987; accepted for publication 4 May 1988)

The solitary-wave interaction mechanism for the good Boussinesq equation is investigated and found to be far more complicated than was previously thought. Three salient features are that solitary waves only exist for a finite range of velocities, that large solitons can turn into so-called antisolitons, and that it is possible for solitons to merge and split. Small solitons, however, appear to be stable. The existence of a potential well is linked to the different behaviors observed between small and large initial conditions.

I. INTRODUCTION

The importance of soliton-producing nonlinear wave equations is well understood among theoretical physicists and applied mathematicians. An equation that produces solitons and has received comparatively little attention in the literature is

$$u_{tt} = -u_{xxxx} + u_{xx} + (u^2)_{xx}. \tag{1.1}$$

This is referred to as the "good" Boussinesq equation (McKean[1]) or the nonlinear beam equation.[2] The related equation

$$u_{tt} = u_{xxxx} + u_{xx} + (u^2)_{xx},$$

known as the "bad" Boussinesq equation, has been studied by Hirota.[3] In a recent article, Manoranjan *et al.*[4] obtained a closed-form expression for the two soliton interactions of (1.1) and carried out numerical experiments to demonstrate the possibility of the breakup of an initial pulse into two solitons.

In this paper we show that the interaction mechanism is more complicated than that reported in Ref. 4. It turns out that when small amplitude solitons of (1.1) collide, they emerge from the nonlinear interaction with no change in shape or velocity. However, the large amplitude solitons change into so-called antisolitons as they come out from the interaction. We show that this difference in behavior is linked to the existence of a potential well for (1.1). Further, the existence of a local minimum for the potential energy enables us to investigate the existence of a class of solutions that remain bounded for all time, along with another class of solutions that blow up in finite time.

Throughout the paper our attention is confined to real-valued solutions u of (1.1) defined in $-\infty < x < \infty$.

II. PRELIMINARIES

A. Conservation laws

It is clear that if u is a smooth solution of (1.1) that vanishes, along with its derivatives, as $|x| \to \infty$, then the quantity

$$J(u) = \int_{-\infty}^{\infty} u_t \, dx$$

is an invariant of motion, whereas

$$I(u) = \int_{-\infty}^{\infty} u \, dx \tag{2.1}$$

varies in time as $I = Jt + \text{const}$. Since we are interested in solutions that remain bounded as t increases, we restrict our attention to initial conditions that satisfy $J = 0$. It is then expedient to introduce a new function w defined by

$$w(x,t) = \int_{-\infty}^{x} u_t(\xi,t) d\xi,$$

which also vanishes with its derivatives as $|x| \to \infty$. In terms of the functions u,w, the equation (1.1) becomes

$$w_t = -u_{xxx} + u_x + (u^2)_x, \quad u_t = w_x, \tag{2.2}$$

a system that conserves the functional I in (2.1) and the functionals M and E given by

$$M(w,u) = \int_{-\infty}^{\infty} wu \, dx, \tag{2.3}$$

and

$$E(w,u) = T(w) + V(u), \tag{2.4}$$

with

$$T(w) = \frac{1}{2}\int_{-\infty}^{\infty} w^2 \, dx, \tag{2.5}$$

and

$$V(u) = \int_{-\infty}^{\infty} \left(\frac{1}{2}u_x^2 + \frac{1}{2}u^2 + \frac{1}{3}u^3\right) dx. \tag{2.6}$$

Note that the system (2.2) is in Hamiltonian form and that E,T,V act as total energy, kinetic energy, and potential energy, respectively.

B. Solitons and antisolitons

We now look for traveling wave solutions of (2.2) of the form $w = w(\xi)$, $u = u(\xi)$, $\xi = x - ct$. With a prime denoting differentiation with respect to ξ, the system (2.2) then reads

$$-cw' = -u''' + (u^2)' + u', \quad -cu' = w'. \tag{2.7}$$

Elimination of w', followed by an integration in which the

 0022-2488/88/091964-05$02.50

integration constant must be zero in view of the boundary conditions for u, leads to

$$-c^2u = u'' - u^2 - u. \tag{2.8}$$

If $c^2 > 1$, then the origin $u = 0$, $u' = 0$ is a center in the phase plane of (2.8), and hence nontrivial solutions such that $u,u' \to 0$ as $|\xi| \to \infty$ are not possible. On the other hand, when $c^2 < 1$, the origin is a saddle point as depicted in the phase plane shown in Fig. 1, and the homoclinic trajectory $O \to A \to B \to O$ represents a soliton. The possible velocities $-1 < c < 1$ can be described in the form $c = \varepsilon(1 - P^2)^{1/2}$, where P is a parameter $0 < P \leqslant 1$, and $\varepsilon = +1$ or -1 determines whether the wave moves to the right or to the left. A simple integration in (2.8) yields the analytic form of the soliton as

$$u(\xi) = (-3P^2/2)\operatorname{sech}^2[(P/2)(\xi - \xi_0)]. \tag{2.9}$$

The (real) integration constant ξ_0 gives the initial location of the wave. It should be noted that the velocity c of the soliton is related to the amplitude $A = 3P^2/2$ by $A = \frac{3}{2}(1 - c^2)$. For the rightward- and leftward-traveling soliton of parameter P, the quantities in (2.1) and (2.3)–(2.6) are given by

$$I_{P,\varepsilon} = -6P, \tag{2.10}$$

$$M_{P,\varepsilon} = -6\varepsilon(1 - P^2)^{1/2}P^3, \tag{2.11}$$

$$E_{P,\varepsilon} = \tfrac{6}{5}P^3(5 - 4P^2), \tag{2.12}$$

$$T_{P,\varepsilon} = 3P^3(1 - P^2), \tag{2.13}$$

$$V_{P,\varepsilon} = \tfrac{3}{5}P^3(5 - 3P^2). \tag{2.14}$$

It should also be pointed out that the functions $w(\xi),u(\xi)$ corresponding to the soliton solve the variational problem

$$\delta E(w,u) = 0, \quad \text{subject to } M(w,u) = M_{P,\varepsilon}. \tag{2.15}$$

This can be verified by comparing the Euler–Lagrange equations of (2.15) with the system (2.7) satisfied by solitons.

Returning to the phase plane in Fig. 1 ($c^2 < 1$), and if we allow singular solutions, the trajectory $O \to C \to \infty \to D \to O$ also provides a traveling wave with velocity c, whose analytic form is found to be

$$u(\xi) = (3P^2/2)\operatorname{cosech}^2[(P/2)(\xi - \xi_0)]. \tag{2.16}$$

This singular solution, which has a double pole at $\xi = \xi_0$, will be referred to as antisoliton. It is helpful to combine Eqs. (2.9) and (2.16) into a single form,

$$u(\xi) = -6P^2e^{\eta}(1 + e^{\eta})^{-2}, \quad \eta = P(\xi - \xi_0) + i\pi\sigma, \tag{2.17}$$

where $i^2 = -1$, and $\sigma = 0$ gives the soliton and $\sigma = 1$ the antisoliton.

The possibility of traveling waves with velocities $c = +1$ or -1 has not been discussed so far. In these cases, a study of (2.8) reveals that there is a singular solution given by $u = 6(\xi - \xi_0)^{-2}$, but no solution of the soliton type exists. The singular solution just mentioned can also be obtained by taking in (2.16) the limit $P \to 0$, i.e., the limit $|c| \to 1$.

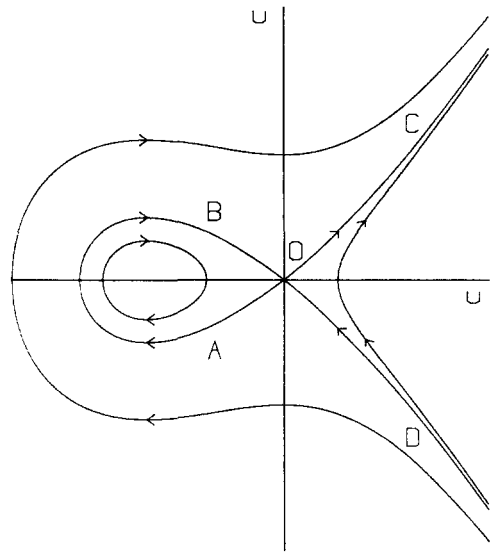


FIG. 1. Phase plane, $c^2 < 1$.

III. SOLITON AND ANTISOLITON INTERACTION

Following a technique used by Hirota,[3] Manoranjan *et al.*[4] constructed the family of solutions of (1.1) given by

$$u = -6(f_{xx}f - f_x^2)/f^2,$$

$$f(x,t) = 1 + \exp(\eta_1) + \exp(\eta_2) + a\exp(\eta_1 + \eta_2), \tag{3.1}$$

with

$$\eta_j = P_j[x - \varepsilon_j v_j t + x_j^0],$$

$$\varepsilon_j = \pm 1, \quad 0 < P_j \leqslant 1, \quad j = 1,2, \tag{3.2}$$

$$a = \frac{(\varepsilon_1 v_1 - \varepsilon_2 v_2)^2 - 3(P_1 - P_2)^2}{(\varepsilon_1 v_1 - \varepsilon_2 v_2)^2 - 3(P_1 + P_2)^2}, \tag{3.3}$$

and

$$v_j = (1 - P_j)^{1/2}, \quad j = 1,2. \tag{3.4}$$

Only real values of the phases $\eta_j, j = 1,2$, featuring in (3.2) were considered in Ref. 4, while in the present study we let η_j have the following complex form:

$$\eta_j = P_j[x - \varepsilon_j v_j t + x_j^0] + i\pi\sigma_j,$$

$$\varepsilon_j = \pm 1, \quad 0 < P_j \leqslant 1, \quad \sigma_j = 0,1, \quad j = 1,2. \tag{3.5}$$

With this choice of η_j, (3.1) is still a valid family of real-valued solutions u. We next show that (3.1) and (3.3)–(3.5) describe the exact interaction of solitons and antisolitons. To simplify matters we only discuss the case $\varepsilon_1 = 1$, $\varepsilon_2 = -1$, but the other possibilities in the choice of ε_j, $j = 1,2$, can be analyzed in a similar fashion (see Ortega[5]).

Elimination of f in (3.1) yields

$$\frac{u}{-6} = \frac{P_1^2e^{\eta_1} + P_2^2e^{\eta_2} + [a(P_1 + P_2)^2 + (P_1 - P_2)^2]e^{\eta_1 + \eta_2} + ae^{\eta_1 + \eta_2}(e^{\eta_1}P_2^2 + e^{\eta_2}P_1^2)}{[1 + e^{\eta_1} + e^{\eta_2} + ae^{\eta_1 + \eta_2}]^2}. \tag{3.6}$$

The behavior of (3.6) depends very much on the value of the number a in (3.3). Several cases must be considered.

(i) $0<a<\infty$. This corresponds to the interior of the regions I_a, I_b, I_c in Fig. 2. More specifically, $0<a<1$ in the interior of the regions I_b, I_c, while $1<a<\infty$ in the interior of the region I_a. By arguing as in Whitham (Ref. 6, Sec. 17.2.), it is found that as $t\to-\infty$, the solution (3.6) becomes just the linear superposition of two traveling waves of the form (2.17) with

$$\eta=\eta_{1-}=P_1[x-v_1t+x_1^0]+i\pi\sigma_1,$$
$$\eta=\eta_{2-}=P_2[x+v_2t+x_2^0+P_2^{-1}\log a]+i\pi\sigma_2.$$

As $t\to\infty$, the solution consists of the traveling waves

$$\eta=\eta_{1+}=P_1[x-v_1t+x_1^0+P_1^{-1}\log a]+i\pi\sigma_1,$$
$$\eta=\eta_{2+}=P_2[x+v_2t+x_2^0]+i\pi\sigma_2.$$

Thus the right moving wave (resp. the left moving wave) emerges from the interaction having a shift in position $\Delta x=P_1^{-1}\log a$ (resp. $\Delta x=-P_2^{-1}\log a$), but with no change in shape or velocity. Note that there are four choices for σ_1, σ_2, so that the interacting waves could be both solitons, both antisolitons, or a soliton along with an antisoliton. The right-going wave is shifted to the left ($\log a>0$) for P_1,P_2 in the region I_a and to the right in the regions I_b and I_c. The shift in the left-going wave occurs in the direction opposite of that in the right-going wave, in agreement with the fact that the total change in the phases η_1,η_2 must be zero (cf. Hirota[3]).

(ii) $a<0$. (Interior of region II.) Here the behavior of (3.6) as $t\to\infty$ is given by

$$\eta=\eta_{1-}=P_1[x-v_1t+x_1^0]+i\pi\sigma_1,$$
$$\eta=\eta_{2-}=P_2[x+v_2t+x_2^0+P_2^{-1}\log|a|]+i\pi\sigma_2^*,$$

where $\sigma_2^*=1-\sigma_2$. For $t\to\infty$ we have

$$\eta=\eta_{1+}=P_1[x-v_1t+x_1^0+P_1^{-1}\log|a|]+i\pi\sigma_1^*,\quad\sigma_1^*=1-\sigma_1,$$
$$\eta=\eta_{2+}=P_2[x+v_2t+x_2^0]+i\pi\sigma_2.$$

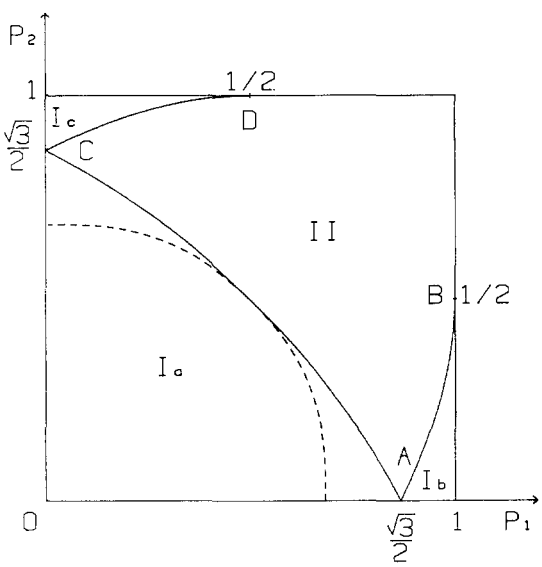


FIG. 2. Soliton and antisoliton interaction for the case $\varepsilon_1=1$, $\varepsilon_2=-1$.

It is apparent that now the waves not only undergo a shift but also change their nature; a soliton/antisoliton entering an interaction emerges as an antisoliton/soliton. Again four cases $\sigma_j=0,1, j=1,2$, are possible.

(iii) $a=0$. (Arcs AB and CD.) We consider first the arc AB where $P_1>P_2$. By taking limits in (3.6) along lines $x=mt+n$, m,n real constants, and comparing with (2.17), we find that as $t\to-\infty$ the solution consists of a single wave with phase

$$P_1[x-v_1t+x_1^0]+i\pi\sigma_1,$$

while for $t\to\infty$ there are two waves,

$$P_2[x-v_2t+x_2^0]+i\pi\sigma_2,$$
$$P_3[x-v_3t+P_3^{-1}(P_1x_1^0-\ldots)]+i\pi\sigma_3,$$

with

$$P_3=P_1-P_2,\quad v_3=(1-P_3^2)^{1/2},$$
$$\sigma_3=\sigma_1+\sigma_2(\mathrm{mod}\,2).$$

In obtaining the last outcoming wave, use must be made of the relation $(P_1-P_2)[1-(P_1-P_2)^2]^{1/2}=P_1v_1+P_2v_2$, which follows from $a=0$ in (3.3). These formulas mean that a single soliton ($\sigma_1=0$) can split into either two solitons ($\sigma_2=0$) or two antisolitons ($\sigma_2=1$). The outgoing waves move in opposite directions and the corresponding parameters satisfy $P_2+P_3=P_1$ [cf. (2.10) and the conservation of (2.1)]. On the other hand, an incoming antisoliton ($\sigma_1=1$) splits into either a left-going soliton and a right-going antisoliton or into a left-going antisoliton and a right-going soliton. It is perhaps useful to observe that if we represent by S and A the soliton and antisoliton, respectively, then the rules

$$S\to S+S,\quad S\to A+A,\quad A\to S+A,\quad A\to A+S,$$

familiar from Boolean algebra, govern the possible interactions.

The arc CD, where $P_2>P_1$, contains the merging of two incoming waves with parameters P_1, P_2-P_1 into a single wave with parameter P_2. The possible interactions can be represented as

$$S+S\to S,\quad A+A\to S,\quad S+A\to A,\quad A+S\to A.$$

(iv) $a=\infty$. (Arc AC.) This case corresponds either to waves of parameters P_1,P_2 merging into a single wave of parameter P_1+P_2, or to the splitting of a single wave with parameter P_1+P_2 into two waves with parameters P_1,P_2. The corresponding analysis can be performed by taking limits in (3.6) for $a\uparrow\infty$, or $a\downarrow-\infty$. However, taking limits directly in (3.6) results in the trivial solution $u=0$, and it is therefore necessary to make the parameters x_1^0 and/or x_2^0 functions of a before letting $|a|\to\infty$. For instance, we could take η_1 as given in (3.5) (x_1^0 a fixed constant) and

$$\eta_2=P_2[x+v_2t+m+P_2^{-1}\log|a|]+i\pi\sigma_2,\quad m\text{ constant}.$$

Although we have only considered two-wave interactions, N-wave interactions with a far more complicated dynamics are also possible and can be studied by using a formula analogous to that of Hirota.[3]

IV. THE POTENTIAL WELL

In the preceding sections we have allowed singular solutions, i.e., solutions with poles. For instance, the $S \to A + A$ splitting solution possesses two poles past the interaction time. In the rest of the paper we exclude this possibility and say that a solution ceases to exist at the time when poles develop. According to this point of view, standard in mathematical analysis, it has been shown in Sec. III that some of the solutions of (2.2) exist for all times whereas other solutions cease to exist at a finite value of t. This difference in behavior, numerically verified in Ref. 4, is explained in the next section in terms of a potential well studied below.

If v is a function of x, $-\infty < x < \infty$, we define the potential energy $V(v)$ as in formula (2.6). According to Sobolev's imbedding theorem (Adams[7]), the expression for $V(v)$ makes sense whenever v is in H^1, i.e., whenever v is square integrable with a square-integrable distributional derivative. It is clear that V can take arbitrarily large positive and negative values. The stationary points of the functional V are easily found to be given by the functions $v_1 = 0$ and $v_{2,b} = -\frac{3}{2}\operatorname{sech}^2[(x-b)/2]$, with b an arbitrary real constant. Note that in view of (2.9), the function $v_{2,b}$ provides the shape of the soliton with parameter $P = 1$, amplitude $A = \frac{3}{2}$, and velocity $c = 0$. This is in agreement with the fact that the functions for which the gradient of the potential vanishes give rise to time-independent solutions of the Hamilton equations (2.2).

The function $v_1 = 0$ provides a local minimum for the functional V, because the Sobolev inequality

$$\left|\int_{-\infty}^{\infty} v^3\,dx\right| \leqslant \int_{-\infty}^{\infty} |v|^3\,dx = \|v\|_{L^3}^3 \leqslant K^3\|v\|_{H^1}^3 = K^3\left(\int_{-\infty}^{\infty}(v^2 + v_x^2)dx\right)^{3/2} \tag{4.1}$$

(K is the imbedding constant) reveals that the cubic term in the potential is negligible in the H^1 neighborhood of $v_1 = 0$.

To study the behavior of V near $v = v_{2,b}$, we make the change $v = v^* + v_{2,b}$. The functional $V(v^* + v_{2,b}) - V(v_{2,b})$ is of the form $Q(v^*) + C(v^*)$, where

$$Q(v^*) = \int_{-\infty}^{\infty}\left[\frac{1}{2}v_x^{*2} + \frac{1}{2}v^{*2} - \frac{3}{2}\operatorname{sech}^2\frac{x-b}{2}\,v^{*2}\right]dx,$$

and C is cubic in v^*. The spectrum of Q is known in closed form (e.g., Whitham,[6] Sec. 17.5), and it turns out that the functional is indefinite, hence $v_{2,b}$ is a saddle point of the potential V.

The depth d of the potential well around the local minimum v_1 is defined by

$$d = \inf\left\{V(v): v \text{ in } H^1,\ v \neq 0,\ \int_{-\infty}^{\infty}[v_x^2 + v^2 + v^3]dx = 0\right\}. \tag{4.2}$$

Note that $d > 0$, because if $v \neq 0$ satisfies the integral constraint in (4.2), then

$$V(v) = \frac{1}{6}\|v\|_{H^1} = -\frac{1}{6}\int_{-\infty}^{\infty} v^3\,dx,$$

which, in view of (4.1), implies that $\|v\|_{H^1}$ [and hence $V(v)$] is bounded away from zero. If we denote the potential well by W, then W consists of functions v with potential energy below the depth d and satisfying the condition

$$-\int_{-\infty}^{\infty}[v_x^2 + v^2 + v^3]dx \leqslant 0. \tag{4.3}$$

When v possesses a square-integrable second derivative, an integration by parts of the v_x^2 term in (4.3) shows that this condition can be interpreted as the requirement that the angle (in the sense of the L^2 metric) between the function v and the force density $\delta V(v)/\delta v$ be obtuse. In symbols W is defined by

$$W = \{v \text{ in } H^1: V(v) < d \text{ and (4.3) holds}\}.$$

In view of (4.1), W is a neighborhood of the origin in the space H^1. Furthermore, for v in W,

$$\|v\|_{H^1}^2 = \int_{-\infty}^{\infty}[v_x^2 + v^2]dx = 6V(v) - 3\int_{-\infty}^{\infty}[v_x^2 + v^2 + v^3]dx \leqslant 6d, \tag{4.4}$$

so that W is bounded. The key issue is that if $w(t),u(t)$ is a solution of the system (2.2) that is smooth enough to conserve the energy $E(w,u)$ in (2.4)–(2.6), and such that initially $E(t=0) < d$ and $u(t=0)$ is in W, then $u(t)$ remains in W for all later times. In fact, if u were to leave the well at a time t_0, then at that time (4.3) would hold with $=$ instead of $\leqslant$, which, since E is below d, is in contradiction with (4.2).

The Euler–Lagrange equation for the constrained minimization problem in (4.2) is easily shown to have no solution other than the function $v_{2,b}$, where V takes the value $\frac{8}{3}$ [see (2.14)]. This makes it plausible that $d = \frac{8}{3}$ and that $v_{2,b}$ provides the mountain pass out of the well. In order to prove rigorously that this is the case, we would have to show that $v_{2,b}$ is not only a critical point for (4.2) but also a minimum, something we have not attempted. Note that, be that as it may, it is certainly true that $d \leqslant \frac{8}{3}$.

V. EXISTENCE AND NONEXISTENCE OF SOLUTIONS

Fourier analysis reveals that the linearization of (2.2) given by

$$w_t = -u_{xxx} + u_x, \quad u_t = w_x,$$

generates a strongly continuous semigroup in the space $L^2 \times H^1$. Since the nonlinear mapping $(w,u) \to ((u^2)_x, 0)$ is indefinitely continuously differentiable in that product space, the results on nonlinear semigroups of Segal[8] show that (2.2) has a generalized solution $(w(t),u(t))$ for each initial data $(w(0),u(0))$ in $L^2 \times H^1$. This solution exists for a positive length of time [depending on $(w(0),u(0))$] and is a continuous function of t and $(w(0),u(0))$. Furthermore, $(w(t),u(t))$ is smooth if $w(0)$ and $u(0)$ are. The functionals M and E in (2.3)–(2.6) are conserved by generalized solu-

tions, since these functionals are continuous in $L^2 \times H^1$ and conserved by smooth solutions. Therefore, if a generalized solution satisfies $E(0) < d$ and $u(0) \in W$, then $u(t)$ remains in the well throughout its interval of existence. Then (4.4) implies that $\|u\|_{H^1}$ remains bounded; hence the solution exists for all times $0 \leqslant t < \infty$ (Segal[8]). On the other hand, solutions for which $u(0)$ is not in W and solutions for which the energy E exceeds the well depth d are not likely to exist for all time. This is exemplified by the case where the initial condition consists of two solitons, well separated and moving towards each other. In Fig. 2 the dashed line represents the locus

$$\tfrac{6}{5}P_1^3(5 - 4P_1^2) + \tfrac{6}{5}P_2^3(5 - 4P_2^2) = \tfrac{6}{5} \geqslant d.$$

Thus, according to (2.12), $E > d$ to the right of the dashed line, a domain which includes region II, which was shown to correspond to $S + S \to A + A$ interactions and therefore exhibits blowup in finite time.

Sufficient conditions for the blowup to occur can be obtained by concavity arguments (Payne[9]). For instance, smooth solutions $(w(t), u(t))$ that vanish at $x = \pm \infty$, together with their derivative, cannot exist for all time if $E < 0$ and $I = 0$. Elimination of w in (2.2) reveals that such solutions satisfy

$$-D^{-2}u_{tt} = u_{xx} - u - u^2,$$

with

$$D^{-2} = \int_{-\infty}^{x}\int_{-\infty}^{x}.$$

Since the operator $-D^{-2}$ acting on the indicated class of functions is symmetric and positive definite, the results in Payne[9], Sec. 8, establish that the existence time of the solution is necessarily finite.

VI. DISCUSSION

We have shown that the "good" Boussinesq equation possesses a highly complicated mechanism for solitary wave interaction. Three salient features are that solitary waves exist only for a finite range of velocities, that interactions can alter the nature of the solitary waves, and that merging and splitting are possible. These properties may, no doubt, be of interest in modeling. In this connection it would be useful to carry out a stability analysis of the solitons of (2.2) (cf. Benjamin[10]). However, solitons with a large value of P are certainly orbitally unstable. This can be seen in Fig. 2, where it is apparent that perturbing the soliton of parameter $P_1 = \sqrt{3}/2$ with any other soliton, no matter how small, always leads to an $S + S \to A + A$ interaction, and, therefore, to blowup in finite time. Also, solitons with $P_1 > \sqrt{3}/2$ can split, under arbitrarily small perturbations, into either two solitons or two antisolitons with parameters P_2 and $P_1 - P_2$, where (P_1, P_2) lies in the arc AB. On the other hand, numerical evidence has led us to conjecture that small-amplitude solitons are orbitally stable. This conjecture can be proved for the periodic problem for (2.2). The details will be reported elsewhere.

[1]H. P. McKean, Commun. Pure Appl. Math. **34**, 599 (1981).
[2]V. E. Zakharov, Zh. Eksp. Teor. Fiz. **65**, 219 (1973).
[3]R. Hirota, J. Math. Phys. **14**, 810 (1973).
[4]V. S. Manoranjan, A. R. Mitchell, and J. L. L. Morris, SIAM J. Sci. Statist. Comput. **5**, 946 (1984).
[5]T. Ortega, Ph.D. thesis, Universidad de Valladolid, 1987.
[6]G. B. Whitham, *Linear and Nonlinear Waves* (Wiley–Interscience, New York, 1974).
[7]R. A. Adams, *Sobolev Spaces* (Academic, New York, 1975).
[8]I. Segal, Ann. Math. **78**, 339 (1963).
[9]L. E. Payne, *Improperly Posed Problems in Partial Differential Equations* (SIAM, Philadelphia, 1975).
[10]T. B. Benjamin, Proc. R. Soc. London Ser. A **328**, 153 (1972).

On a class of eigenfunction identities for Hill operators

Robert L. Sachs
Department of Mathematics, The Pennsylvania State University, University Park, Pennsylvania 16802

(Received 20 November 1986; accepted for publication 27 April 1988)

Two very general identities are proved involving the eigenfunctions of two Hill operators and the spectrum of a third. Various specializations lead to a number of identities involving squares and fourth powers of the periodic eigenfunctions.

I. INTRODUCTION

Hill operators are ordinary differential operators of the form

$$H := H(q) = -\frac{d^2}{dx^2} + q(x), \tag{1}$$

where the "potential" $q(x)$ is smooth and periodic in x (see Ref. 1). Motivated in large part by the beautiful theory of the Korteweg–de Vries equation, the geometry of the space of all Hill operators has been developed rather fully.[2–4] The picture is as follows: The space of Hill operators (or, equivalently, the space of potentials) decomposes into a disjoint union of "isospectral manifolds" of varying dimensions. Each isospectral manifold is determined by the spectrum of H acting on $L^2(\mathbb{R})$, and all potentials on the same isospectral manifold have the same spectrum. In special cases these manifolds are finite dimensional ("finite gap potentials") while generically they are infinite dimensional ("infinite gap potentials"). The isospectral manifolds are tori parametrized by the Dirichlet equivalues of H.

Since the Korteweg–de Vries flow does not alter the spectrum of H (see Ref. 5), this flow and its hierarchy of related flows provide a large number of vector fields tangential to the isospectral manifold. For technical reasons, McKean and Trubowitz[3] used another set of vector fields to span the tangent space in the infinite-dimensional case. These are the Hamiltonian vector fields generated by the discriminant $\Delta(\lambda;q)$ described below. A surprising identity shows that the KdV vector fields are indeed in the span of this other set:

$$1 = \sum_{j=0}^{\infty} \varepsilon_j f_{2j}^2(x), \tag{2}$$

where the numbers $\varepsilon_j \geqslant 0$ depend only on the spectrum of $H(q)$, and $f_{2j}(x)$ is the normalized periodic or antiperiodic eigenfunction of $H(q)$ with eigenvalue λ_{2j}. The identity (2) has also been used by Moser and Trubowitz[6] to establish a link between Hill operators and the Neumann system of constrained oscillators.

In this paper, two classes of generalizations of (2) are derived. These identities involve two potentials and one spectrum; so they are quite general. Certain specializations lead to some curious identities that are still generalizations of (2). These include a "bilinear" version summed over *all* j,

$$1 = \sum_{j=0}^{\infty} (-1)^j |\dot{\Delta}(\lambda_j;q)| \cdot f_j^2(s;q) f_j^2(t;q), \tag{3}$$

which specializes further to a quartic one,

$$1 = \sum_{j=0}^{\infty} (-1)^j |\dot{\Delta}(\lambda_j;q)| f_j^4(s;q), \tag{4}$$

and a quadratic one

$$1 = \sum_{j=0}^{\infty} (-1)^j |\dot{\Delta}(\lambda_j;q)| f_j^2(s;q), \tag{5}$$

by either $s = t$ or integration over t. There are also the corresponding sums over even j, given below in Sec. IV. These are more complicated identities involving various λ derivatives.

The geometrical significance of these identities awaits further study but they do indicate yet another aspect of the rich structure of Hill operators. The proof of these identities uses contour integration, as in a proof of (2) given by Deift and Trubowitz.[7]

II. NOTATION AND DERIVATION OF TWO GENERAL IDENTITIES

Given $q(x)$ periodic with period 1 and mean value 0,

$$\int_0^1 q(x)dx = 0, \tag{6}$$

which eliminates the shifting $q \to q + \alpha, \lambda_j \to \lambda_j + \alpha$ for $\alpha \in \mathbb{R}$, we define the standard basis of eigenfunctions normalized at 0 [here and elsewhere, a prime (′) denotes d/dx]:

$$\begin{aligned} &H(q)y_j(x,\lambda;q) = \lambda y_j(x,\lambda;q), \quad j = 1,2, \\ &y_1(0,\lambda;q) = y_2'(0,\lambda;q) = 1, \\ &y_1'(0,\lambda;q) = y_2(0,\lambda;q) = 0. \end{aligned} \tag{7}$$

When $q \equiv 0$, clearly $y_1(x,\lambda;q) = \cos\sqrt{\lambda}\cdot x$ and $y_2(x,\lambda;q) = (\sin\sqrt{\lambda}\cdot x)/\sqrt{\lambda}$. For general q, restricted as above, these are asymptotic values as $|\lambda| \to \infty$ (see Ref. 1).

In terms of these functions, the discriminant $\Delta(\lambda;q)$ may be defined as follows:

$$\Delta(\lambda;q) = y_1(1,\lambda;q) + y_2'(1,\lambda;q). \tag{8}$$

This is the trace of the Floquet or transfer matrix and it determines the spectrum of H on $L^2(\mathbb{R})$. Namely, the roots of $\Delta^2 - 4 = 0$ are real and constitute the end points of the real intervals whose union is the spectrum of H acting on $L^2(\mathbb{R})$. Denoting these roots of $\Delta^2 - 4 = 0$ by

$$\lambda_0 < \lambda_1 \leqslant \lambda_2 < \lambda_3 \leqslant \lambda 4 < \cdots, \tag{9}$$

a standard asymptotic estimate[1] for the λ_j's is

$$\lambda_{2n-1}, \lambda_{2n} \simeq n^2\pi^2 + O(1), \quad \text{as } n \to \infty. \tag{10}$$

The open intervals $(\lambda_{2n-1}, \lambda_{2n})$ are the so-called spectral gaps and they may be empty. As n increases, their length

 0022-2488/88/091969-03$02.50

decreases at a rate related to the smoothness of q. For later use, note that (for large n) $(n+\frac{1}{2})^2\pi^2$ is strictly between λ_{2n} and λ_{2n+1}.

The KdV flows preserve the spectrum of H; so they preserve the λ_j's, which in turn implies that they preserve $\Delta(\lambda)$. Thus $\Delta(\lambda)$ is in fact the same for all isospectral potentials. McKean and Trubowitz[3] proved that the equation $\Delta^2(\lambda;q)-4=0$, along with conditions at the double roots $\lambda_{2j-1}=\lambda_{2j}$ (namely, $\dot{\Delta}=d\Delta/d\lambda=0$), defines the isospectral manifold. They also exhibited the vector fields generated by $\Delta(\lambda)$.

Our final notation is for the Bloch eigenfunctions. These are the eigenfunctions that diagonalize the Floquet matrix—i.e., the Bloch functions $f_\pm(x,\lambda;q)$ have the property

$$f_\pm(x+1,\lambda;q)=\mu_\pm(\lambda;q)\cdot f_\pm(x,\lambda;q), \tag{11}$$

where $\mu_\pm(\lambda;q)$ are the Floquet multipliers. The μ's satisfy the quadratic equation

$$\mu^2-\Delta(\lambda;q)\cdot\mu+1=0, \tag{12}$$

and when $\Delta^2-4=0$, $\mu_+=\mu_-=\pm 1$. We will normalize $f_\pm$ by writing

$$f_\pm(x,\lambda;q)=y_1(x,\lambda;q)+m_\pm(\lambda;q)y_2(x,\lambda;q). \tag{13}$$

The term $m_\pm(\lambda;q)$ is easily expressed in terms of $\mu_\pm$ and $y_j(1;\lambda;q)$.

The following two identities will lead to many specializations, including those mentioned above. In their most general form, they involve two potentials and the spectrum of $H(q)$ for a third one. These three functions may lie on three distinct isospectral manifolds.

Identity 1: Let q_1 and q_2 be *any* pair of smooth periodic functions of period 1 with mean value 0. Let $\lambda_0<\lambda_1\leqslant\lambda_2<\cdots$ be the spectrum of $H(Q)$ for some smooth Q also of period 1 and mean value 0. Then

$$\lim_{n\to\infty}\frac{1}{2\pi i}\int_{|\lambda|=(n+1/2)^2\pi^2}\frac{y_2(1,\lambda;q_1)y_2(1,\lambda;q_2)}{1-[\Delta(\lambda;Q)]^2/4}\,d\lambda=1. \tag{14}$$

Identity 2: Let q_1,q_2,Q be as in Identity 1. Then also

$$\lim_{n\to\infty}\frac{1}{2\pi i}\int_{|\lambda|=(n+1/2)^2\pi^2}y_2(1,\lambda;q_1)y_2(1,\lambda,q_2)\times\left((\lambda-\lambda_0)\prod_{j\geq 1}^{\infty}\left[\frac{\lambda_{2j}(Q)-\lambda}{j^2\pi^2}\right]^2\right)^{-1}d\lambda=1. \tag{15}$$

Proof of Identities: Both numerators and denominators are functions with well-known asymptotic behavior as $|\lambda|\to\infty$ (see Ref. 1):

$$\begin{aligned}&y_2(1,\lambda;q_j)=\sin\sqrt{\lambda}/\sqrt{\lambda}\,[1+O(1/|\lambda|)],\\&\Delta(\lambda;Q)=2\cos\sqrt{\lambda}\,[1+O(|\lambda|^{-1/2})],\\&\lambda_{2j}=j^2\pi^2+O(j^{-2}).\end{aligned} \tag{16}$$

So the integrand becomes $f(1/\lambda)d\lambda$ near $\lambda=\infty$, and the sequence of circles avoids the poles.

III. SOME IDENTITIES DERIVED FROM IDENTITY 1

Identity 1 above leads immediately to a residue sum by Cauchy's theorem. In general there are poles of order 2 at worst and order 1 "generically." These poles are at the roots of $\Delta^2(\lambda;Q)-4$, namely, $\lambda_0<\lambda_1\leqslant\lambda_2\cdots$. So we have

$$\sum_{j=0}^{\infty}\operatorname{Res}[\lambda=\lambda_j]\left[\frac{y_2(1,\lambda;q_1)y_2(1,\lambda;q_2)}{1-[\Delta(\lambda;Q)/2]^2}\right]=1. \tag{17}$$

If all λ_j's are distinct (so that all the poles are simple), this yields

$$\sum_{j=0}^{\infty}\frac{y_2(1,\lambda_j;q_1)y_2(1,\lambda_j;q_2)}{[-\Delta(\lambda_j)\dot{\Delta}(\lambda_j)/2]}=1. \tag{18}$$

But

$$\Delta(\lambda_j)=\begin{cases}2, & \text{if } j\equiv 0,3 \bmod 4,\\ -2, & \text{if } j\equiv 1,2 \bmod 4,\end{cases}$$

and

$$\operatorname{sgn}\dot{\Delta}(\lambda_j)=\begin{cases}1, & \text{if } j\equiv 0,1 \bmod 4,\\ +1, & \text{if } j\equiv 2,3 \bmod 4.\end{cases}$$

Thus for purely simple spectrum we have

$$\sum_{j=0}^{\infty}(-1)^j\frac{y_2(1,\lambda_j;q_2)y_2(1,\lambda_j;q_2)}{|\dot{\Delta}(\lambda_j;Q)|}=1. \tag{19}$$

At double poles, the residues are readily computed as well:

$$\frac{d}{d\lambda}\left[\frac{(\lambda-\lambda_j)^2y_2(1,\lambda;q_1)y_2(1,\lambda;q_2)}{(1-\Delta^2/4)}\right]_{\lambda=\lambda_j}. \tag{20}$$

In certain cases below, these terms will vanish.

Proposition 3: Let q_1 and q_2 be isospectral with Q. Then $y_2(1,\lambda_j;q_1)=y_2(1,\lambda;q_2)=0$ at double roots $\lambda_{2j-1}=\lambda_j$. Hence the only residues are at simple roots.

Proof: When there is no gap, there must be a fixed Dirichlet eigenvalue—i.e., $y_2(1,\lambda_{2j};q_k)$ vanishes for $k=1,2$. In this case, the double root in the denominator of (19) is actually a removable singularity.

A similar argument yields the next proposition.

Proposition 4: If q_1 and Q are isospectral, then all poles are in fact simple.

There are many ways to specialize Identity 1 above. We will limit our discussion to those that we believe will be of most interest and that lead to manageable formulas.

Corollary 5: If $\lambda_0<\lambda_1<\lambda_2<\cdots$ is the spectrum for a Hill operator, $H(Q)$, with all gaps open, then

$$\sum_{j=0}^{\infty}(-1)^j\frac{\sin^2(\sqrt{\lambda_j})}{\lambda_j|\dot{\Delta}(\lambda_j;Q)|}=1. \tag{21}$$

Proof: Set $q_1=q_2=0$.

Corollary 6: If $\lambda_0<\lambda_1<\cdots$ is as above, then for any q_1 with mean value zero,

$$\sum_{j=0}^{\infty}\frac{(-1)^jy_2(1,\lambda_j;q_1)\sin(\sqrt{\lambda_j})}{\sqrt{\lambda_j}\,|\dot{\Delta}(\lambda_j;Q)|}=1. \tag{22}$$

Proof: Set $q_2=0$.

Corollary 7: For any q_1 periodic with mean value zero,

$$y_2(1,0;q_1)+2\sum_{j=0}^{\infty}(-1)^jy_2(1,j^2\pi^2;q_1)=1. \tag{23}$$

Proof: Set $Q=q_2=0$.

Corollary 8: For any q_1,q_2 with mean value 0,

$$\sum_{j=0}^{\infty}\operatorname{Res}_{\lambda=j^2\pi^2}\frac{y_2(1,\lambda;q_1)y_2(1,\lambda;q_2)}{\sin^2(\sqrt{\lambda})}=1. \tag{24}$$

Proof: Set $Q=0$.

Corollary 9: For any Q periodic with mean value 0,

$$\sum_{j=0}^{\infty}(-1)^j|\dot{\Delta}(\lambda_j;Q)|\cdot f_j^2(s;Q)\cdot f_j^2(t;Q)=1. \tag{25}$$

Proof: Let $q_1=Q(\cdot+s)$, $q_2=Q(\cdot+t)$. Then using the facts

$$y_2(1,\lambda;Q(\cdot+s))=y_2(1,\lambda;Q)\,f_+(s,\lambda;Q)\,f_-(s,\lambda;Q),$$
$$y_2(1,\lambda_j;Q)\,f_+(s,\lambda_j;Q)\,f_-(s,\lambda_j;Q)=-\dot{\Delta}(\lambda_j)\,f_j^2(s;Q), \tag{26}$$

where f_j is the normalized eigenfunction at λ_j, (25) follows. The double roots do not contribute; so any ambiguity in f_j is irrelevant.

Corollary 10: For any Q periodic with mean value 0,

$$\sum_{j=0}^{\infty}(-1)^j|\dot{\Delta}(\lambda_j)|\cdot f_j^4(s;Q)=1. \tag{27}$$

Proof: Set $q_1=q_2=Q(\cdot+s)$.

Corollary 11: For any Q periodic with mean value 0,

$$\sum_{j=0}^{\infty}(-1)^j|\dot{\Delta}(\lambda_j)|f_j^2(s;Q)=1. \tag{28}$$

Proof: Integrate (25) in t from 0 to 1.

Corollary 12: For any Q periodic with mean value 0,

$$\sum_{j=0}^{\infty}(-1)^j|\dot{\Delta}(\lambda_j;Q)|=1. \tag{29}$$

Proof: Integrate (28) in s from 0 to 1.

IV. SOME IDENTITIES DERIVED FROM IDENTITY 2

Identity 2 also leads to a residue sum by Cauchy's theorem. In this case, all poles except λ_0 are double; so the residue formula is more complicated. Before proceeding, we introduce some further notation:

$$\begin{aligned}
\varepsilon_0^{-1}&:=\prod_{k=1}^{\infty}\left(\frac{\lambda_{2k}-\lambda_0}{k^2\pi^2}\right)^2,\\
\varepsilon_j^{-1}&:=\frac{(\lambda_{2j}-\lambda_0)}{j^2\pi^2}\cdot\prod_{\substack{k=1\\k\neq j}}^{\infty}\left(\frac{\lambda_{2k}-\lambda_{2j}}{k^2\pi^2}\right)^2,\quad j\geqslant 1,\\
\delta_0&:=\varepsilon_0^{-1},\\
\delta_j&:=-\frac{d}{d\lambda}\left[\left(\frac{\lambda-\lambda_0}{j^2\pi^2}\right)\cdot\prod_{\substack{k=1\\k\neq j}}^{\infty}\left(\frac{\lambda_{2k}-\lambda}{k^2\pi^2}\right)^2\right]_{\lambda=\lambda_{2j}},\quad j\geqslant 1.
\end{aligned} \tag{30}$$

Note that $\varepsilon_j>0$. Using these definitions, Identity 2 becomes

$$\begin{aligned}
1&=\sum_{j=0}^{\infty}\mathrm{Res}_{\lambda=\lambda_{2j}}\left\{y_2(1,\lambda;q_1)\cdot y_2(1,\lambda;q_2)\right.\\
&\quad\left.\times\left((\lambda-\lambda_0)\prod_{k=1}\left(\frac{\lambda_{2k}-\lambda}{k^2\pi^2}\right)^2\right)^{-1}\right\}\\
&=\sum_{j=1}^{\infty}\varepsilon_j\cdot\frac{d}{d\lambda}(y_2 1,\lambda;q_1)y_2(1,\lambda;q_2)|_{\lambda=\lambda_{2j}}\\
&\quad+\sum_{j=0}^{\infty}\varepsilon_j^2\cdot\delta_j\cdot y_2(1,\lambda_{2j};q_1)y_2(1,\lambda_{2j};q_2).
\end{aligned} \tag{31}$$

Note that $Q=0$ reduces Identity 2 to Identity 1, since in this case (and this case only) there are no gaps. The other reductions, such as $q_1=0$ or $q_1=q_2=0$, are still allowed. These yield the following corollaries.

Corollary 13: For Q,q_2 periodic with mean value 0,

$$\begin{aligned}
1&=\sum_{j=1}^{\infty}\varepsilon_j\left[(\cos\sqrt{\lambda_{2j}}\ \sin\sqrt{\lambda_{2j}}\,2\kappa_{2j}\sqrt{\lambda_{2j}})\cdot y_2(1,\lambda_{2j};q_2)\right.\\
&\quad\left.+\frac{\sin\sqrt{\lambda_{2j}}}{\sqrt{\lambda_{2j}}}\dot{y}_2(1,\lambda_{2j};q_2)\right]\\
&\quad+\sum_{j=0}^{\infty}\varepsilon_j^2\,\delta_j\cdot\frac{\sin\sqrt{\lambda_{2j}}}{\sqrt{\lambda_{2j}}}\cdot y_2(1,\lambda_{2j};q_2).
\end{aligned} \tag{32}$$

Proof: Set $q_1\equiv 0$.

Corollary 14: For Q periodic with mean value 0,

$$1=\sum_{j=1}^{\infty}\delta_j\frac{d}{d\lambda}\left[\frac{\sin^2\sqrt{\lambda}}{\sqrt{\lambda}}\right]_{\lambda=\lambda_{2j}}+\sum_{j=0}^{\infty}\varepsilon_j^2\,\delta_j\cdot\frac{\sin^2\sqrt{\lambda_{2j}}}{\lambda_{2j}}. \tag{33}$$

Proof: Set $q_1=q_2=0$.

Corollary 15: For Q periodic with mean value 0,

$$\begin{aligned}
1&=\sum_{j=1}^{\infty}\varepsilon_j\frac{d}{d\lambda}\left[y_2^2(1,\lambda;Q)\,f_+(s,\lambda;Q)\,f_-(s,\lambda;Q)\right.\\
&\quad\left.\cdot f_+(t,\lambda;Q)\,f_-(t,\lambda;Q)\right]_{\lambda=\lambda_{2j}}\\
&\quad+\sum_{j=0}^{\infty}\varepsilon_j^2\,\delta_j[\dot{\Delta}(\lambda_{2j};Q)]^2\cdot f_{2j}^2(s;Q)\cdot f_{2j}^2(t;Q).
\end{aligned} \tag{34}$$

Proof: Set $q_1=Q(\cdot+s)$, $q_2=Q(\cdot+t)$.

Corollary 16: For Q periodic with mean value 0,

$$\begin{aligned}
1&=\sum_{j=1}^{\infty}\varepsilon_j\frac{d}{d\lambda}\left[y_2^2(1,\lambda;Q)\cdot f_+^2(s,\lambda;Q)\,f_-^2(s,\lambda;Q)\right]_{\lambda=\lambda_{2j}}\\
&\quad+\sum_{j=0}^{\infty}\varepsilon_j^2\,\delta_j[\dot{\Delta}(\lambda_{2j};Q)]^2\cdot f_{2j}^4(s;Q).
\end{aligned} \tag{35}$$

Proof: Set $q_1=q_2=Q(\cdot+s)$.

Remark: The double roots $\lambda_{2j-1}=\lambda_{2j}$ will not contribute to the sums in Corollary 15 and Corollary 16 since $y_2(1,\lambda;Q)$ vanishes as does $\dot{\Delta}(\lambda;Q)$.

[1]W. Magnus and S. Winkler, *Hill's Equation* (Wiley, New York, 1966; Dover, New York, 1979).
[2]H. P. McKean and P. van Moerbeke, Invent. Math. **30**, 217 (1975).
[3]H. P. McKean and E. Trubowitz, Commun. Pure Appl. Math. **29**, 143 (1976).
[4]S. P. Novikov, Funct. Anal. Appl. **8**, 236 (1974).
[5]P. D. Lax, Commun. Pure Appl. Math. **28**, 141 (1975).
[6]J. Moser, *Integrable Hamiltonian System and Spectral Theory* (Lezione Fermiane, Pisa, 1981).
[7]P. Deift and E. Trubowitz, Commun. Pure Appl. Math. **34**, 713 (1981).

Proper affine collineations in Robertson–Walker space-times

C. D. Collinson
School of Mathematics, The University of Hull, Hull HU6 7RX, England

(Received 26 January 1988; accepted for publication 27 April 1988)

It is shown that the only Robertson–Walker space-times admitting a proper affine collineation are static, hence proving the uniqueness of the collineation found recently by Maartens [J. Math. Phys. **28**, 2051 (1987)].

I. INTRODUCTION

An affine collineation is generated by a vector field ξ satisfying

$$\underset{\xi}{\pounds}\Gamma^{\alpha}_{\beta\gamma} = \xi_{(\alpha;\beta);\gamma} = 0. \tag{1.1}$$

Such a collineation is called proper if it is *not* a homothetic collineation, for which the vector field ξ would satisfy

$$\underset{\xi}{\pounds} g_{\alpha\beta} = \xi_{(\alpha;\beta)} = \psi g_{\alpha\beta}, \quad \psi_{,\gamma} = 0.$$

Bedran, Lesche[1] and Maartens[2] have investigated solutions of (1.1) for the Robertson–Walker space-times with metric

$$ds^2 = dt^2 - R^2(t)(dr^2/(1-kr^2) + r^2\,d\theta^2 + r^2\sin^2\theta\,d\varphi^2). \tag{1.2}$$

Bedran and Lesche confine attention to solutions ξ that are isotropic and homogeneous and Maartens points out that the calculations are incomplete in the static case $\dot{R} = 0$, for which a proper affine collineation is found.

In this paper the first integrability condition of (1.1) will be exploited to prove the following theorems.

Theorem 1: Nonempty Robertson–Walker space-times admitting a proper affine collineation are necessarily static.

Theorem 2: Nonempty, static Robertson–Walker space-times admit one and only one independent *proper* affine collineation (in the sense that any other proper collineation is obtained by adding a homothetic Killing vector to the generator).

II. PROOF OF THE THEOREMS

The first integrability condition of (1.1) is[3]

$$\underset{\xi}{\pounds} R^{\alpha}{}_{\beta\gamma\delta} = 0.$$

This equation defines a curvature collineation and Katzin, Levine, and Davis[4] have shown that a necessary condition for a curvature collineation to exist is that

$$h_{\alpha\epsilon}R^{\epsilon}{}_{\beta\gamma\delta} + h_{\beta\epsilon}R^{\epsilon}{}_{\alpha\gamma\delta} = 0, \tag{2.1}$$

where

$$h_{\alpha\beta} = \underset{\xi}{\pounds} g_{\alpha\beta} = \xi_{(\alpha;\beta)}.$$

This equation has been investigated in some detail, first by Collinson[5] for empty space-times and then by Hall[6] and others for nonempty space-times. In particular Hall shows that four different forms of the general solution $h_{\alpha\beta}$ of the equation arise (see his Theorem 1). The first form arises when two independent vectors k exist satisfying the equation

$$R^{\epsilon}{}_{\beta\gamma\delta}k_{\epsilon} = 0. \tag{2.2}$$

The second form, namely

$$h_{\alpha\beta} = \varphi g_{\alpha\beta} + \lambda k_{\alpha}k_{\beta} \tag{2.3}$$

arises when only one vector k exists satisfying (2.2). The third form requires the existence of null eigenvectors of the Ricci tensor and the fourth and final form is

$$h_{\alpha\beta} = \varphi g_{\alpha\beta} \tag{2.4}$$

The Robertson–Walker space-times are conformally flat so that (2.2) is equivalent to

$$R^{\epsilon}{}_{\beta}k_{\epsilon} = 0, \tag{2.5}$$

and

$$k_{\alpha}(R_{\beta\gamma} - \tfrac{1}{3}Rg_{\beta\gamma}) = k_{\beta}(R_{\alpha\gamma} - \tfrac{1}{3}Rg_{\alpha\gamma}). \tag{2.6}$$

Also

$$R_{\alpha\beta} = -(p+\rho)u_{\alpha}u_{\beta} + \tfrac{1}{2}(\rho - p)g_{\alpha\beta}, \tag{2.7}$$

where

$$p + \rho = 2\left[-\left(\frac{\ddot{R}}{R}\right) + \left(\frac{\dot{R}^2}{R^2}\right) + \frac{k}{R^2}\right], \tag{2.8}$$

and

$$\rho - p = 2\left[\left(\frac{\ddot{R}}{R}\right) + 2\left(\frac{\dot{R}}{R}\right)^2 + \frac{2k}{R^2}\right]. \tag{2.9}$$

Substituting (2.7) into (2.5) and (2.6) yields

$$-(p+\rho)u_{\beta}u^{\epsilon}k_{\epsilon} + \tfrac{1}{2}(\rho - p)k_{\beta} = 0, \tag{2.10}$$

and

$$k_{\alpha}\left[-(p+\rho)u_{\beta}u_{\gamma} + \tfrac{1}{6}(\rho + 3p)g_{\beta\gamma}\right] = k_{\beta}\left[-(p+\rho)u_{\alpha}u_{\gamma} + \tfrac{1}{6}(\rho + 3p)g_{\alpha\gamma}\right]. \tag{2.11}$$

Contracting (2.10) with u^{β} gives

$$-\tfrac{1}{2}u^{\epsilon}k_{\epsilon}(3p + \rho) = 0.$$

There are two possibilities. If $u^{\epsilon}k_{\epsilon} = 0$ then (2.10) and the contraction of (2.11) with u^{α} gives

$$\tfrac{1}{2}(\rho - p)k_{\beta} = 0 \quad \text{and} \quad -\tfrac{1}{6}(5\rho + 3p)k_{\beta}u_{\gamma} = 0.$$

For a nonempty space-time (i.e., not both ρ and p are zero) these equations imply the $k_{\beta} = 0$. If $u^{\epsilon}k_{\epsilon} \neq 0$ then $3p + \rho = 0$ and (2.10) yields, for nonempty space-times, $k_{\beta} = u^{\epsilon}k_{\epsilon}u_{\beta}$ or, with a rescaling, $k_{\beta} = u_{\beta}$. Equation (2.11) is then identically satisfied. From this it follows that the first form of $h_{\alpha\beta}$ referred to above cannot exist and that the second form, (2.3), only exists if $3p + \rho = 0$ and $k_{\alpha} = u_{\alpha}$.

If k_{α} is a null eigenvector of the Ricci tensor then

$$R^{\epsilon}{}_{\beta}k_{\epsilon}k^{\beta} = 0.$$

This is the contraction of (2.10) with k^{β}, namely

 0022-2488/88/091972-02$02.50

$$-(p+\rho)(u^{\epsilon}k_{\epsilon})^2 = 0.$$

With k null, $u^{\epsilon}k_{\epsilon}$ is nonzero and so $p+\rho=0$. Then

$$R_{\alpha\beta} = -pg_{\alpha\beta} \quad \text{and} \quad R_{\alpha\beta\gamma\delta} = \tfrac{1}{3}p(g_{\alpha\gamma}g_{\beta\delta} - g_{\alpha\delta}g_{\beta\gamma}).$$

Equation (2.1) is then easily shown to lead, with $p \neq 0$, to the form (2.4).

Now $h_{\alpha\beta} = \xi_{(\alpha;\beta)}$ and substituting (2.4) into (1.1) gives $\varphi_{,\gamma} = 0$. It follows that the form (2.4) corresponds to a homothetic collineation. The only case in which a proper affine collineation may exist is when $3p+\rho=0$ and then

$$\xi_{(\alpha;\beta)} = \varphi g_{\alpha\beta} + \lambda u_{\alpha}u_{\beta}. \tag{2.12}$$

Using (2.8) and (2.9) the condition $3p+\rho=0$ is equivalent to $\ddot{R}=0$. Substituting (2.12) into (1.1) yields

$$\varphi_{,\gamma}g_{\alpha\beta} + \lambda_{,\gamma}u_{\alpha}u_{\beta} + \lambda u_{\alpha;\gamma}u_{\beta} + \lambda u_{\alpha}u_{\beta;\gamma} = 0. \tag{2.13}$$

Contracting (2.13) on $\alpha\beta$ gives

$$4\varphi_{,\gamma} + \lambda_{,\gamma} = 0,$$

while contracting with $u^{\alpha}u^{\beta}$ gives

$$\varphi_{,\gamma} + \lambda_{,\gamma} = 0.$$

It follows that both φ and λ must be constant and that (2.13) reduces to

$$\lambda(u_{\alpha;\gamma}u_{\beta} + u_{\beta;\gamma}u_{\alpha}) = 0.$$

Contracting this with u^{β} yields

$$\lambda u_{\alpha;\gamma} = 0. \tag{2.14}$$

Now $u^{\alpha} = \delta^{\alpha}_{0}$ so that

$$u^{\alpha}{}_{;\gamma} = \Gamma^{\alpha}_{0\gamma} = \tfrac{1}{2}g^{\alpha\beta}g_{\beta\gamma,0} = (\dot{R}/R)(\delta^{\alpha}_{\gamma} - u^{\alpha}u_{\gamma}).$$

It follows that if $\dot{R} \neq 0$, (2.14) requires λ to be zero, in which case (2.12) reduces to the form (2.4), thus proving Theorem 1. Theorem 2 follows from the constancy of λ, where the proper affine collineation referred to is that already found by Maartens.

[1] M. L. Bedran and B. Lesche, J. Math. Phys. **27**, 2360 (1986).
[2] R. Maartens, J. Math. Phys. **28**, 2051 (1987).
[3] K. Yano, *The Theory of Lie Derivatives and its Application* (North-Holland, Amsterdam, 1955).
[4] G. H. Katzin, J. Levine, and W. R. Davis, J. Math. Phys. **10**, 617 (1969).
[5] C. D. Collinson, J. Math. Phys. **11**, 818 (1970).
[6] G. S. Hall, Gen. Relativ. Gravit **15**, 581 (1983).

Matrix hierarchies and vector bundles over Riemann surfaces

G. F. Helminck
Universiteit Twente, Faculteit der Toegepaste Wiskunde, P.O. Box 217, 7500 AE Enschede, The Netherlands

(Received 22 January 1988; accepted for publication 13 April 1988)

To a compact Riemann surface, a point on the surface, a local coordinate around this point, and a holomorphic transfer function of degree m is associated a point in the Grassmann manifold of $l^2(\mathbb{Z})^m$. This leads to solutions of the nonlinear hierarchies corresponding to maximal Abelian subalgebras of $\mathrm{End}(\mathbb{C}^m)$.

I. INTRODUCTION

In this paper we show how to associate to a compact Riemann surface X a point x_∞ on X, a local coordinate z^{-1} around x_∞, and a holomorphic transfer function ϕ of degree m on the unit circle, a subspace W_ϕ in the Grassmann manifold $\mathrm{Gr}(l^2(\mathbb{Z})^m)$. As is shown in Ref. 1, one can construct solutions of several hierarchies of nonlinear partial differential equations, starting from certain open subsets of $\mathrm{Gr}(l^2(\mathbb{Z})^m)$. This accounts for the occurrence of holomorphic vector bundles over compact Riemann surfaces in this context. In Sec. II we give a description of the hierarchies. In Sec. III the subspace W_ϕ is constructed, and in Sec. IV we express the virtual dimension of W_ϕ in the dimensions of cohomology groups.

II. THE HIERARCHIES

In Ref. 1 we associated to each maximal commutative subalgebra D of $Gl_m(\mathbb{C})$ and an invertible element A of D, a hierarchy of nonlinear partial differential equations whose solution space we denoted by $\mathbf{H}(D,A)$. The equations are formulated in terms of pseudodifferential operators with coefficients from an algebra B, consisting of $m\times m$ matrices depending on the parameters $t_{i\alpha}$, $i\geqslant 0$ and $1\leqslant\alpha\leqslant r = \dim(D)$.

Let E_α, $1\leqslant\alpha\leqslant r$, be a basis of D. Since D is an algebra, we have

$$E_\alpha E_\beta = \sum_{\gamma=1}^{r} i_{\alpha\beta\gamma}E_\gamma.$$

Then we are looking for operators in $B[\xi,\xi^{-1}]$ of the form

$$L = A\xi + \sum_{j<0} l_j\xi^j$$

and

$$U_\alpha = E_\alpha + \sum_{j<0} u_{j,\alpha}\xi^j$$

that satisfy the following set of relations:

$$[L,U_\alpha] = 0, \quad [U_\alpha,U_\beta] = 0, \quad \text{for all } \alpha \text{ and } \beta, \tag{2.1}$$

$$U_\alpha U_\beta = \sum_{\gamma=1}^{r} i_{\alpha\beta\gamma}U_\gamma, \tag{2.2}$$

$$\partial_{i\alpha}(L) = [(L^iU_\alpha)_+,L], \tag{2.3}$$

$$\partial_{i\alpha}(U_\beta) = [(L^iU_\alpha)_+,U_\beta] \tag{2.4}$$

Here $\partial_{i\alpha}$ denotes the partial derivative w.r.t. $t_{i\alpha}$, and P_+, for $P\in B[\xi,\xi^{-1}]$, denotes the differential operator part of P.

It was further shown in Ref. 1 that one could produce a considerable collection of solutions of this hierarchy starting from an open subset $Q(D,A)$ of the Grassmann manifold $\mathrm{Gr}(\mathbf{H})$ with $\mathbf{H} = l^2(\mathbb{Z})^m$. The elements of $\mathbf{H}$ we see as sequences (a_k), $k\in\mathbb{Z}$ and $a_k\in\mathbb{C}^m$, and the inner product on $\mathbf{H}$ has the form

$$(a_k,b_k) = \sum_{k\in\mathbb{Z}} a_k\bar{b}_k^t.$$

The space $\mathrm{Gr}(\mathbf{H})$ is defined with respect to the splitting $\mathbf{H} = \mathbf{H}_{>0}\oplus\mathbf{H}_{<0}$, with $\mathbf{H}_{>0} = \{(a_k)\in\mathbf{H},\ a_k = 0 \text{ for all } k<0\}$ and $\mathbf{H}_{<0} = \mathbf{H}_{\geqslant 0}^{\perp}$. It consists of all closed subspaces of $\mathbf{H}$ that are the image of an embedding w: $\mathbf{H}_{>0}\to\mathbf{H}$ that decomposes w.r.t. $\mathbf{H} = \mathbf{H}_{>0}\oplus\mathbf{H}_{<0}$ as $w = \binom{w_+}{w_-}$, with w_+ Fredholm and w_- bounded. The index of w_+ is called the *virtual dimension* of the image of w.

If Q is the open set of $\mathrm{Gr}(l^2(\mathbb{Z})^m)$ consisting of all subspaces that project bijectively onto $\mathbf{H}_{>0}$, then the open set $Q(D,A)$ of $\mathrm{Gr}(l^2(\mathbb{Z})^m)$, from which we can construct solutions of Eqs. (2.1)–(2.4), consists of the subspaces obtained by letting the group of flows corresponding to D and A act on Q. It consists of operators of the form

$$\exp\left(\sum_{\substack{i\geqslant 0\\ 1<\alpha<r}} t_{i\alpha}E_\alpha A^i\,\mathrm{diag}(\Lambda^i,\ldots,\Lambda^i)\right),$$

where Λ denotes the shift operator in $l_2(\mathbb{Z})$; i.e., if $\alpha = (b_n)\in l^2(\mathbb{Z})$, then $(\Lambda\alpha)_n = b_{n-1}$. We denote by $(L_W,U_{\alpha,W})$ the solution of the hierarchy corresponding to $W\in Q(D,A)$.

III. THE CONSTRUCTION OF THE SUBSPACE

Let X be a compact Riemann surface of genus g. We choose a point x_∞ on X, an open neighborhood U of x_∞, and an isomorphism z of U with an open neighborhood

$$\{x|x\in\mathbb{P}^1(\mathbb{C}),\ |x|>1/(1+\epsilon_z),\ \epsilon_z>0\}$$

of the closed unit disk around infinity such that $z(x_\infty) = \infty$. In other words $1/z$ is a local coordinate around x_∞. We will identify subsets of U and functions on these subsets with their image under z and with functions on this image. If μ is a map with the same properties as z, there exists a power series $\Sigma_{i\geqslant 0}\, c_i\lambda^i$ satisfying first of all

$$\sum_{i\geqslant 0} |c_i|\,|1+\epsilon|^i < \infty,$$

for some $\epsilon>0$; second, for all $x\in\mathbb{P}^1(\mathbb{C})$,

 0022-2488/88/091974-05$02.50

$$|x|\geqslant 1/(1+\epsilon), \quad \sum_{i\geqslant 0} c_i x^{-i}\neq 0;$$

and finally

$$\mu^{-1} = z^{-1}\left(\sum_{i\geqslant 0} c_i z^{-i}\right). \tag{3.1}$$

A. Notation

Let X_∞ be the inverse image under z of $\{x|x\in\mathbb{P}^1(\mathbb{C}), |x|\geqslant 1\}$, let X_0 be the closure of $X\setminus X_\infty$, and write $X_{0\infty}$ for $X_0\cap X_\infty$. We fix an open neighborhood V of X_0, namely, the complement of the inverse image under z of $\{x|x\in\mathbb{P}^1(\mathbb{C}), |x|\geqslant 1+\epsilon_z\}$.

In what follows any capital letter with subscript 0, like T_0, will mean an open connected subset of V containing X_0; any capital letter with subscript ∞ will denote an open connected neighborhood of X_∞ in U; and any capital letter with subscript 0∞ denotes an open connected subset of $U\cap V$ containing $X_{0\infty}$. In particular, for a given $U_{0\infty}$, we write U_0 for $U_{0\infty}\cup X_0$ and U_∞ for $U_{0\infty}\cup X_\infty$.

Let f: $U_{0\infty}\to\mathbb{C}^m$ be analytic. On a neighborhood of S^1, f equals its Laurent series w.r.t. z^{-1}, i.e.,

$$f=\sum_{k\in\mathbb{Z}} a_k z^k, \quad \text{with } a_k\in\mathbb{C}^m.$$

This series converges absolutely on a region $1-\epsilon<|z(x)|<1+\epsilon$; therefore the (a_k) belongs to $l^2(\mathbb{Z})^m$. In this way all analytic functions from $U_{0\infty}$ to C^m are embedded into $l^2(\mathbb{Z})^m$. Let dx be the Haar measure on S^1 with $\int_{S^1} dx = 1$. Then the l^2 norm of f, $\|f\|_2$, is given by

$$\|f\|_2^2 = \int_{S^1} f(x)(\overline{f(x)})^t\, dx = \sum_{k\in\mathbb{Z}} a_k \bar{a}_k^t,$$

where we have written elements of $\mathbb{C}^m$ as rows, and t means transposition.

B. Transfer functions

Definition (3.1): A *transfer function* ϕ of degree m is a holomorphic map from some $U_{0\infty}$ to $Gl_m(\mathbb{C})$.

To such a function we can associate a holomorphic vector bundle of degree m over X. One simply glues the trivial bundles on U_0 and U_∞ according to

$$(x,v)\to(x,\phi(x)v), \quad \text{with } x\in U_{0\infty},\ v\in\mathbb{C}^m.$$

We denote this bundle by L_ϕ and the projection on the base space by π. The sheaf of holomorphic sections of (L_ϕ,π) we denote by $\mathbb{L}_\phi$. If $\mathbf{O}$ is the sheaf of holomorphic functions on X, then $\mathbb{L}_\phi$ is a locally free $\mathbf{O}$ module of rank m.

Remark 1: In the same way as above one associates to a transfer function also a holomorphic bundle over $\mathbb{P}^1(\mathbb{C})$. As Grothendieck[2] has shown, its structure is rather simple. Over $\mathbb{P}^1(\mathbb{C})$ a transfer function ϕ has the following decomposition:

$$\phi(x) = \phi_+(x)\,\mathrm{diag}(x^{l_1},\ldots,x^{l_m})\phi_-(x),$$

where ϕ_+ is an analytic map $U_\infty\to Gl_m(\mathbb{C})$, ϕ_- is an analytic map $U_0\to Gl_m(\mathbb{C})$, and the l_i belong to $\mathbb{Z}$ and satisfy $l_{i+1}\geqslant l_i$. This implies that the bundle over $\mathbb{P}^1(\mathbb{C})$ associated to ϕ is a direct sum of line bundles.

Remark 2: Since U and V are noncompact Riemann surfaces, every holomorphic bundle over U or V is trivial (see Ref. 3, p. 204). Hence if L is any holomorphic vector bundle over X of degree m, then the restrictions of L to U and V are trivial, and there is a transfer function ϕ on $U\cap V$ such that L and L_ϕ are isomorphic as holomorphic bundles over X.

Remark 3: If W_1 and W_2 are open nonempty subsets of X, with $W_1\subset W_2$, then restriction to W_1 defines an injection of $\mathbb{L}_\phi(W_2)$ into $\mathbb{L}_\phi(W_1)$. For each closed nonempty A, the sections of $\mathbb{L}_\phi$ over A are the elements of

$$\varinjlim_{\substack{W\supset A\\ W\text{ open}}} \mathbb{L}_\phi(W).$$

Using the identification just described this space is nothing but the union of the $\mathbb{L}_\phi(W)$ with $W\supset A$.

Remark 4: If W is any subspace of $l^2(\mathbb{L})^m$, then we write W^{an} for the subspace of W consisting of the analytic functions in W, i.e.,

$$\left\{(a_k)\in W \,\middle|\, \sum_{k\in\mathbb{Z}} a_k z^k\in\mathbb{L}_{I_m}(S^1)\right\},$$

where I_m is the transfer function $x\to \mathrm{Id}_{\mathbb{C}^m}$. Every transfer function ϕ defines through the composition of functions an automorphism of $\{l^2(\mathbb{Z})^m\}^{\mathrm{an}}$, which is continuous in the L^2 norm. Hence it extends to a continuous automorphism ϕ of $l^2(\mathbb{Z})^m$.

C. The space W_ϕ

In this section we introduce the space W_ϕ. It is the natural extension to $m>1$ of the spaces considered in Ref. 4. For an arbitrary ϕ we embed $\mathbb{L}_\phi(X_0)$ into $l^2(\mathbb{Z})^m$ as follows: First restrict these sections to a neighborhood of S^1 in $U\cap V$, express them in the standard trivialization of L_ϕ around x_∞, and identify the sections of that trivialization in the obvious way with holomorphic $\mathbb{C}^m$-valued functions on a neighborhood of S^1. Clearly $\mathbb{L}_{I_m}(X_0)$ and $\mathbb{L}_\phi(X_0)$ are in $l^2(\mathbb{Z})^m$ related through $\mathbb{L}_\phi(X_0)=\phi\,(\mathbb{L}_{I_m}(X_0))$, and further we have $\mathbb{L}_{I_m}(X_0)=\mathbb{L}_{I_1}(X_0)^m$. For W_ϕ we take the closure in $l_2(\mathbb{Z})^m$ of $\mathbb{L}_\phi(X_0)$, and from the foregoing we see that $W_\phi=\phi(W_{I_m})=\phi((W_{I_1})^m)$. This relation also implies that if ϕ_0: $U_0\to Gl_m(\mathbb{C})$ is holomorphic, then $W_{\phi\phi_0}=W_\phi$ for all ϕ.

Remark 1: If $m=1$, then (as shown in Ref. 4) multiplying ϕ with a holomorphic function ϕ_∞: $V_\infty\to\mathbb{C}^*$, with $\phi_\infty(x_\infty)=1$, does not affect the solution of the Kadomtsev–Petviashvili (KP) hierarchy, i.e., $L_{W_{\phi_\infty\phi}}=L_{W_\phi}$. Thus one could associate to a line bundle a solution of the KP hierarchy. However, for $m>1$ this does not necessarily hold, for to show it one needs that the ϕ_∞ action commutes with the operators that give the flows on the Grassmannian; in other words, it holds for holomorphic ϕ_∞: $V_\infty\to\exp(D)$, with $\phi_\infty(x_\infty)=\mathrm{Id}$.

Remark 2: From (3.1) one deduces directly that a transition on another local coordinate around x_∞ corresponds to applying an automorphism of $l^2(\mathbb{Z})^m$ to all the W_ϕ, of the form

$$\begin{pmatrix} B & & & 0 \\ & B & & \\ & & \ddots & \\ 0 & & & B \end{pmatrix}, \quad B \in \mathrm{Gl}(l^2(\mathbb{Z})),$$

$$B = \begin{pmatrix} \ddots & & & & & & \\ & 1 & & & & & \\ & & 1 & & 0 & & \\ & & & 1 & & & \\ & & * & & 1 & & \\ & & & & & 1 & \\ & & & & & & \ddots \end{pmatrix}.$$

In this paragraph we show that W_ϕ belongs to $\mathrm{Gr}(l^2(\mathbb{Z})^m)$, and we compute its virtual dimension. With remark 1 of Sec. III B one can reduce this to the case $\phi = I_m$. Using the notation of the first section, let Φ^+ be the automorphism of $\mathbf{H}$ corresponding to ϕ_+ and Φ_- that corresponding to ϕ_-.

With respect to $\mathbf{H} = \mathbf{H}_{>0} \oplus \mathbf{H}_{<0}$ we have the decompositions

$$\Phi_+ = \begin{pmatrix} a_1 & 0 \\ c_1 & d_1 \end{pmatrix}, \quad \mathrm{diag}(\Lambda^{l_1},\ldots,\Lambda^{l_m}) = \begin{pmatrix} \alpha & \beta \\ \gamma & \delta \end{pmatrix},$$

$$\Phi_- = \begin{pmatrix} a_2 & b_2 \\ 0 & d_2 \end{pmatrix},$$

with β and γ finite dimensional and α Fredholm of index $-\Sigma_{i=1}^m l_i$. Assume that W_{I_m} belongs to $\mathrm{Gr}_s(\mathbf{H})$. Then W_{I_m} is the image of a continuous embedding w of $\mathbf{H}_{>0}$ in $\mathbf{H}$ that decomposes w.r.t. $\mathbf{H} = \mathbf{H}_{>0} \oplus \mathbf{H}_{<0}$ as $\binom{w_+}{w_-}$, with w_+ Fredholm of index s. Now

$$W_\phi = \Phi_+ \,\mathrm{diag}(\Lambda^{l_1},\ldots,\Lambda^{l_m})\Phi_-(W_{I_m}),$$

and by substituting the decompositions above one gets that W_ϕ is the image of the embedding given by

$$\begin{pmatrix} a_1\alpha a_2 & a_1\alpha b_2 + a_1\beta d_2 \\ c_1\alpha a_2 + d_1\gamma\alpha_2 & (c_1\alpha + d_1\gamma)b_2 + (c_1\beta + d_1\delta)d_2 \end{pmatrix}\begin{pmatrix} w_+ \\ w_- \end{pmatrix}.$$

Hence

$$W_\phi \in \mathrm{Gr}_{s+\mathrm{index}(\alpha)}(\mathbf{H}) = \mathrm{Gr}_{s-\Sigma_{i=1}^m l_i}(\mathbf{H}).$$

In the next section we will express the virtual dimension of W_ϕ in the dimensions of some cohomology groups as was done in Ref. 4 for the case $m = 1$.

IV. THE VIRTUAL DIMENSION OF W_ϕ

Let U_0, U_∞, and $U_{0\infty}$ be as in Sec. III A. Since noncompact Riemann surfaces are Stein (see Ref. 5), U_0, U_∞, and $U_{0\infty}$ are a-cyclic for any coherent $\mathbf{O}$ module $\mathbf{F}$, and according to Leray's theorem (see Ref. 5) we have, for all i, $H^i(X,\mathbf{F}) = H^i(U,\mathbf{F})$, where U is the covering (U_0,U_∞) of X. In particular, we get for the locally free $\mathbf{O}$ module $\mathbb{L}_\phi$ the exact sequence

$$0 \to \mathbb{L}_\phi(X) \xrightarrow{i} \mathbb{L}_\phi(U_0) \oplus \mathbb{L}_\phi(U_\infty)$$

$$\xrightarrow{\pi} \mathbb{L}_\phi(U_{0\infty}) \to H^1(X,\mathbb{L}_\phi) \to 0.$$

Here $i(f) = f|U_0 + f|U_\infty$ and $\pi(g_0,g_\infty) = g_0|U_{0\infty} - g_\infty|U_{0\infty}$. This exactness is preserved if we take the direct limit for U_0 tending to X_0 and U_∞ tending to X_∞, and thus we get

$$0 \to \mathbb{L}_\phi(X) \xrightarrow{i} \mathbb{L}_\phi(X_0) \oplus \mathbb{L}_\phi(X_\infty)$$

$$\xrightarrow{\pi} \mathbb{L}_\phi(S^1) \to H^1(X,\mathbb{L}_\phi) \to 0. \tag{4.1}$$

Embed the spaces $\mathbb{L}_\phi(X)$, $\mathbb{L}_\phi(X_0)$, $\mathbb{L}_\phi(X_\infty)$, and $\mathbb{L}_\phi(S^1)$ into $l^2(\mathbb{Z})^m$ as described in (2.3). In passing to the completion of those four spaces in $l^2(\mathbb{Z})^m$ we get the sequence

$$0 \to \mathbb{L}_\phi(X) \xrightarrow{i} W_\phi \oplus \mathbf{H}_{<0} \xrightarrow{\bar\pi} \mathbf{H},$$

with $\bar\pi$ the extension of π to $W_\phi \oplus \mathbf{H}_{<0}$, i.e., $\bar\pi(w \oplus h) = w - h$, $w \in W_\phi$, $h \in \mathbf{H}_{<0}$. We claim now that $\ker(\pi) = \ker(\bar\pi)$ and that $\mathrm{coker}(\pi) \cong \mathrm{coker}(\bar\pi)$. Assuming this for the moment, we can conclude directly (a) that W_ϕ belongs to the Grassmann manifold of $\mathbf{H}$ and, in particular (b) to which component it belongs. Specifically, from this assumption one sees directly that $\dim(\ker(\pi))$ is equal to the dimension of the kernel of the natural projection of W_ϕ to $\mathrm{diag}(\Lambda,\ldots,\Lambda)\mathbf{H}_{>0}$, and that the cokernel of this last map is isomorphic to the cokernel of π. Summarizing we have found the following theorem.

Theorem (4.1): W_ϕ lies in the Grassman manifold of $\mathbf{H}$, and its virtual dimension is

$$\dim H^0(X,\mathbb{L}_\phi) - m - \dim H^1(X,\mathbb{L}_\phi)$$
$$= c(\det(L_\phi)) - mg.$$

The last equality is the Riemann–Roch theorem for holomorphic vector bundles over a compact Reimann surface (see Ref. 6, p. 64). Furthermore, $\det(L_\phi)$ denotes the line bundle associated to $\det(\phi)$, and $c(\det(L_\phi))$ denotes its Chern class. What remains to be shown is that Eq. (4.1) is not affected by completion.

The kernel of π consists of (g,g) with $g \in \mathbb{L}_\phi(X_0) \cap \mathbf{H}^{\mathrm{an}}_{<0}$, and the kernel of $\bar\pi$ consists of (g,g) with $g \in W_\phi \cap \mathbf{H}_{<0}$. Thus we have to show $\mathbb{L}_\phi(X_0) \cap \mathbf{H}^{\mathrm{an}}_{<0} = W_\phi \cap \mathbf{H}_{\phi<0}$.

As for the cokernels, the exactness of Eq. (4.1) implies that there is a finite-dimensional space V in $\mathbf{H}^{\mathrm{an}}_{>0}$ such that $\mathbf{H}^{\mathrm{an}} = V \oplus \{\mathbb{L}_\phi(X_0) + \mathbf{H}^{\mathrm{an}}_{<0}\}$. Hence it is sufficient to show that $W_\phi + \mathbf{H}_{<0}$ is closed and that $\{W_\phi + \mathbf{H}_{<0}\} \cap V = \{0\}$.

A key role in the proof of these properties is played by the following.

Lemma (4.2): Let $\{f_j\} = \{\Sigma_{n\in\mathbb{Z}} a_{nj} z^n\}$ be a Cauchy sequence in $\mathbb{L}_1(X_0)$ w.r.t. the L^2 norm on S^1. Then $\{f_j\}$ converges uniformly on compact subsets of the interior X_0^0 of X_0 to a holomorphic f on X_0^0, with Laurent series $\Sigma_{n\in\mathbb{Z}} a_n z^n$ around x_∞, and in $l^2(\mathbb{Z})$ we have

$$\lim_{j\to\infty}\{a_{nj}\} = \{a_n\}.$$

Proof: For the domain X_0 we have an analog of the Cauchy integral formula. The role of $dx/(x-z)$ is taken over by a meromorphic differential $\omega(P,Q)$, $P,Q \in X$, satisfying the following propositions (see Ref. 7, Appendix §3).

Proposition (4.3): Take $P \neq Q$, $P \neq x_\infty$, and $Q \neq x_\infty$. Let s and t be local coordinates around P and Q, respectively;

then there are neighborhoods N of P and M of Q and a holomorphic function g on $s(N)\times t(M)$ such that

$$\omega(P^1,Q^1)=g(s(P^1),t(Q^1))ds,$$

for all $P^1\in N$ and $Q^1\in M$.

Proposition (4.4): If f is a function on X_0, holomorphic on X_0^0 and continuous on $X_{0\infty}$, then for all Q in X_0^0 we have

$$f(Q)=\frac{1}{2\pi i}\int_{X_{0\infty}} f(P)\omega(P,Q)\ .$$

Let f_j be as in the formulation of Lemma (4.2). By combining Propositions (4.3) and (4.4) we get for a suitable neighborhood M of Q in X_0^0 that, for all Q^1 in M,

$$|f_i(Q^1)-f_j(Q^1)|\leqslant\|f_i-f_j\|_2\cdot C,$$

with C a constant independent of f_i and f_j. Clearly this local uniform convergence implies the uniform convergence on general compacta, and the pointwise limit of the f_j defines a holomorphic f on X_0^0. The last statement of Lemma (4.2) is a direct consequence of the uniform convergence on a strip

$$\{P\,|P\in X,\ 1-\epsilon_2\leqslant|z(P)|\leqslant 1-\epsilon_1,\ \epsilon_2>\epsilon_1>0\}.$$

This proves the lemma. □

Consider now the case of the kernels. Take $w\in W_\phi\cap\mathbf{H}_{<0}$. Then there is a Cauchy sequence $\{w_j\}$ in $\mathbb{L}_{I_m}(X_0)$ such that

$$w=\lim_j \phi(w_j).$$

By applying the lemma to the components of the w_j, one sees that $\phi^{-1}w$ is the $\mathbf{L}^2$-boundary value of a holomorphic $\tilde{w}$: $X_0^0\to\mathbb{C}^m$. Since ϕ is holomorphic on a neighborhood of S^1, $\phi(\tilde{w})$ is holomorphic on a strip

$$\{P\in X_0|1>|z(P)|>1-\epsilon,\epsilon>0\},$$

and w is the L^2-boundary value of $\phi(\tilde{w})$. On the other hand, $w\in\mathbf{H}_{<0}$ and thus is the L^2-boundary value of a holomorphic function on the interior of X_∞. Now the edge of the wedge theorem (see Ref. 8, p. 75) implies that w is holomorphic on a neighborhood of $X_{0\infty}$, and $w\in W_\phi^{\mathrm{an}}\cap\mathbf{H}_{<0}^{\mathrm{an}}$. By applying Lemma (4.2) once more one sees that $W_{I_m}^{\mathrm{an}}=\mathbb{L}_{I_m}(X_0)$, and thus $W_\phi^{\mathrm{an}}=\mathbb{L}_\phi(X_0)$. This proves the equality of the kernels.

Let $V\subset\mathbf{H}_{\geqslant 0}^{\mathrm{an}}$ be such that $V\oplus\{\mathbb{L}_\phi(X_0)+\mathbf{H}_{<0}^{\mathrm{an}}\}$ is equal to $\mathbf{H}^{\mathrm{an}}$. Assume for the moment that $W_\phi+\mathbf{H}_{<0}$ is closed; then

$$\mathbf{H}=V+\{W_\phi+\mathbf{H}_{<0}\}=V+\mathrm{Im}(\bar{\pi}).$$

We show first that $V\cap\{W_\phi+\mathbf{H}_{<0}\}=\{0\}$. We have just proved that $W_\phi\cap\mathbf{H}_{<0}=\mathbb{L}_\phi(X_0)\cap\mathbf{H}_{<0}^{\mathrm{an}}$. Let Y be the orthocomplement of $W_\phi\cap\mathbf{H}_{<0}$ in $\mathbf{H}_{<0}$; then we have $\mathrm{Im}(\bar{\pi})=W_\phi\oplus Y$. Take any v in $V\cap\mathrm{Im}(\bar{\pi})$ and decompose it as $v=\phi(w_1)+y$, with $w_1\in W_{I_m}$ the L^2-boundary value of a holomorphic $\tilde{w}$: $X_0^0\to\mathbb{C}^m$. Again $\phi(w_1)$ is the L^2-boundary value of the holomorphic function $\phi(\tilde{w})$ on a strip

$$\{P\in X_0|1>z(P)|>1-\epsilon,\epsilon>0\},$$

with Laurent series $\Sigma_{n\in\mathbb{Z}}a_nz^n$. Since $v=\phi(w_1)+y$, with $y\in\mathbf{H}_{<0}$ and $v\in\mathbf{H}_{\geqslant 0}^{\mathrm{an}}$, one sees that $v=\Sigma_{n\geqslant 0}a_nz^n$ and that $\phi(\tilde{w})$ extends holomorphically to a neighborhood of $X_{0\infty}$. In other words, $\phi(w_1)\in\mathbb{L}_\phi(X_0)$, but then also $y\in\mathbf{H}_{<0}^{\mathrm{an}}$. Thus $v\in V\cap\{\mathbb{L}_\phi(X_0)+\mathbf{H}_{<0}^{\mathrm{an}}\}=\{0\}$. □

Finally, we show that $W_\phi+\mathbf{H}_{<0}$ is closed. Since it is the inverse image of the image of W_ϕ under the natural projection $p_+:\mathbf{H}\to\mathbf{H}_{\geqslant 0}$, we only need to prove that W_ϕ projects onto a closed subspace of $\mathbf{H}_{\geqslant 0}$. Let i_{W_ϕ} be the canonical embedding of W_ϕ into $\mathbf{H}$. We first show that $K=(\mathrm{Id}-p_+)\circ i_{W_\phi}: W_\phi\to\mathbf{H}_{<0}$ is compact.

Recall that each w in W_ϕ is the L^2-boundary value of a holomorphic $\tilde{w}$ on a strip

$$\{x\in X_0|1-\epsilon<|z(x)|<1,\epsilon>0\},$$

with ϵ only depending on ϕ. Choose a λ in $(1-\epsilon,1)$. For each w in W_ϕ, let $R_\lambda(w)$ be the evaluation of $\tilde{w}$ on

$$\{x|x\in\mathbb{P}^1(\mathbb{C}),\ |x|=\lambda\},$$

i.e., for all $t\in S^1$,

$$R_\lambda(w)(t)=\tilde{w}(\lambda t). \tag{4.2}$$

Clearly $R_\lambda(w)$ belongs to $\mathbf{H}^{\mathrm{an}}$, and its Laurent series around x_∞ is $\Sigma_{n\in\mathbb{Z}}(a_n\lambda^n)z^n$ if $\tilde{w}=\Sigma_{n\in\mathbb{Z}}a_nz^n$. This R_λ is a bounded map for Lemma (4.2) implying that

$$|R_\lambda(w)(x)|\leqslant\mathrm{const}\times\|w\|_2,$$

for all $x\in S^1$. In particular, one sees that

$$a_n\to(a_n\lambda^n)_{n<0} \tag{4.2}$$

defines a bounded map from W_ϕ to $\mathbf{H}_{<0}$.

Consider now the operator $C:\mathbf{H}_{<0}\to\mathbf{H}_{<0}$,

$$C((a_n))=(a_n\lambda^{-n})\ .$$

This C is a compact operator, for it is the limit w.r.t. the operator norm of the finite-dimensional bounded map C_m: $\mathbf{H}_{<0}\to\mathbf{H}_{<0}$, $m\geqslant 1$, given by

$$C_m((a_n))=(b_n)\ ,$$

with $b_n=0$ if $|n|>m$, and $b_n=a_n\lambda^{-n}$ for $|n|\leqslant m$. Thus the composition of Eq. (4.2) and C is compact, and that is exactly K.

From the foregoing we know that the kernel of i_W-K consists of $\mathbb{L}_\phi(X_0)\cap\mathbf{H}_{<0}^{\mathrm{an}}$. Let Y be the orthogonal complement in W_ϕ of the finite-dimensional space $\mathbb{L}_\phi(X_0)\cap\mathbf{H}_{<0}^{\mathrm{an}}$. Then $(i_{W_\phi}-K)(W_\phi)=(i_{W_\phi}-K)(Y)$, and $i_{W_\phi}-K$ is injective on Y.

According to Ref. 9, $(i_{W_\phi}-K)(Y)$ is closed if and only if there is a positive c such that

$$\|(i_{W_\phi}-K)(y)\|_2/\|y\|_2\geqslant c\ .$$

Let K_m be the composition of Eq. (4.2) and C_m; then the K_m converge to K in the operator norm. For all m, $i_{W_\phi}-K_m$ is injective on Y; therefore the image of Y under $i_{W_\phi}-K_m$ is closed if and only if

$$\|(i_{W_\phi}-K_m)(y)\|_2/\|y\|_2\geqslant c_m>0\ .$$

Assume that this holds for all m; then we have, for a sufficiently large N, that $\forall n\geqslant N$,

$$\frac{\|(i_{W_\phi}-K_n)y\|_2}{\|y\|_2}\geqslant\frac{\|(i_{W_\phi}-K_N)y\|_2-\|(K_N-K_n)y\|_2}{\|y\|_2}$$
$$\geqslant C_N-\|K_N-K_n\|>c>0\ .$$

But then we also have, for all $n\geqslant N$,

$$\frac{\|(i_{W_\phi}-K)y\|_2}{\|y\|_2} \geqslant \frac{\|(i_{W_\phi}-K_n)y\|_2 - \|(K-K_n)(y)\|_2}{\|y\|_2}$$
$$> c - \|K-K_n\| > 0, \quad \text{for } n \to \infty .$$

Hence we are left to prove that the image of $i_{W_\phi} - K_m$ is closed. Because $K_m(\mathbf{H}_{<0})$ is finite dimensional, $Y + K_m(\mathbf{H}_{<0})$ is a closed subspace of $\mathbf{H}$. But since $(i_{W_\phi} - K_m)(Y)$ has finite codimension in $Y + K_m(\mathbf{H}_{<0})$, it has to be closed, too. This proves the assertion.

[1]G. F. Helminck and G. F. Post, "Generalized multicomponent KP hierarchies," Universiteit Twente Memorandum 708, April 1988.

[2]A. Grothendieck, "Sur la classification des fibrés holomorphes sur la sphère de Riemann," Am. J. Math. **79**, 121 (1957).

[3]O. Forster, *Riemannsche Fläche* (Springer, Berlin, 1977).

[4]G. Segal and G. Wilson, "Loop groups and KdV-equations," Publ. Math. IHES **61**, 5 (1985).

[5]H. Grauert and R. Remmert, *Theory of Stein Spaces* (Springer, Berlin, 1979).

[6]R. C. Gunning, *Vector Bundles over Riemann Surfaces* (Princeton U.P., Princeton, NJ, 1967).

[7]A. Hurwitz and R. Courant, *Funktionen Theorie, Grundlehren der Mathematischen Wissenschaften in Einzeldarstellungen*, Vol. 3 (Springer, Berlin, 1964), 4th ed.

[8]J. W. de Roever, "Complex Fourier transformation and analytic functionals with unbounded carrier," Mathematical Center, Amsterdam, Tract 89, 1978.

[9]M. Schechter, *Principles of Functional Analysis* (Academic, New York, 1975).

The Ricci curvature of Diff $S^1/SL(2,\mathbb{R})$

M. J. Bowick and A. Lahiri
Physics Department, Syracuse University, Syracuse, New York 13244-1130

(Received 23 February 1988; accepted for publication 27 April 1988)

Previous calculations of the Ricci curvature for the manifold $\mathrm{Diff}(S^1)/S^1$ are extended to $\mathrm{Diff}(S^1)/\mathrm{SL}(2,\mathbb{R})$. These manifolds are distinguished by being coadjoint orbits of $\mathrm{Diff}(S^1)$ which admit compatible symplectic and complex structures, making them Kähler manifolds.

I. INTRODUCTION

In Refs. 1 string theory was formulated in terms of the holomorphic geometry of certain bosonic and ghost Fock bundles over the space of complex structures of loop space. This work was a geometrization or globalization of the work of Frenkel, Garland, and Zuckerman,[2] who gave the conditions for the consistency of string theory in terms of the Lie algebra cohomology for the Virasoro algebra with coefficients in the Fock space of the string. Since the natural complex structure of loop space depends on the choice of parametrization of the loop (circle), up to an overall rigid rotation, the possible complex structures to be considered are labeled by the orbit of a given complex structure under the action of $\mathrm{Diff}(S^1)$—the group of orientation-preserving diffeomorphisms of the circle. This orbit is isomorphic to $\mathrm{Diff}(S^1)/S^1$, for the complex structure given in Refs. 1. $\mathrm{Diff}(S^1)/S^1$ is a beautiful infinite-dimensional manifold. It admits a compatible symplectic and complex structure itself that renders it Kähler.[1,3] The appearance of $\mathrm{Diff}(S^1)/S^1$ can be understood from the Kostant–Kirillov theory of coadjoint orbits.[4] The coadjoint representation of $\mathrm{Diff}(S^1)$ is a representation of $\mathrm{Diff}(S^1)$ on its smooth dual space $(\mathrm{Diff}(S^1))^*$. At the algebraic level this space is the space of quadratic differentials on the circle. The coadjoint orbits of $\mathrm{Diff}(S^1)$ have all been classified.[5–9] In accordance with the general (finite-dimensional) theory they all admit nondegenerate symplectic structures. As manifolds they are all isomorphic to $\mathrm{Diff}(S^1)/H$, where H is the common isotropy group of any point on the orbit. Only two of these orbits admit an integrable complex structure that makes them Kähler manifolds.[1,3,8] These are $\mathrm{Diff}(S^1)/S^1$ and $\mathrm{Diff}(S^1)/\mathrm{SL}(2,\mathbb{R})$.

In Refs. 1 it was shown that the critical dimension of the bosonic string (or, equivalently, the conformal anomaly) can be understood as arising from the nonzero curvature of the (anti)canonical line bundle of $\mathrm{Diff}(S^1)/S^1$. For a Kähler manifold this curvature is given by the Ricci form of the manifold itself. For $\mathrm{Diff}(S^1)/S^1$ the result is[1]

$$R(L_{-m},L_n) = (-\tfrac{26}{12}m^3 + \tfrac{1}{6}m)\delta_{m,n}, \tag{1}$$

where $R(L_{-m},L_n)$ is the $\bar{m}n$ component of the Ricci form. It is the "26" in this result that yields the critical dimension of 26 for the bosonic string. This result can be supersymmetrized[10,11,12] and has been extended to the case of strings moving on a group manifold.[13]

In this paper we extend this calculation, for completeness, to the other Kähler coadjoint orbit, $\mathrm{Diff}(S^1)/\mathrm{SL}(2,\mathbb{R})$. In terms of the setting of Refs. 1 this space would correspond to the orbit under $\mathrm{Diff}(S^1)$ of any projective $\mathrm{SL}(2,\mathbb{R})$-invariant complex structure. As the cocycle of the Virasoro algebra is cohomologous to the $\mathrm{SL}(2,\mathbb{R})$-invariant cocycle, it is convenient to choose $\mathrm{SL}(2,\mathbb{R})$-invariant structures in applications to string or conformal field theory. $\mathrm{Diff}(S^1)/\mathrm{SL}(2,\mathbb{R})$ has also been considered in Ref. 14, where it is interpreted geometrically as the space of all Lorentz metrics on the single-sheeted hyperboloid obtained from the standard [$\mathrm{SL}(2,\mathbb{R})$-invariant] Lorentz metric by a global conformal diffeomorphism.

II. PRESENTATION

For $\mathrm{Diff}(S^1)/\mathrm{SL}(2,\mathbb{R})$ the type (1,0) and (0,1) eigenspaces of the complex structure J are the vector fields on the circle L_m with $m>1$ and $m<1$, respectively. That is, at the origin,

$$JL_m = \begin{cases} iL_m, & m>1, \\ -iL_m, & m<-1, \end{cases} \tag{2}$$

with $L_m = z^{-m+1}\,d/dz$.

The Kähler metric is given by

$$\omega(L_m,L_n) = a(m^3-m)\delta_{m,-n} = \omega(m), \tag{3}$$

where we introduce the notation $\omega(m)$ for later convenience. This is the $\mathrm{SL}(2,\mathbb{R})$-invariant form of the Kähler metric given in Refs. 1 for $\mathrm{Diff}(S^1)/S^1$. The complex structure gives the following decomposition of the corresponding complexified Lie algebra (at the origin):

$$\underline{\mathrm{Diff}\,S^1} = \underline{L}_+ + \underline{L}_- + \underline{\mathrm{SL}(2,\mathbb{R})}_{\mathbb{C}} \tag{4}$$

where $\underline{L}_+$ and $\underline{L}_-$ correspond to the type (1,0) and (0,1) vector fields, respectively, and $\mathrm{SL}(2,\mathbb{R})$ is spanned by L_1, L_0, and L_{-1}.

To compute the curvature we introduce the Toeplitz operator[1,15,16]

$$\varphi(X) = \mathcal{L}_X - \nabla_X, \tag{5}$$

where ∇_X denotes the covariant derivative in the direction of vector field X with respect to the Kähler connection, and $\mathcal{L}_X$ is the Lie derivative. The operator $\varphi(X)$ can be thought of as operating within the space $\underline{L}_+$ since both $\mathcal{L}_X$ and ∇_X preserve the complex structure and thus the decomposition (4). The tensorial character of the Toeplitz operators allows their action on a vector field Y to be computed at the origin by extending Y away from the origin in any convenient manner. Specifically, one can extend Y away from the origin

 0022-2488/88/091979-03$02.50

such that the covariant derivative ∇_X always vanishes and $\varphi(X) = \mathscr{L}_X$. For $X \in \mathrm{SL}(2,\mathbb{R})$ this is guaranteed since X vanishes at the origin. That is, if we denote the identity of $\mathrm{Diff}(S^1)$ by e, then $X(e) = 0$ for $X \in \mathrm{SL}(2,\mathbb{R})$, so that $\nabla_X = 0$. For $X \in \underline{L}_-$, let $(Y)_+$ denote the (1,0) component of the vector field $\bar{Y}$. It is a holomorphic vector field and hence

$$\nabla_X (Y)_+ = \bar{\partial}_X ((Y)_+) = 0 \tag{6}$$

at the origin. Here we have used the fact that a Kähler connection agrees with the $\bar{\partial}$ operator in the antiholomorphic directions, i.e., $\nabla_X = \bar{\partial}_X$ for $X \in \underline{L}_-$. For $X \in \underline{L}_+$, $\varphi(X)$ is defined by its skew-Hermitian character. Thus we conclude

$$\varphi(X) = \begin{cases} \mathrm{ad}\, X, & X \in \mathrm{SL}(2,\mathbb{R}), \\ \pi_+ \circ \mathrm{ad}\, X, & X \in \underline{L}_-, \\ -\varphi(\bar{X})^+, & X \in \underline{L}_+, \end{cases} \tag{7}$$

where π_+ denotes the projection onto the subspace $\underline{L}_+$, "ad" is the adjoint action, and $X \to \bar{X}$ is involution that maps $\underline{L}_+ \to \underline{L}_-$. In terms of components,

$$\begin{aligned} &\varphi(L_k) = \mathrm{ad}\, L_k, \quad k = 1,0,-1, \\ &\varphi(L_{-m}) = \pi_+ \circ \mathrm{ad}\, L_{-m}, \quad m > 1, \\ &\varphi(L_m) = -\varphi(L_{-m})^+, \quad m > 1. \end{aligned} \tag{8}$$

The Riemann tensor is given by

$$\begin{aligned} R(X,Y) &= [\nabla_X, \nabla_Y] - \nabla_{[X,Y]} \\ &= [\varphi(X), \varphi(Y)] - \varphi([X,Y]), \end{aligned} \tag{9}$$

using $[\mathscr{L}_X, \mathscr{L}_Y] = \mathscr{L}_{[X,Y]}$ and $[\mathscr{L}_X, \nabla_Y] = \nabla_{[X,Y]}$, which follows from left invariance under vector fields X and Y. Thus the Riemann tensor (and subsequently the Ricci form) can be computed entirely from our knowledge of the action of the Toeplitz operators (7) and (8).

This gives

$$\begin{aligned} &\varphi(L_m) L_p = -(p+2m) \frac{\omega(p)}{\omega(p+m)} L_{m+p}, \quad m > 1, \\ &\varphi(L_{-m}) L_p = -\theta(p-m-1)(p+m) L_{-m+p}, \\ &\varphi(L_k) L_p = (k-p) L_{k+p}, \quad k = 1,0,-1, \end{aligned} \tag{10}$$

with $\omega(p) = a(p^3 - p)$. The Riemann tensor is then

$$R(L_{-m}, L_n) L_p = R_{\bar{m}n}{}^q{}_p, \tag{11}$$

with

$$\begin{aligned} R_{\bar{m}n}{}^q{}_p = &\left[(2m+q)(m+q+n) \frac{\omega(m+q-n)}{\omega(m+q)} - (q+n)(2m+q-n) \frac{\omega(q-n)}{\omega(q)} \theta(q-n-1) \right. \\ &\left. - (m+n) \begin{cases} 2m-2n+q, & m \geqslant n \\ (q+n-m) \dfrac{\omega(q-n+m)}{\omega(q)}, & m \leqslant n \end{cases} \right] \delta_{q+m,n+p}, \end{aligned} \tag{12}$$

where $m,n,p,q > 1$.

The Ricci form can now be computed in two ways.[1,16] With a Kähler manifold one can trace over diagonal terms $p = q$. Since a Kähler manifold is also Riemannian one can equivalently trace over $q = n$. The latter trace has the advantage of rendering the absolute convergence of the Ricci form obvious. The diagonal terms of (12) are

$$\begin{aligned} R_{\bar{m}n}{}^n{}_p = \delta_{mp} \Bigg\{ &(2m+q)(m+2n) \frac{\omega(m)}{\omega(m+n)} \\ &- (m+n) \begin{cases} 2m-n, & m \geqslant n \\ (2n-m) \dfrac{\omega(m)}{\omega(n)}, & m \leqslant n \end{cases} \Bigg\}. \end{aligned} \tag{13}$$

The Ricci form is then

$$\begin{aligned} R_{\bar{m}p} = \delta_{mp} \Bigg[&\sum_{n=2}^{\infty} \left\{ (2m+n)(2n+m) \frac{\omega(m)}{\omega(n+m)} \right. \\ &\left. - (m+n)(2n-m) \frac{\omega(m)}{\omega(n)} \right\} \\ &+ \sum_{n=2}^{m} (m+n)(2n-m) \frac{\omega(m)}{\omega(n)} \\ &- \sum_{n=2}^{m} (m+n)(2m-n) \Bigg]. \end{aligned} \tag{14}$$

Since the overall trace is convergent we can shift arguments $(n \to n+m)$ in the second sum to give

$$\begin{aligned} R_{\bar{m}p} &= \delta_{mp} \Bigg[-\sum_{n=2}^{m+1} (m+n)(2-m) \frac{\omega(m)}{\omega(n)} \\ &\quad + \sum_{n=2}^{m} (m+n)(2n-m) \frac{\omega(m)}{\omega(n)} \\ &\quad - \sum_{n=2}^{m} (m+n)(2m-n) \Bigg] \\ &= -\delta_{mp} \sum_{n=2}^{m+1} (m+n)(2m-n) \\ &= -\tfrac{13}{6}(m^3 - m) \delta_{mp}. \end{aligned} \tag{15}$$

As required this is the $\mathrm{SL}(2,\mathbb{R})$-invariant analog of the result (1) for $\mathrm{Diff}(S^1)/S^1$, i.e., we obtain the expected critical dimension 26 in string applications and the $\mathrm{SL}(2,\mathbb{R})$-invariant cocycle of the Virasoro algebra (or conformal anomaly) from the Ricci curvature of $\mathrm{Diff}(S^1)/\mathrm{SL}(2,\mathbb{R})$.

ACKNOWLEDGMENT

M. J. B.'s work is supported in part by the U.S. Department of Energy (DOE) under Contract No. DE-FG02-85ER40231 (Outstanding Junior Investigator Program).

[1]M. J. Bowick and S. G. Rajeev, Phys. Rev. Lett. **58**, 535 (1987); Nucl. Phys. B **293**, 348 (1987); "The complex geometry of string theory and loop space," in *Frontiers in Particle Theory, Proceedings of the Johns Hopkins Workshop on Current Problems in Particle Theory II*, edited by Yi-shi Duan, G. Domokos, and S. Kovesi-Domokos (World Scientific, Singapore, 1988); M. J. Bowick, "The geometry of string theory," in *Eighth Workshop on Grand Unification*, edited by K. C. Wali (World Scientific, Singapore, 1988).
[2]I. B. Frenkel, H. Garland, and G. J. Zuckerman, Proc. Natl. Acad. Sci. USA **83**, 8442 (1986).
[3]A. A. Kirillov and D. V. Yuriev, Funct. Anal. Appl. **20**, 322 (1986); A. A. Kirillov, *ibid.* **21**, 122 (1987); A. A. Kirillov and D. V. Yuriev, *ibid.* **21**, 35 (1987) (in Russian).
[4]V. Guillemin and S. Sternberg, *Symplectic Techniques in Physics* (Cambridge U.P., Cambridge, 1984).
[5]V. F. Lazutkin and T. F. Pankratova, Funct. Anal. Appl. **9**, 306 (1975).
[6]A. A. Kirillov, Lec. Notes Math. **970**, 101 (1982).
[7]G. Segal, Commun. Math. Phys. **80**, 307 (1981).
[8]E. Witten, Commun. Math. Phys. **144**, 1 (1988).
[9]I. Bakas, Nucl. Phys. B **302**, 189 (1988).
[10]K. Pilch and N. Warner, J. Class. Quant. Gravit. **4**, 1183 (1987).
[11]D. Harari, D. K. Hong, P. Ramond, and V. Rodgers, Nucl. Phys. B **294**, 556 (1987).
[12]Z. Y. Zhao, K. Wu, and T. Saito, Phys. Lett. B **199**, 37 (1987).
[13]J. Mickelsson, Commun. Math. Phys. **112**, 653 (1987).
[14]B. Kostant and S. Sternberg, Ann. Phys. (NY) **176**, 49 (1987).
[15]D. Freed, in *Infinite-Dimensional Groups with Applications*, edited by V. Kac (Springer, Berlin, 1985); Ph.D. thesis, University of California at Berkeley, 1985.
[16]B. Zumino, "The geometry of the Virasoro group for physicists," to be published in *Proceedings of the Cargese Summer School on Particle Physics*, 1987.

Wiener analysis of a binary hysteresis system

J. Nakayama and E. Omori
Department of Electronics, Kyoto Insitute of Technology, Matsugasaki, Kyoto 606, Japan

(Received 29 December 1987; accepted for publication 11 May 1988)

Nonlinear transformation of an independent Gaussian sequence by a binary hysteresis is studied by Wiener's method [*Nonlinear Problems in Random Theory* (MIT, Cambridge, MA, 1958)]. The output sequence of the hysteresis is represented by a series of Wiener–Hermite functionals, which are random and orthogonal (uncorrelated) to each other. From the hysteresis equation, a set of linear equations for the Wiener–Hermite functionals is derived without approximation. The equations directly show that the hysteresis is decomposed into equivalent subsystems consisting of static nonlinear elements, multipliers, adders, and linear filters. They are solved by use of random numbers generated in a computer in order to get a sample output sequence approximated by a finite sum of the Wiener–Hermite functionals. Another set of equations is derived for the mean-square values of the Wiener–Hermite functionals. In terms of its numerical solutions, the convergence property of the Wiener–Hermite functional series is discussed.

I. INTRODUCTION

In this paper we study the nonlinear transformation of a Gaussian random sequence by a binary hysteresis, shown in Fig. 1. Such a hysteresis has received much interest in control and biophysics because it is a simple model of a transfer relay in a control system[1] and a basic element in a neuron system.[2]

We deal with such a problem by the Wiener method,[3–5] which has been applied to nonlinear stochastic problems such as the theory of turbulence,[6,7] wave propagation and scattering in random media,[8–10] analysis and identification of physical and biological systems,[2–4,11–15] and stochastic differential equations.[16] In principle, the Wiener theory gives a mathematical method to represent the output of nonlinear systems with the Gaussian input signal in terms of Wiener–Hermite functionals. The coefficients of such a functional expansion are deterministic functions called Wiener kernels, in terms of which systems are identified.[3,11–13] Exact Wiener kernels are known for several nonlinear systems: a memoryless nonlinear element,[4] the so-called sandwich system,[14–15] and a bilinear system.[14] But no exact Wiener analysis has been made for a hysteresis system. Since Wiener kernels are difficult to determine exactly for a system with a strong nonlinearity and memory, it is often said that Wiener theory should be tried in case of a mild nonlinearity.[5]

But even though the binary hysteresis has a strong nonlinearity with memory, this paper demonstrates that an exact Wiener analysis may be carried out in a special case where the input is an independent, zero-mean, Gaussian sequence. First, we derive a hysteresis equation describing the input–output relationship. We develop the output sequence of the hysteresis in terms of Wiener–Hermite functionals. Using a theorem on the Wiener–Hermite functionals,[17] we then transform the hysteresis equation into a set of linear equations for the Wiener–Hermite functionals. Since Wiener kernels have been believed to be essentially necessary for calculating Wiener–Hermite functionals, many studies have been made to determine them for various problems.[2,5–16] However, the linear equations enable us to calculate directly the Wiener–Hermite functionals without knowledge of Wiener kernels. Taking advantage of this, we carry out a computer simulation in order to get a sample output sequence approximated by a finite sum of the Wiener–Hermite functionals. Then we derive another set of equations for the quadratic mean value of the Wiener–Hermite functionals, in terms of which the convergence property of the Wiener–Hermite functional series is discussed.

We summarize basic notations and definitions of the Wiener–Hermite functionals in the Appendix, because various notations are available.[2–5]

II. HYSTERESIS EQUATION

Let $X_p(\omega)$ $(p = 0, \pm 1, \pm 2, \ldots)$ be an input random sequence for the hysteresis shown in Fig. 1, where p denotes the discrete time and ω is a probability parameter describing a sample point in the sample space Ω. (However, we often drop the probability parameter ω in the equations below.) We assume the input X_p is a Gaussian sequence with zero mean and unit variance:

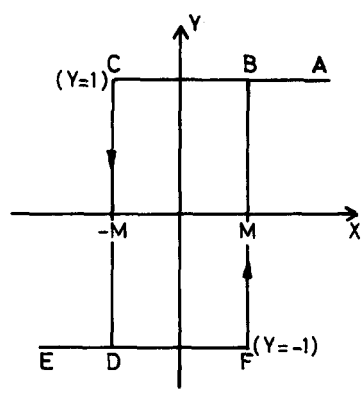


FIG. 1. The input–output characteristic of a binary hysteresis, which transforms a Gaussian sequence into a binary sequence. The threshold level is denoted by $\pm M$.

 0022-2488/88/091982-08$02.50

$$\langle X_p \rangle = 0, \quad \langle X_p^2 \rangle = 1 \quad (p = 0, \pm 1, \pm 2, \ldots), \tag{1}$$

where the angle brackets denote the ensemble average over the sample space Ω. The output of the hysteresis is denoted by $\{Y_p(\omega)\}$, which is a binary sequence taking ± 1. We implicitly assume, however, that the sample space Ω is of function space type.[18] Thus a sample point ω and a function $g(\omega)$ are regarded as an infinite-dimensional vector given by a sample sequence $\{X_p\}$ and as a functional of $\{X_p\}$, respectively:

$$\begin{aligned} &\omega = (\ldots, X_2, X_1, X_0, X_{-1}, \ldots), \\ &g(\omega) = g(\ldots, X_2, X_1, X_0, X_{-1}, \ldots). \end{aligned} \tag{2}$$

Here ω_p, the pth component of ω, is given by $\omega_p = X_p(\omega)$. Furthermore, a shift of a sample sequence by q, $\{X_p\} \to \{X_{p+q}\}$, induces a shift T^q of a sample point ω in Ω, namely, $X_{p+q}(\omega) = X_p(T^q\omega)$. In the case of an independent Gaussian sequence,[18] the shift T^q is a measure-preserving transformation with a group property: $T^0 = 1$ (identity); $T^{p+q} = T^p \cdot T^q$. Once the shift T is so defined, $g(T^p\omega)$ becomes a stationary random function of p for any random variable $g(\omega)$, where $T^p\omega$ and $g(T^p\omega)$ are defined as

$$\begin{aligned} &T^p\omega = (\ldots, X_{p+2}, X_{p+1}, X_p, X_{p-1}, \ldots), \\ &g(T^p\omega) = g(\ldots, X_{p+2}, X_{p+1}, X_p, X_{p-1}, \ldots). \end{aligned} \tag{3}$$

Further, we define even and odd functionals of ω. Setting $-\omega = (\ldots, -X_2, -X_1, -X_0, -X_{-1}, \ldots)$ by (2), we may define even and odd functionals by the relations

$$\begin{aligned} g(\omega) &= g(-\omega) \quad \text{(even functional)}, \\ &= -g(-\omega) \quad \text{(odd functional)}. \end{aligned} \tag{4}$$

Because the hysteresis is with memory, the present output Y_p is not uniquely determined by the present input X_p. When the one-step past output Y_{p-1} is equal to -1 or 1, however, the present output Y_p is uniquely determined by the curve EDFBA or EDCBA in Fig. 1. Thus we get

$$\begin{aligned} &Y_p = C(X_p - M) \quad (Y_{p-1} = -1), \\ &Y_p = C(X_p + M) \quad (Y_{p-1} = 1), \end{aligned} \tag{5}$$

where $\pm M$ is the threshold of the hysteresis and $C(x)$ is a clip function:

$$\begin{aligned} C(x) &= 1 \quad (x > 0) \\ &= -1 \quad (x < 0). \end{aligned} \tag{6}$$

Since $Y_p(\omega) = \pm 1$, from (5) we obtain the hysteresis equation:

$$\begin{aligned} Y_p &= (\tfrac{1}{2})(Y_{p-1} + 1)C(X_p + M) \\ &\quad - (\tfrac{1}{2})(Y_{p-1} - 1)C(X_p - M) \\ &= D(X_p)Y_{p-1} + S(X_p). \end{aligned} \tag{7}$$

Here $D(x)$ and $S(x)$ are discontinuous functions given by (see Fig. 2)

$$\begin{aligned} &D(x) = (\tfrac{1}{2})[C(x + M) - C(x - M)], \\ &S(x) = (\tfrac{1}{2})[C(x + M) + C(x - M)], \end{aligned} \tag{8}$$

which are even and odd functions, respectively:

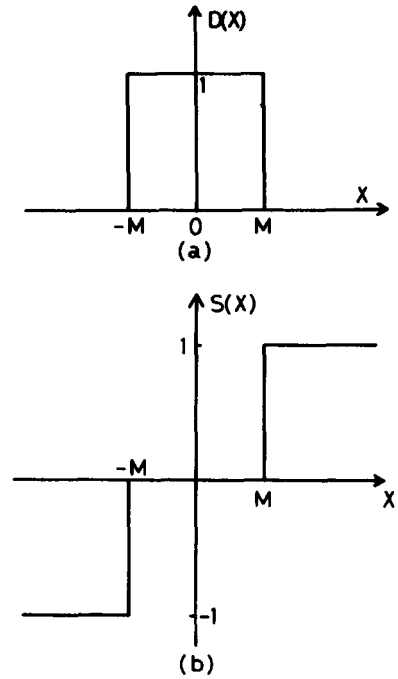


FIG. 2. Discontinuous functions $D(x)$ and $S(x)$. They are even and odd functions, respectively.

$$\begin{aligned} &D(x) = D(-x), \quad S(x) = -S(-x), \\ &D(x) = D^2(x), \quad S(x) = S^3(x), \\ &D^2(x) + S^2(x) = 1, \quad D(x)S(x) = 0. \end{aligned} \tag{9}$$

Note that the hysteresis equation (7) is formally linear with respect to the output $\{Y_p\}$, even though the hysteresis is nonlinear and discontinuous in the input–output relationship.

A formal solution of (7) is easily obtained by the iteration as[19]

$$\begin{aligned} Y_p(\omega) &= S(X_p) + D(X_p)S(X_{p-1}) \\ &\quad + D(X_p)D(X_{p-1})S(X_{p-2}) + \cdots. \end{aligned} \tag{10}$$

This shows that, due to the causality, the present output is a functional of the present and past input. Thus, rewriting (3), we may put

$$Y_p(\omega) = g(X_p, X_{p-1}, X_{p-2}, \ldots) = g(T^p\omega) = Y_0(T^p\omega), \tag{11}$$

which means that $Y_p(\omega)$ can be obtained from $Y_0(\omega)$ by the measure-preserving transformation.

We note that (10) is valid for any input sequence with any correlation function. It is difficult, however, to calculate from (10) statistical properties of the output Y_p for a correlated input sequence. But when X_p is an independent Gaussian sequence enjoying

$$\langle X_p X_q \rangle = \delta(p,q), \quad p,q = 0, \pm 1, \pm 2, \ldots, \tag{12}$$

$\delta(p,q)$ being Kronecker's delta, we may carry out an exact Wiener analysis as is described in the next section.

III. WIENER–HERMITE EXPANSION OF THE OUTPUT SEQUENCE

Two functional expansion techniques are well known: Volterra functional series[5,12] and Wiener–Hermite expan-

sion (notations and definitions of the latter are summarized in the Appendix). Because the hysteresis is discontinuous in the input–output relationship, Volterra functional series may not be applicable. On the other hand, one may represent the hysteresis by the Wiener–Hermite expansion as in (A11), because $\{Y_p\}$ has a finite variance.

Because the hysteresis is symmetrical with respect to the origin (see Fig. 1), Y_p becomes an odd functional as (4), namely,

$$\begin{aligned} Y_p &= g(X_p, X_{p-1}, X_{p-2}, \cdots) \\ &= -g(-X_p, -X_{p-1}, -X_{p-2}, \cdots) . \end{aligned} \tag{13}$$

Since odd-degree Wiener–Hermite functionals are odd by (A15), we write Y_p in terms of only odd-degree Wiener–Hermite functionals as

$$Y_p = \sum_{n=0}^{\infty} y_{2n+1}(p) , \tag{14}$$

$$y_{2n+1}(p) = I_{2n+1}[G_{2n+1}(\cdot - p), \omega] . \tag{15}$$

Here $I_{2n+1}[G_{2n+1}(\cdot - p), \omega]$ is the $(2n+1)$th degree Wiener–Hermite functional defined by (A12). (For convenience, however, the notations $y_{2n+1}(p)$ and $I_{2n+1}[G_{2n+1}(\cdot - p), \omega]$ are doubly used for the Wiener–Hermite functional.) Due to the causality, the Wiener kernel $G_{2n+1}(p_1, p_2, \cdots, p_{2n+1})$ should be zero if any of its arguments is positive, that is,

$$\begin{aligned} &G_{2n+1}(p_1, p_2, \cdots, p_{2n+1}) = 0 , \\ &\qquad \text{for any } p_i > 0 \quad (i = 1,2,3,\cdots,2n+1). \end{aligned} \tag{16}$$

Next we write $D(X_p)$ and $S(X_p)$ as

$$D(X_p) = \sum_{n=0}^{\infty} d_{2n} H_{2n}(X_p) = \sum_{n=0}^{\infty} I_{2n}[D_{2n}(\cdot - p), \omega] , \tag{17}$$

$$S(X_p) = \sum_{n=0}^{\infty} s_{2n+1} H_{2n+1}(X_p) , \tag{18}$$

where the symmetrical Wiener kernel D_{2n} and constants d_{2n} and s_{2n+1} are given by

$$\begin{aligned} &D_{2n}(p_1 - p, p_2 - p, \cdots, p_{2n} - p) \\ &\quad = d_{2n} \delta(p_1, p) \delta(p_2, p) \cdots \delta(p_{2n}, p) , \end{aligned} \tag{19}$$

$$d_{2n} = \frac{1}{(2n)!} \int_{-M}^{M} D(x) \cdot H_{2n}(x) G(x) dx , \tag{20}$$

$$s_{2n+1} = \frac{1}{(2n+1)!} \int_{-\infty}^{\infty} S(x) \cdot H_{2n+1}(x) G(x) dx , \tag{21}$$

which depend on the threshold M of the hysteresis. Here $G(x)$ is the Gaussian distribution (A1). By (9) and Parseval's relation, we obtain the useful relations

$$\begin{aligned} &d_0 = \sum_{n=0}^{\infty} (2n)!\, d_{2n}^2 , \\ &\sum_{n=0}^{\infty} (2n)!\, d_{2n}^2 + \sum_{n=0}^{\infty} (2n+1)!\, s_{2n+1}^2 = 1 . \end{aligned} \tag{22}$$

In order to analyze the hysteresis equation, we then develop the product term $Y_{p-1} \cdot D(X_p)$ in (7) in terms of Wiener–Hermite functionals. This can be done easily by (A19) as

$$\begin{aligned} &Y_{p-1} \cdot D(X_p) \\ &\quad = \left(\sum_{m=0}^{\infty} I_{2m+1}[G_{2m+1}(\cdot - p + 1), \omega] \right) \\ &\qquad \times \left(\sum_{n=0}^{\infty} I_{2n}[D_{2n}(\cdot - p), \omega] \right) \\ &\quad = \sum_{m=0}^{\infty} \sum_{n=0}^{\infty} \sum_{k=0}^{m} \frac{(2m+1-2k+2n)!\,(2n+2k)!}{(2m+1-2k)!\,(2k)!\,(2n)!} \\ &\qquad \times I_{2m+1}[\{G_{2m+1-2k+2n}(\cdot - p + 1), \\ &\qquad D_{2n+2k}(\cdot - p)\}_{2n}, \omega] . \end{aligned} \tag{23}$$

Here $\{G_{2m+1-2k+2n}(\cdot - p + 1), D_{2n+2k}(\cdot - p)\}_{2n}$ is a $2n$-dimensional inner product defined by (A10), which can be calculated as follows:

$$\begin{aligned} &\{G_{2m+1-2k+2n}(\cdot - p + 1), D_{2n+2k}(\cdot - p)\}_{2n} \\ &\quad = \sum_{p_1', p_2', \cdots, p_{2n}' = -\infty}^{\infty} G_{2m+1-2k+2n}(p_1' - p + 1, p_2' - p + 1, \cdots, p_{2n}' - p + 1, p_1 - p + 1, \cdots, p_{2m-2k+1} - p + 1) \\ &\qquad \times D_{2n+2k}(p_1' - p, p_2' - p, \cdots, p_{2n}' - p, p_{2m+2-2k} - p, \cdots, p_{2m+1} - p) \qquad (24) \\ &\quad = G_{2m+1-2k+2n}(1,1,\cdots,1, p_1 - p + 1, \cdots, p_{2m-2k+1} - p + 1) \\ &\qquad \times d_{2n+2k} \delta(p_{2m+2-2k}, p) \delta(p_{2m+3-2k}, p) \cdots \delta(p_{2m+1}, p) , \qquad (25) \\ &\quad = \begin{cases} 0 & (n > 0) , \\ G_{2m+1-2k}(p_1 - p + 1, \cdots, p_{2m-2k+1} - p + 1) \times d_{2k} \delta(p_{2m+2-2k}, p) \delta(p_{2m+3-2k}, p) \cdots \delta(p_{2m+1}, p) & (n = 0) . \end{cases} \qquad (26) \end{aligned}$$

Equation (24) is due to the definition (A10). Substituting (19) into the right-hand side of (24) and carrying out the summation, one easily finds (25), in which the number of 1 in the variables in $G_{2m+1-2k+2n}$ is equal to $2n$. By (16), however, the right-hand side of this equation vanishes for any positive integer n. Thus we finally obtain (26), which simplifies the right-hand side of (23) as

$$\begin{aligned} &Y_{p-1} \cdot D(X_p) \\ &\quad = \sum_{m=0}^{\infty} \sum_{k=0}^{m} I_{2m+1}[\{G_{2m+1-2k}(\cdot - p + 1), \\ &\qquad D_{2k}(\cdot - p)\}_0, \omega] , \end{aligned} \tag{27}$$

which can be simplified further. By (A12) and (26), we obtain

$$I_{2m+1}[\{G_{2m+1-2k}(\cdot-p+1),D_{2k}(\cdot-p)\}_0,\omega]$$
$$=\sum_{p_1,p_2,\ldots,p_{2m-2k+1}=-\infty}^{p-1} G_{2m+1-2k}(p_1-p+1,\ldots,$$
$$p_{2m-2k+1}-p+1)\cdot d_{2k}$$
$$\times H^{(2m+1)}[X_{p_1},X_{p_2},\ldots,X_{p_{2m-2k+1}},X_p,\ldots,X_p], \quad (28)$$

where the upper limit of the summation becomes $p-1$ by the causality (16). Because variables $p_1,p_2,\ldots,p_{2m-2k+1}$ in (28) are always smaller than p, it follows from (A1), (A4), and (A9) that

$$H^{(2m+1)}[X_{p_1},X_{p_2},\ldots,X_{p_{2m-2k+1}},X_p,\ldots,X_p]$$
$$=H^{(2m-2k+1)}[X_{p_1},X_{p_2},\ldots,X_{p_{2m-2k+1}}]$$
$$\times H_{2k}(X_p) \quad (p_1,p_2,p_3,\ldots,p_{2m-2k+1}<p). \quad (29)$$

Therefore we obtain from (28) and (29)

$$I_{2m+1}[\{G_{2m+1-2k}(\cdot-p+1),D_{2k}(\cdot-p)\}_0,\omega]$$
$$=I_{2m-2k+1}[G_{2m+1-2k}(\cdot-p+1),\omega]$$
$$\times d_{2k}\cdot H_{2k}(X_p)$$
$$=y_{2m-2k+1}(p-1)\times d_{2k}\cdot H_{2k}(X_p). \quad (30)$$

Inserting this relation into (27), we finally obtain a simple expression of the product,

$$Y_{p-1}\cdot D(X_p)$$
$$=\sum_{m=0}^{\infty}\sum_{k=0}^{m} y_{2m-2k+1}(p-1)\times d_{2k}\cdot H_{2k}(X_p). \quad (31)$$

This simple result is essentially due to the fact that, for any p, X_p and Y_{p-1} are statistically independent from each other in the case of the independent sequence $\{X_p\}$. This property will be used often below. Substituting (14), (18), and (31) into (7), we get

$$\sum_{m=0}^{\infty}\left[y_{2m+1}(p)-\sum_{k=0}^{m} y_{2m-2k+1}(p-1)\right.$$
$$\left.\cdot d_{2k}\cdot H_{2k}(X_p)-s_{2m+1}\cdot H_{2m+1}(X_p)\right]=0, \quad (32)$$

which holds in the ensemble mean-square sense. Because Wiener–Hermite functionals with different degrees are orthogonal (uncorrelated) to each other, we obtain from (32) a set of linear equations for $y_{2m+1}(p)$:

$$y_{2m+1}(p)=\sum_{k=0}^{m} y_{2m-2k+1}(p-1)$$
$$\times d_{2k}\cdot H_{2k}(X_p)+s_{2m+1}\cdot H_{2m+1}(X_p)$$
$$(m=0,1,2,3,\ldots), \quad (33)$$

where coefficients $d_{2k}\cdot H_{2k}(X_p)$ are time dependent and random.

We may represent (33) by an equivalent network in Fig. 3, which illustrates how the hysteresis with an independent Gaussian input is decomposed into subsystems consisting of static nonlinear elements, multipliers, adders, and linear filters. Note that this network is valid only for the Gaussian sequence with (1) and (12).

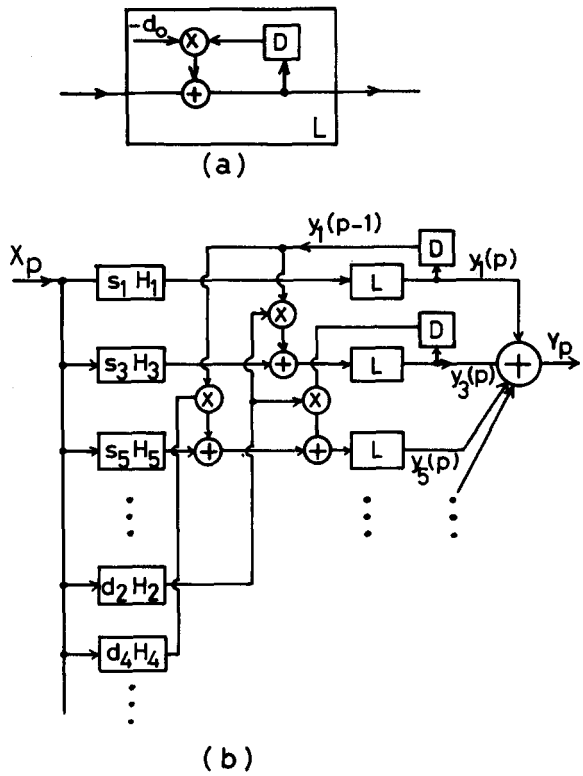


FIG. 3. (a) Linear filter L and (b) equivalent network of the hysteresis with an independent zero-mean Gaussian sequence $\{X_p\}$. The one-step delay line is denoted by D, and H_n is the nth-degree Hermite polynomial, representing a static nonlinear element.

In many cases Wiener kernels are definitely needed to calculate Wiener–Hermite functionals. Therefore theoretical and experimental works have been carried out for determining Wiener kernels for various nonlinear problems.[2–16] Since Wiener kernels are multivariable functions, it is not easy to obtain higher-degree Wiener kernels in a general case. In our case, however, without knowledge of Wiener kernels we may easily calculate Wiener–Hermite functionals $y_{2m+1}(p)$ from (33) by the iteration. Taking advantage of this, we will calculate higher-degree Wiener–Hermite functionals later to discuss the convergence property of the Wiener–Hermite expansion (14).

We should note that, if we take a spectral representation of the Wiener–Hermite functionals, asymmetrical Wiener kernels in the spectral domain can be easily obtained from (33). Discussions on the Wiener kernels, however, will be omitted here.

IV. CORRELATION FUNCTIONS

We have derived a set of linear equations for the Wiener–Hermite functionals, from which we will obtain correlation functions of the Wiener–Hermite functionals in this section. However, we start with the average and the quadratic mean of the output sequence $\{Y_p\}$.

Since the Wiener–Hermite functional $y_n(p)=I_n[G_n(\cdot-p),\omega]$ has zero averages except when $n=0$, we obtain from (14)

$$\langle Y_p\rangle=\sum_{n=0}^{\infty}\langle y_{2n+1}(p)\rangle=0, \quad (34)$$

which is to be expected from the symmetry. Next we calculate the mean square of Y_p. Since $\langle [H_k(X_p)]^2 \rangle = k!$ and, for any given p, m, and k, $y_{2m-2k+1}(p-1)$ and $H_{2k}(X_p)$ are statistically independent from each other, taking the square of (33) and averaging the yield we find

$$\sigma_{2m+1}^2 = \sum_{k=0}^{m} \sigma_{2m-2k+1}^2 \times (2k)! \cdot d_{2k}^2 + (2m+1)! \cdot s_{2m+1}^2 \quad (m = 0,1,2,3,\ldots), \tag{35}$$

where σ_{2m+1}^2 is the quadratic mean of the $(2m+1)$-degree Wiener–Hermite functionals,

$$\sigma_{2m+1}^2 = \langle [y_{2m+1}(p)]^2 \rangle. \tag{36}$$

Taking a summation of (35), we obtain

$$\begin{aligned} \sum_{m=0}^{\infty} \sigma_{2m+1}^2 &= \sum_{m=0}^{\infty} \left\{ \sum_{k=0}^{m} \sigma_{2m-2k+1}^2 \times (2k)! \cdot d_{2k}^2 + (2m+1)! \cdot s_{2m+1}^2 \right\} \\ &= \left(\sum_{m=0}^{\infty} \sigma_{2m+1}^2 \right) \left(\sum_{k=0}^{\infty} (2k)! \cdot d_{2k}^2 \right) + \sum_{m=0}^{\infty} (2m+1)! \cdot s_{2m+1}^2. \end{aligned} \tag{37}$$

Using (22) and (37), we easily find

$$\langle |Y_p|^2 \rangle = \left\langle \left[\sum_{m=0}^{\infty} y_{2m+1}(p) \right]^2 \right\rangle = \sum_{m=0}^{\infty} \sigma_{2m+1}^2 = 1. \tag{38}$$

Let us obtain the correlation function R_{2m+1} of $y_{2m+1}(p)$. Multiplying both sides of (33) by $y_{2m+1}(q)$ and averaging, we obtain, for $q < p$,

$$\begin{aligned} R_{2m+1}(p-q) &= \langle y_{2m+1}(p) y_{2m+1}(q) \rangle \\ &= d_0 R_{2m+1}(p-q-1) \quad (q<p). \end{aligned} \tag{39}$$

Solving this with the initial condition $R_{2m+1}(0) = \sigma_{2m+1}^2$, we get

$$R_{2m+1}(p-q) = \sigma_{2m+1}^2 d_0^{|p-q|}, \tag{40}$$

which means that all Wiener–Hermite functionals have an exponential correlation function, with the common exponent $\ln(d_0)$. Making a summation of (40) and using (38), we obtain the autocorrelation function R_{YY} of Y_p,

$$R_{YY}(p-q) = \langle Y_p Y_q \rangle = \sum_{m=0}^{\infty} R_{2m+1}(p-q) = d_0^{|p-q|}. \tag{41}$$

From (41) we may conclude that the hysteresis transforms the independent Gaussian sequence $\{X_p\}$ into $\{Y_p\}$, a binary sequence with an exponential correlation. A continuous-time binary process taking ± 1 with an exponential correlation is known as the random telegraph signal.[4] Thus we may say that the hysteresis transforms an independent Gaussian sequence to the discrete-time random telegraph signal.

We calculate the cross-correlation $R_{YX}(p-q)$ of Y_p and X_q, which relates to the first-degree Wiener kernel. Multiplying both sides of (33) by $X_q = H_1(X_q)$, with $m = 0$ and taking the ensemble average, we get

$$\begin{aligned} R_{YX}(p-q) = G_1(q-p) &= \langle y_1(p) \cdot H_1(X_q) \rangle \\ &= d_0 R_{YX}(p-q-1) + s_1 \delta(p,q). \end{aligned} \tag{42}$$

where $G_1(q-p)$ is the first-degree Wiener kernel. Since Y_p and X_q are statistically independent from each other when $q > p$, we obtain from (42)

$$R_{YX}(p-q) = \begin{cases} 0 & (q>p), \\ s_1 \cdot d_0^{(p-q)} & (q \leqslant p), \end{cases} \tag{43}$$

which is a one-side exponential function. We note that (41) and (43) can be directly obtained from (10).

V. NUMERICAL RESULTS

Since $\langle |Y_p|^2 \rangle$ is finite, the Wiener–Hermite expansion (14) always converges in the ensemble mean-square sense. This does not mean that (14) converges fast. In fact, the convergence is expected to be slow because the hysteresis has a strong nonlinearity. Since no examples have been given definitely for the convergence property of the Wiener–Hermite expansion in the case of strong nonlinearity, it is interesting to check this point.

By $Y_p^{(N)}$ we denote an approximate output by a finite sum of Wiener–Hermite functionals,

$$Y_p^{(N)} = \sum_{m=0}^{N} y_{2m+1}(p). \tag{44}$$

Then the mean-square error $e^2(N)$ due to the approximation becomes

$$e^2(N) = \langle |Y_p - Y_p^{(N)}|^2 \rangle = 1 - \sum_{m=0}^{N} \sigma_{2m+1}^2. \tag{45}$$

In order to estimate $e^2(N)$, we first solve (35) for σ_{2m+1}^2 by the iteration starting with $m = 0$. Obtaining σ_{2m+1}^2 up to $m = 25$, we then calculate the mean-square error $e^2(N)$. A numerical example is shown in Fig. 4, in which $\log[e^2(N)]$ is plotted against $\log(N)$. Roughly speaking, the figure shows that $\log[e^2(N)]$ is asymptotically proportional to $-\frac{1}{2} \cdot \log(N)$. Thus we get a rough estimation,[20]

$$\begin{aligned} &e^2(N) \sim O(N^{-1/2}), \\ &[Y_p - Y_p^{(N)}] \sim O(N^{-1/4}), \quad N \to \infty, \end{aligned} \tag{46}$$

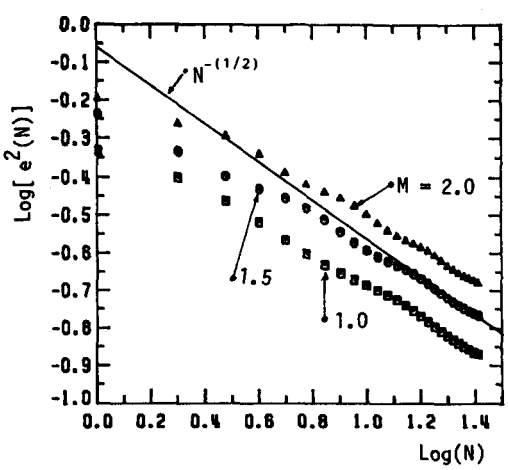


FIG. 4. Convergence property of the Wiener–Hermite expansion. The ensemble mean-square error is denoted by $e^2(N)$, and M is a threshold of the hysteresis.

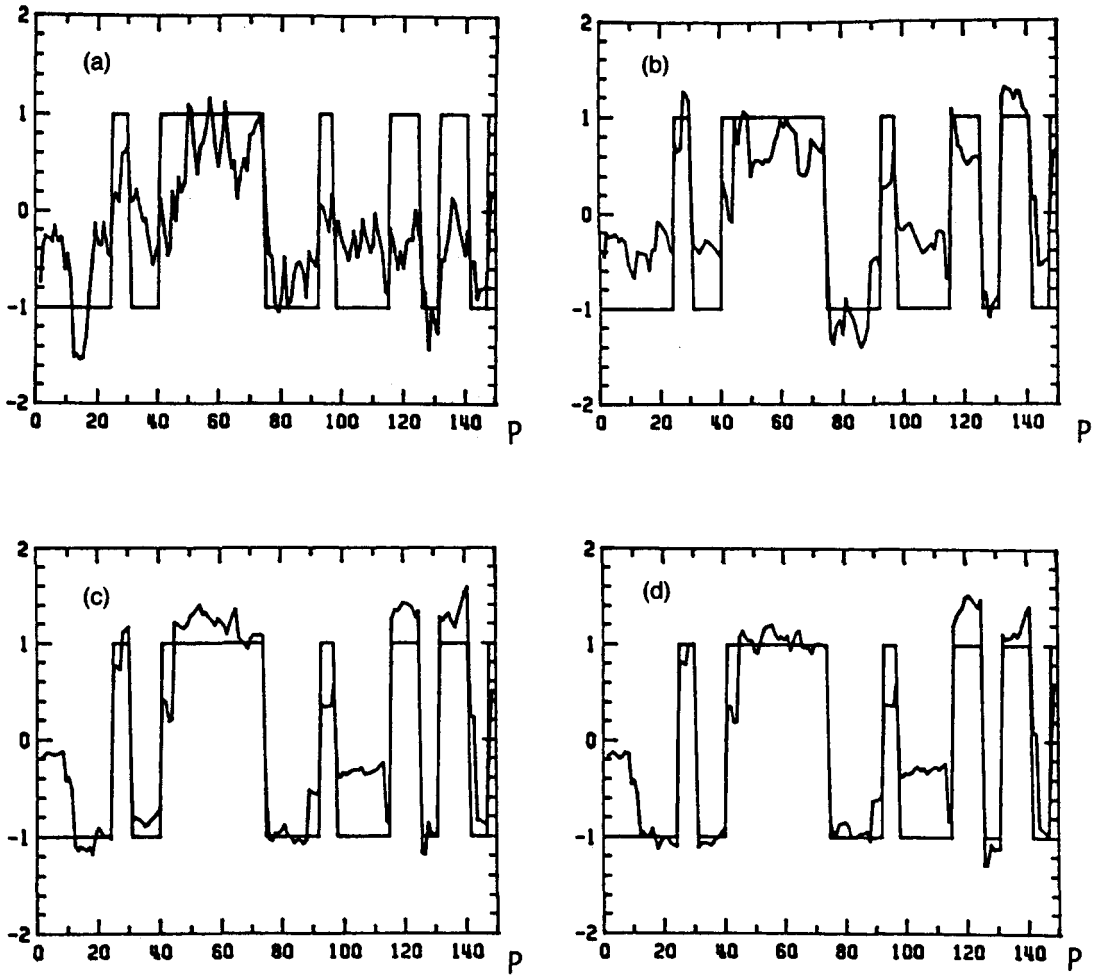


FIG. 5. Simulated output of the hysteresis (binary curve) and an approximate output (fluctuating curve) by a finite sum of Wiener–Hermite functionals for $M = 1.5$. (a) First-term approximation with $N = 0$; (b) four-term approximation with $N = 3$; (c) 12-term approximation with $N = 11$; and (d) 26-term approximation with $N = 25$.

which means that the convergence of the Wiener–Hermite expansion (14) is very slow in our case.

Next we describe a computer simulation. Using Gaussian random numbers $\{X_p\}$ generated in a computer, we solve (33) to get a sample sequence $y_{2m+1}(p)$, in terms of which an approximate output $Y_p^{(N)}$ by the Wiener–Hermite expansion may be computed. A simulated result so obtained is illustrated in Fig. 5. In the first term approximation with $N = 0$ the approximate output $Y_p^{(N)}$ fluctuates frequently against the discrete time p, and it is much different from Y_p, the true output. In the four-term approximation with $N = 3$ the approximation becomes better, and the switching part from 1 to -1, or from -1 to 1, in the true output Y_p is partially reconstructed by $Y_p^{(N)}$. In a higher-degree approximation with $N = 11$ or 25 the approximation becomes much better, as is shown in Figs. 5(c) and 5(d). But $Y_p^{(N)}$ is still far from the binary sequence Y_p. These computer results imply again that the Wiener–Hermite expansion converges slowly.

VI. CONCLUSIONS

We have demonstrated an exact Wiener analysis for a symmetrical binary hysteresis in the special case where the input is an independent, zero-mean, Gaussian sequence. We have derived a set of linear equations for the Wiener–Hermite functionals, from which Wiener–Hermite functionals could be directly simulated without knowledge of Wiener kernels. Also we have obtained an equivalent network for the hysteresis.

However, the Wiener analysis becomes complicated in the case of a correlated Gaussian sequence. This problem will be left for future study.

ACKNOWLEDGMENTS

The authors would like to thank S. Sato, Department of Biophysical Engineering, Osaka University, for suggesting this problem. Warm thanks go to Y. Yoshida and H. Ogura, Kyoto Institute of Technology, for their encouragement.

APPENDIX: BASIC NOTATIONS AND DEFINITIONS

1. Hermite polynomials of single variable

We define $H_n(x)$, Hermite polynomials of single variable, by

$$H_n(x) = [G(x)]^{-1}\cdot\left(-\frac{d}{dx}\right)^n\cdot G(x) \quad (n=0,1,2,\ldots),$$

$$G(x) = (1/\sqrt{2\pi})\exp(-\tfrac{1}{2}x^2), \tag{A1}$$

where $G(x)$ is the one-dimensional Gaussian distribution.

Examples:

$$\begin{aligned} &H_0(x) = 1,\\ &H_1(x) = x,\\ &H_2(x) = x^2 - 1,\\ &H_3(x) = x^3 - 3x,\\ &\vdots\,. \end{aligned} \tag{A2}$$

Orthogonality relation:

$$\begin{aligned}\langle H_n(x)\cdot H_m(x)\rangle &= \int_{-\infty}^{\infty} H_n(x)\cdot H_m(x)\cdot G(x)dx\\ &= n!\cdot\delta(n,m),\end{aligned} \tag{A3}$$

where $\delta(n,m)$ stands for Kronecker's delta, and the angular brackets denote the average with respect to $G(x)$.

2. Multivariate Hermite polynomials

We define the multivariate Hermite polynomial $H^{(n)}$ by

$$\begin{aligned}&H^{(n)}[X_{p_1},X_{p_2},\ldots,X_{p_n}]\\ &\quad= [G_{2k+1}(\mathbf{X})]^{-1}\prod_{\nu=1}^{n}\left(-\frac{\partial}{\partial X_{p_\nu}}\right)G_{2k+1}(\mathbf{X}),\end{aligned} \tag{A4}$$

Here $\mathbf{X} = (X_k,X_{k-1},X_{k-2},\ldots,X_0,\ldots,X_{-k})$ is a $(2k+1)$-dimensional vector and $G_{2k+1}(\mathbf{X})$ is the $(2k+1)$-dimensional Gaussian distribution,

$$G_{2k+1}(\mathbf{X}) = \left[\frac{1}{\sqrt{2\pi}}\right]^{(2k+1)}\cdot\exp\left(-\frac{1}{2}\sum_{p=-k}^{k}X_p^2\right), \tag{A5}$$

where k is taken to be arbitrarily large.

Examples of multivariate Hermite polynomials:

$$\begin{aligned} &H^{(0)}[X_p] = 1, \quad H^{(1)}[X_p] = X_p,\\ &H^{(2)}[X_p,X_q] = X_pX_q - \delta(p,q),\\ &H^{(3)}[X_p,X_q,X_r] = X_pX_qX_r - X_r\cdot\delta(p,q)\\ &\qquad\qquad - X_p\cdot\delta(r,q) - X_q\cdot\delta(r,p).\end{aligned} \tag{A6}$$

Orthogonality relation:

$$\begin{aligned}&\langle H^{(n)}[X_{p_1},X_{p_2},\ldots,X_{p_n}]\cdot H^{(m)}[X_{q_1},X_{q_2},\ldots,X_{q_m}]\rangle\\ &\quad= \delta(n,m)\delta_{pq}^{(m)},\end{aligned} \tag{A7}$$

where the angle brackets denote the average with respect to the multidimensional Gaussian distribution (A5) with an arbitrarily large k, and $\delta_{pq}^{(m)}$ stands for the sum of all distinct products of m Kronecker deltas of the form $\delta(p_\alpha,q_\beta)$, $p = (p_1,p_2,\ldots,p_n)$, $q = (q_1,q_2,\ldots,q_m)$, all p_α and q_β appearing just once in each product; for example,

$$\delta_{pq}^{(2)} = \delta(p_1,q_1)\cdot\delta(p_2,q_2) + \delta(p_2,q_1)\cdot\delta(p_1,q_2).$$

Even/odd symmetry:

$$\begin{aligned}&H^{(n)}[X_{p_1},X_{p_2},\ldots,X_{p_n}]\\ &\quad= (-1)^n\cdot H^{(n)}[-X_{p_1},-X_{p_2},\ldots,-X_{p_n}].\end{aligned} \tag{A8}$$

Factorization: The multivariate Hermite polynomial is reduced to a product of Hermite polynomials. For example,

$$H^{(3)}[X_p,X_q,X_r] = \begin{cases} H_3(X_p), & p=q=r,\\ H_2(X_p)\cdot H_1(X_r), & p=q\neq r,\\ H_1(X_p)\cdot H_1(X_q)\cdot H_1(X_r), & p\neq q,\quad q\neq r,\quad r\neq p,\\ H^{(2)}[X_p,X_q]\cdot H_1(X_r), & q\neq r,\quad p\neq r.\end{cases} \tag{A9}$$

3. The k-dimensional inner product

For the symmetrical functions $G_m(p_1,p_2,\ldots,p_m)$ and $D_n(p_1,p_2,\ldots,p_n)$ we define the k-dimensional inner product $\{G_m(\cdot-p),D_n(\cdot-q)\}_k$ by

$$\begin{aligned}&\{G_m(\cdot-p),D_n(\cdot-q)\}_k\\ &\quad= \sum_{p_1',p_2',\ldots,p_k'=-\infty}^{\infty} G_m(p_1'-p,p_2'-p,\ldots,p_k'-p,\\ &\qquad p_1-p,\ldots,p_{m-k}-p)D_n(p_1'-q,p_2'-q,\ldots,p_k'-q,\\ &\qquad p_{m-k+1}-q,\ldots,p_{m+n-2k}-q),\\ &\qquad k\leqslant\min(m,n),\end{aligned} \tag{A10}$$

which is an $m+n-2k+2$ variable function but asymmetrical in general. When $m=n=k$, $\{G_m(\cdot-p),D_m(\cdot-q)\}_m$ becomes a function of the difference $p-q$.

4. Wiener–Hermite expansion

Notations and definitions for the Wiener–Hermite expansion are described here. However, see Refs. 3–5 for mathematical details.

We consider a functional $g(T^p\omega)$ of an independent Gaussian sequence $\{X_p\}$, where ω is given by a sample sequence as (2) and $T^p\omega$ is another sample sequence obtained by the shift T. When a functional $g(T^p\omega)$ has a finite variance, i.e., $\langle|g(T^p\omega)|^2\rangle<\infty$, it has an orthogonal development called a Wiener–Hermite expansion, that is,

$$g(T^p\omega) = \sum_{n=0}^{\infty} I_n[G_n(\cdot-p),\omega], \tag{A11}$$

where $I_n[G_n(\cdot-p),\omega]$ is the nth degree Wiener–Hermite functional

$$\begin{aligned}&I_n[G_n(\cdot-p),\omega]\\ &\quad= \sum_{p_1,p_2,\ldots,p_n=-\infty}^{\infty} G_n(p_1-p,p_2-p,\ldots,p_n-p)\\ &\qquad\times H^{(n)}[X_{p_1},X_{p_2},\ldots,X_{p_n}] \quad (n=0,1,2,\ldots).\end{aligned} \tag{A12}$$

Here the dot in $I_n[G_n(\cdot-p),\omega]$ stands for the summation variables $(p_1,p_2,\ldots,p_n)$, $H^{(n)}$ is the multivariate Hermite polynomial defined by (A4), and $G_n(p_1,p_2,\ldots,p_n)$ is a deterministic function called the Wiener kernel, which is assumed to be symmetrical with respect to its arguments. (When $n=0$, however, $I_0[G_0(\cdot-p),\omega]$ is a constant equal to G_0.) The Wiener–Hermite functionals enjoy the following properties.

Shift transformation: By the definition (3) of the shift T^p and (A12), the Wiener–Hermite functionals satisfy

$$I_n[G_n(\cdot),T^p\omega] = I_n[G_n(\cdot-p),\omega] \quad (p=0,\pm1,\pm2,\ldots). \tag{A13}$$

The orthogonality relation:

$$\langle I_m[G_m(\cdot-p),\omega]\cdot I_n[D_n(\cdot-q),\omega]\rangle = \delta(n,m)\cdot m!\{G_m(\cdot-p),D_m(\cdot-q)\}_m, \tag{A14}$$

where $\{G_m(\cdot-p),D_m(\cdot-q)\}_m$ denotes the m-dimensional inner product defined in (A10). Putting $n=0$, we find that $I_m[G_m(\cdot-p),\omega]$ has zero averages except for $m=0$.

Even/odd symmetry: By (A8) and (A12), odd- (even-) degree Wiener–Hermite functionals are odd (even) in the sense of (4):

$$I_n[G_n(\cdot-p),\omega] = (-1)^n I_n[G_n(\cdot-p),-\omega] \quad (n=0,1,2,\ldots). \tag{A15}$$

Convergence in the ensemble mean square sense: The Wiener–Hermite expansion (A11) holds in the ensemble mean-square sense, namely,

$$\lim_{N\to\infty}\left\langle\left|g(T^p\omega)-\sum_{n=0}^{N} I_n[G_n(\cdot-p),\omega]\right|^2\right\rangle = 0. \tag{A16}$$

By (A7), (A14), and Parseval's relation, we get

$$\langle|g(T^p\omega)|^2\rangle = \sum_{m=0}^{\infty}\langle|I_m[G_m(\cdot-p),\omega]|^2\rangle = \sum_{m=0}^{\infty} m!\{G_m(\cdot-p),G_m(\cdot-p)\}_m. \tag{A17}$$

5. Wiener–Hermite expansion of a product

Let $g(T^p\omega)$ and $f(T^q\omega)$ be random functions described by Wiener–Hermite expansions:

$$g(T^p\omega) = \sum_{m=0}^{\infty} I_m[G_m(\cdot-p),\omega], \qquad f(T^q\omega) = \sum_{n=0}^{\infty} I_n[F_n(\cdot-q),\omega], \tag{A18}$$

where $G_m(\cdot-p)$ and $F_n(\cdot-q)$ are symmetrical kernels. Then the product of $g(T^p\omega)$ and $f(T^q\omega)$ is given by another Wiener–Hermite expansion as

$$g(T^p\omega)\times f(T^q\omega) = \sum_{m=0}^{\infty}\sum_{n=0}^{\infty}\sum_{k=0}^{m}\frac{(n+m-k)!\,(n+k)!}{(m-k)!\,k!\,n!} I_m \times[\{G_{n+m-k}(\cdot-p),F_{n+k}(\cdot-q)\}_n,\omega]. \tag{A19}$$

This relation is a slightly modified version of Theorem 2 in Ref. 17.

[1] J. E. Gibson, *Nonlinear Automatic Control* (McGraw-Hill, New York, 1963).
[2] P. Z. Marmarelis and V. Z. Marmarelis, *Analysis of Physiological Systems* (Plenum, New York, 1978).
[3] N. Wiener, *Nonlinear Problems in Random Theory* (MIT, Cambridge, MA, 1958).
[4] H. Ogura, *Stochastic Processes* (Corona, Tokyo, 1978) (in Japanese).
[5] W. J. Rugh, *Nonlinear System Theory* (Johns Hopkins U.P., Baltimore, MD, 1981).
[6] A. Siegel, T. Imamura, and W. C. Meecham, "Wiener–Hermite expansion in model turbulence in the late decay stage," J. Math. Phys. **6**, 707 (1965).
[7] H. D. Hogge and W. C. Meecham, "The Wiener–Hermite expansion applied to decaying isotropic turbulence using a renormalized time-dependent base," J. Fluid Mech. **85**, 325 (1978).
[8] H. Ogura, "Theory of waves in a homogeneous random medium," Phys. Rev. A **11**, 942 (1975).
[9] J. Nakayama, H. Ogura, and M. Sakata, "Scattering of a scalar wave from a slightly random surface," J. Math. Phys. **22**, 471 (1981).
[10] J. Nakayama, K. Mitzutani, and M. Tsuneoka, "Scattering of electromagnetic waves from a perfectly conductive slightly random surface: Depolarization in backscatter," J. Math. Phys. **27**, 1436 (1986).
[11] J. F. Barrett, "Hermite functional expansion and the calculation of output auto-correlation and spectrum for any time-invariant nonlinear system with noise input," J. Electron. Control **16**, 107 (1963).
[12] S. Yasui, "Stochastic functional Fourier series, Volterra series and nonlinear system analysis," IEEE Trans. Automat. Control. **24**, 230 (1979).
[13] M. Schtzen, "Nonlinear system modeling based on the Wiener theory," Proc. IEEE **69**, 1557 (1981).
[14] S. Yasui, "Wiener-like Fourier kernels for nonlinear system identification and synthesis (Nonanalytic cascade, bilinear, and feedback cases)," IEEE Trans. Automat. Control. **AC-27**, 677 (1982).
[15] E. Isobe and S. Sato, "An integro-differential formula on the Wiener kernels and its application to sandwich system identification," IEEE Trans. Automat. Control **AC-29**, 595 (1984).
[16] E. Isobe and S. Sato, "Wiener kernels of a system described by Ito stochastic differential equation," Inst. Elect. Commun. Eng. Jpn. Trans. **J66A**, 374 (1983).
[17] J. Nakayama, "A formula on the Wiener–Hermite expansion," J. Math. Phys. **27**, 2559 (1987).
[18] J. L. Doob, *Stochastic Processes* (Wiley, New York, 1953).

[19] Equation (10) gives another orthogonal development, because a product of any two different terms in the right hand vanishes by the relation $D(x)S(x)=0$ in (9).
[20] Numerical overflow takes place in a FORTRAN routine for factorials in (21). This makes it difficult to calculate $y_{2m+1}(p)$ when $(2m+1)>51$.

Long wave equations and metrizable Lie algebras

B. A. Kupershmidt
The University of Tennessee Space Institute, Tullahoma, Tennessee 37388

(Received 17 September 1987; accepted for publication 18 May 1988)

An infinite commuting hierarchy of three-Hamiltonian integrable extensions of the dispersive water waves (DWW) hierarchy is constructed for each finite-dimensional Lie algebra $\mathscr{G}$ with an invariant symmetric bilinear form on it. The construction is based on special two-cocycles on a one-parameter family of differential Lie algebras attached to $\mathscr{G}$.

I. INTRODUCTION

The classical equations of one-dimensional dispersiveless free surface long waves

$$\begin{aligned} &u_t + uu_x + h_x = 0, \\ &h_t + (uh)_x = 0, \quad u = u(x,t), \quad h = h(x,t) \end{aligned} \tag{1.1}$$

lie at the intersection of two of the richest, but still not well understood, integrable systems known to date. The first of these is also a dispersiveless two-dimensional analog of (1.1), called the Benney system,[1]

$$\begin{aligned} &u_t + uu_x + h_x - u_y \int_0^y u_x\, dy = 0, \\ &h_t + \left(\int_0^h u\, dy\right)_x = 0, \quad u = u(x,y,t), \quad h = h(x,t), \end{aligned} \tag{1.2}$$

which, upon introducing the moments $A_m := \int_0^h u^m\, dy$, results in the separate system of evolution equations for the A_m's,

$$A_{m,t} + A_{m+1,x} + mA_{m-1}A_{0,x} = 0. \tag{1.3}$$

The system (1.3) is integrable, i.e., it possesses an infinite number of conserved densities[1] and an infinite number of commuting higher flows resulting from the first Hamiltonian structure of (1.3).[2,3] In addition, the system (1.2) is itself separate Hamiltonian and integrable, and the system (1.3) has a second Hamiltonian structure.[4] The second of these integrable systems is also one-dimensional but *dispersive* (i.e., with higher x derivatives) analog of (1.1), called the Broer–Kaup (BK) system[5,6]

$$\begin{aligned} &u_t + uu_x + h_x - \tfrac{1}{2}u_{xx} = 0, \\ &h_t + (uh)_x + \tfrac{1}{2}h_{xx} = 0, \quad u = u(x,t), \quad h = h(x,t). \end{aligned} \tag{1.4}$$

The system (1.4) is also integrable: it has an infinite number of conserved densities[6,7] and higher commuting flows,[7] three Hamiltonian structures,[7] and two different Lax representations.[6,7] The classical Korteweg–de Vries (KdV) hierarchy can be represented as a factor of the BK hierarchy, and not only in the differential equations sense, but also in the Lax and Hamiltonian senses.[7]

Although a dispersive generalization of the system (1.3) is known, being just the famous Kadomtsev–Petviashvili (KP) hierarchy,[4] it has not yet been found for the truly two-dimensional Benny system (1.2); were such a generalization to become known it could be rightfully called *Integrabilitorum princeps.*

The purpose of this paper is to construct, for *arbitrary metrizable Lie algebra* $\mathscr{g}$, an infinite commuting hierarchy of integrable three-Hamiltonian extensions of the BK hierarchy, the latter corresponding to the case of $\mathscr{g} = \{0\}$. Recall that a finite-dimensional Lie algebra $\mathscr{g}$ over a commutative ring $\tilde{\mathscr{k}}$ is called metrizable if there exists a nondegenerate symmetric bilinear invariant form on it. (For example, any complex semisimple Lie algebra is metrizable.) We fix $\mathscr{g}$ and assume that $\tilde{\mathscr{k}} \supset \mathbb{Q}$, and that there exists a basis $(e_1,...,e_N)$ of $\mathscr{g}$ in which the structure constants $t^{\sigma}_{\mu\nu}$ of $\mathscr{g}$ are cyclic symmetric,

$$t^{\sigma}_{\mu\nu} = t^{\mu}_{\nu\sigma}, \quad 1 \leqslant \mu,\nu,\sigma \leqslant N = \dim \mathscr{g}; \tag{1.5}$$

e.g., an orthonormal basis for $\mathscr{g}$ will do. We fix this basis of $\mathscr{g}$.

I now set up the notation, state the main results, and describe the layout of the paper. The basic facts about differential rings and the Hamiltonian formalism are taken for granted; they can be found, e.g., in Refs. 8 (Chap. I) and 9 (Chaps. VII and VIII). (I will recall most of the needed facts in the main body of the paper.)

Let K be a commutative differential ring with a derivation $\partial: K \to K$, such that $\tilde{\mathscr{k}} \subset \mathscr{k} := \operatorname{Ker} \partial|_K$. Let $C = K[u^{(j)}, h^{(j)}, r_\mu^{(j)}, s_\mu^{(j)}]$, $j \in \mathbb{Z}_+$, $1 \leqslant \mu \leqslant N$, be a differential algebra. Consider the following system of evolution equations in C:

$$\begin{aligned} &u_t = (u^2 + 2h - u_x)_x, \quad h_t = (2uh + h_x + \mathbf{r}'\mathbf{s})_x, \\ &\mathbf{r}_t = -[\mathbf{r},\mathbf{s}] + (u\mathbf{r})_x, \quad \mathbf{s}_t = (u\mathbf{s} + 2\mathbf{r})_x, \end{aligned} \tag{1.6}$$

where $\mathbf{r} = (r_1,...,r_N)^t$, $\mathbf{s} = (s_1,...,s_N)^t$ are $\mathscr{g}$-valued functions of x and t, and the commutator $[\mathbf{r},\mathbf{s}]$ is taken in $\mathscr{g}$. The system (1.6) is a $\mathscr{g}$-generalization of the BK system (1.4) to which it collapses when $\mathbf{r} = \mathbf{s} = \mathbf{0}$ (and also for $\mathscr{g} = \{0\}$), provided one changes the fluid dynamical time t in (1.4) into $-t/2$ (to make forthcoming formulas more aesthetically appealing).

Set

$$H_0 = u/2, \tag{1.7}$$

$$H_1 = h, \tag{1.8}$$

$$H_2 = uh + \mathbf{r}'\mathbf{s}, \tag{1.9}$$

$$H_3 = h^2 + h(u^2 - u^{(1)}) + u\mathbf{r}'\mathbf{s} + \mathbf{r}'\mathbf{r}, \tag{1.10}$$

and define the following matrix differential operators:

 0022-2488/88/091990-11$02.50

$$B^1=\left(\begin{array}{cc|cc} 0 & \partial & & \\ \partial & 0 & \multicolumn{2}{c}{0} \\ \hline & & \tfrac{1}{2}[\mathbf{s},\,] & \partial\mathbf{1} \\ \multicolumn{2}{c|}{0} & \partial\mathbf{1} & 0 \end{array}\right), \tag{1.11}$$

$$B^2=\left(\begin{array}{c|c|c|c} 2\partial & \partial(u-\partial) & 0 & 0 \\ \hline (u+\partial)\partial & h\partial+\partial h & \mathbf{r}'\partial & \mathbf{s}'\partial \\ \hline 0 & \partial\mathbf{r} & -[\mathbf{r},\,] & 0 \\ \hline 0 & \partial\mathbf{s} & 0 & 2\partial\mathbf{1} \end{array}\right), \tag{1.12}$$

$$B^3=\left(\begin{array}{c|c|c|c} 2(u\partial+\partial u) & \partial(u-\partial)^2+2(h\partial+\partial h) & 2\mathbf{r}'\partial & 2\mathbf{s}'\partial \\ \hline (u+\partial)^2\partial & (u+\partial)(h\partial+\partial h) & (u+\partial)\mathbf{r}'\partial & (u+\partial)\mathbf{s}'\partial \\ +2(h\partial+\partial h) & +(h\partial+\partial h)(u-\partial) & -[\mathbf{s},\mathbf{r}]' & +2\mathbf{r}'\partial \\ & +\mathbf{r}'\mathbf{s}\partial+\partial\mathbf{r}'\mathbf{s} & & \\ \hline 2\partial\mathbf{r} & \partial\mathbf{r}(u-\partial)-[\mathbf{r},\mathbf{s}] & 0 & -2[\mathbf{r},\,] \\ \hline 2\partial\mathbf{s} & \partial\mathbf{s}(u-\partial)+2\partial\mathbf{r} & -2[\mathbf{r},\,] & -2[\mathbf{s},\,] \end{array}\right). \tag{1.13}$$

It is easy to see that the $\mathcal{G}$–BK system (1.6) can be written in the form

$$\begin{pmatrix} u \\ h \\ \mathbf{r} \\ \mathbf{s} \end{pmatrix}_t = B^1\delta(H_3)=B^2\delta(H_2)=B^3\delta(H_1), \tag{1.14}$$

where $\delta(H)$ is the column vector of functional derivatives of an element $H\in C$.

Theorem 1.1: (i) The matrices B^1 and B^2 are Hamiltonian; (ii) the matrix B^3 is Hamiltonian.

Since the matrices B^1 and B^2 are affine, the (i) part is equivalent to constructing the corresponding one-parameter family of differential Lie algebras, say $\mathcal{G}(\lambda)$ [formula (2.3)], and a few (more exactly, five) generalized two-cocycles on $\mathcal{G}(\lambda)$ [formulas (2.7)–(2.11)]. This will be done in Sec. II. Since the matrix B^3 is badly nonlinear, the direct check of its Hamiltonian property would be a modern version of constructing a regular 65537-gon with a ruler and a compass. We circumvent this problem in Sec. VI by *linearizing* the matrix B^3, i.e., by exhibiting a nonlinear differential change of variables that transforms B^3 into an affine matrix; after that, purely Lie-algebraic arguments will suffice to prove the (ii) part.

Denote

$$C'=\mathbb{Q}[u^{(j)},h^{(j)},r_\mu^{(j)},s_\mu^{(j)}],\quad j\in\mathbb{Z}_+,\quad 1\leqslant\mu\leqslant N.$$

Theorem 1.2: There exists an infinite sequence of elements $H_m\in C'$, $m\in\mathbb{Z}_+$, starting with (1.7)–(1.10), such that

$$B^1\delta(H_{m+1})=B^2\delta(H_m),\quad m\in\mathbb{Z}_+, \tag{1.15a}$$

$$B^2\delta(H_{m+1})=B^3\delta(H_m),\quad m\in\mathbb{Z}_+. \tag{1.15b}$$

In other words, the $\mathcal{G}$–BK system (1.14) is just one member of an infinite commuting three-Hamiltonian hierarchy of integrable systems. [It is quite likely that the coefficients of H_{m+1}'s (when the structure constants $t_{\mu\nu}^{\sigma}$ are treated as formal variables; see Sec. III below) belong to $\mathbb{Z}$ and not to $\mathbb{Q}$, but my methods are insufficient to prove this.] Equation (1.15b) is obvious for $m=0$ and follows from (1.15a) for $m\in\mathbb{N}$ by a straightforward computation performed in Sec. III. Equation (1.15a) presents, as is usual in bi-Hamiltonian analysis (outside classical mechanics where B^1 is invertible), serious difficulties. They are resolved in two steps. First, in Sec. IV we consider a new one-parameter family of differential Lie algebras $\mathcal{L}(\lambda)$ [formula (4.1)] and three two-cocycles on $\mathcal{L}(\lambda)$ [formulas (4.3)–(4.5)]. Out of this data, over the differential ring $C_1=K[u^{(j)},h^{(j)},q^{(j)},p^{(j)}]$, $j\in\mathbb{Z}_+$, we extract the following two matrix differential operators:

$$\overline{B}^1=\left(\begin{array}{cc|cc} 0 & \partial & & \\ \partial & 0 & \multicolumn{2}{c}{0} \\ \hline & & 0 & q\partial+\partial q \\ \multicolumn{2}{c|}{0} & q\partial+\partial q & p\partial+\partial p \end{array}\right), \tag{1.16}$$

$$\overline{B}^2=\left(\begin{array}{c|c|c|c} 2\partial & \partial(u-\partial) & 0 & 0 \\ \hline (u+\partial)\partial & h\partial+\partial h & q\partial+\partial q & p\partial+\partial p \\ \hline 0 & q\partial+\partial q & 0 & 0 \\ \hline 0 & p\partial+\partial p & 0 & q\partial+\partial q \end{array}\right). \tag{1.17}$$

Later on, from the matrices $\overline{B}^1$ and $\overline{B}^2$, we derive the following matrix:

$$\overline{B}^3 = \begin{pmatrix} 2(u\partial+\partial u) & \partial(u-\partial)^2 + 2(h\partial+\partial h) & 2(q\partial+\partial q) & 2(p\partial+\partial p) \\ (u+\partial)^2\partial + 2(h\partial+\partial h) & (u+\partial)(h\partial+\partial h) + (h\partial+\partial h)(u-\partial) + (p\partial+\partial p) & (u+\partial)(q\partial+\partial q) & (u+\partial)(p\partial+\partial p) + (q\partial+\partial q) \\ 2(q\partial+\partial q) & (q\partial+\partial q)(u-\partial) & 0 & 0 \\ 2(p\partial+\partial p) & (p\partial+\partial p)(u-\partial) + (q\partial+\partial q) & 0 & 0 \end{pmatrix}. \tag{1.18}$$

Set

$$\overline{H}_0 = u/2, \tag{1.19}$$

$$\overline{H}_1 = h, \tag{1.20}$$

$$\overline{H}_2 = uh + p, \tag{1.21}$$

$$\overline{H}_3 = h^2 + h(u^2 - u^{(1)}) + up + q, \tag{1.22}$$

and introduce the following grading into C_1:

$$\begin{aligned} &r\!k(u^{(j)}) = 1+j, \quad r\!k(h^{(j)}) = 2+j, \\ &r\!k(q^{(j)}) = 4+j, \quad r\!k(p^{(j)}) = 3+j, \quad r\!k(K) = 0. \end{aligned} \tag{1.23}$$

Theorem 1.3: There exists a unique (modulo Im ∂) sequence of homogeneous elements $\overline{H}_m \in C_1' = \mathbb{Q}[u^{(j)}, h^{(j)}, q^{(j)}, p^{(j)}]$, $j \in \mathbb{Z}_+$, $m \in \mathbb{Z}_+$, $rk(\overline{H}_m) = m+1$, starting with (1.21)–(1.24), such that

$$\overline{B}^1\delta(\overline{H}_{m+1}) = \overline{B}^2\delta(\overline{H}_m), \quad m \in \mathbb{Z}_+, \tag{1.24}$$

$$\overline{B}^2\delta(\overline{H}_{m+1}) = \overline{B}^3\delta(\overline{H}_m), \quad m \in \mathbb{Z}_+. \tag{1.25}$$

This will be proven in Sec. IV. Second, in Sec. V we consider a differential homomorphism (i.e., over K and commuting with ∂) $\Phi: C_1 \to C$, given on the generators of C_1 by the formulas

$$\Phi(u) = u, \quad \Phi(h) = h, \quad \Phi(q) = \mathbf{r}'\mathbf{r}, \quad \Phi(p) = \mathbf{r}'\mathbf{s}. \tag{1.26}$$

Lemma 1.1: Each of the pair of matrices $(\overline{B}^l, B^l)$, $l = 1,2,3$, is Hamiltonically related with respect to the homomorphism Φ, i.e. (see Ref. 10, Proposition 6.1),

$$D(\Phi)B^l D(\Phi)^\dagger = \Phi(\overline{B}^l), \tag{1.27.l}$$

$l = 1,2,3$, where Φ is the vector $(\Phi(u), \Phi(h), \Phi(q), \Phi(p))^t$, $D(\Phi)$ is the Fréchet derivative of Φ, and the dagger stands for "adjoint."

To check (1.27) is a simple but long exercise that is left to the reader. Since Φ is injective, from the Hamiltonian character of B^3 (proved in Sec. VI) and (1.27.3), it follows that $\overline{B}^3$ is also Hamiltonian. Since $\overline{B}^1$ and $\overline{B}^2$ are Hamiltonian by construction [being associated with the Lie algebra $\mathscr{L}(\lambda)$], (1.24) and (1.25) describe an infinite commuting hierarchy of integrable three-Hamiltonian systems. For the Hamiltonian $\overline{H}_3$, $\overline{B}^1\delta(\overline{H}_3)$ represents the system

$$\begin{aligned} &u_t = (u^2 + 2h - u_x)_x, \quad h_t = (2uh + h_x + p)_x, \\ &q_t = uq_x + 2u_x q, \quad p_t = up_x + 2u_x p + q_x, \end{aligned} \tag{1.28}$$

which is another generalization of the BK system (1.4).

Theorem 1.4: Set $H_m = \Phi(\overline{H}_m) \in C'$, $m \in \mathbb{Z}_+$. Then the thus defined sequence $\{H_m\}$ satisfies (1.15a).

This proves the (i) part of Theorem 1.2. Theorem 1.4 is proved in Sec. V.

II. LIE ALGEBRAS $\mathscr{g}(\lambda)$ AND ASSOCIATED HAMILTONIAN MATRICES

Denote by $\mathscr{D} = \mathscr{D}(K)$ the following Lie algebra structure on K:

$$[X_1, X_2] = X_1X_2^{(1)} - X_1^{(1)}X_2, \quad X_1, X_2 \in K, \tag{2.1}$$

where $(\cdot)^{(l)} = \partial^l(\cdot)$. Denote by V_c, $c \in \mathscr{k}$, the following $\mathscr{D}$-module structure on K:

$$X: f \mapsto X_f^{(1)} + cX^{(1)}f, \quad X \in \mathscr{D}, \quad f \in K. \tag{2.2}$$

The Lie algebra $\mathscr{g}(\lambda)$ is a free $(2N+2)$-dimensional K-module $K \oplus K \oplus (K \otimes \mathscr{g}) \oplus (K \otimes \mathscr{g})$ (all tensor products are over $\mathscr{k}$) with the commutator

$$\left[\begin{pmatrix} \varphi_1 \\ X_1 \\ f_1 \otimes a_1 \\ g_1 \otimes b_1 \end{pmatrix}, \begin{pmatrix} \varphi_2 \\ X_2 \\ f_2 \otimes a_2 \\ g_2 \otimes b_2 \end{pmatrix}\right] = \begin{pmatrix} X_1\varphi_2^{(1)} - X_2\varphi_1^{(1)} \\ X_1X_2^{(1)} - X_1^{(1)}X_2 \\ X_1f_2^{(1)} \otimes a_2 - X_2f_1^{(1)} \otimes a_1 + f_1f_2 \otimes [a_1, a_2] \\ X_1g_2^{(1)} \otimes b_2 - X_2g_1^{(1)} \otimes b_1 - (\lambda/2)f_1f_2 \otimes [a_1, a_2] \end{pmatrix}, \quad \varphi_l, X_l, f_l, g_l \in K, \quad a_l, b_l \in \mathscr{g}, \quad l = 1,2. \tag{2.3}$$

In other words, $\mathscr{g}(\lambda)$ is semidirect product Lie algebra $\mathscr{D} \leftthreetimes [V_0 \oplus (V_0 \otimes \tilde{\mathscr{g}}(\lambda))]$, where $\tilde{\mathscr{g}}(\lambda)$ is the following Lie algebra structure on $\mathscr{g} \oplus \mathscr{g}$:

$$\left[\begin{pmatrix} a_1 \\ b_1 \end{pmatrix}, \begin{pmatrix} a_2 \\ b_2 \end{pmatrix}\right] = \begin{pmatrix} [a_1,a_2] \\ -(\lambda/2)[a_1,a_2] \end{pmatrix}, \quad a_l, b_l \in \mathscr{g}, \quad l = 1,2; \tag{2.4}$$

here and below, λ is a formal parameter commuting with everything.

Recall (Ref. 9, Chap. VIII) that if $\mathscr{N} = K^n$ has a differential Lie algebra structure $[\ ,\ ]$, then a bilinear differential operator $\omega: \mathscr{N} \times \mathscr{N} \to K$ is called a (generalized) two-cocycle on $\mathscr{N}$ if

$$\omega(X,Y) \sim -\omega(Y,X), \quad \forall X,Y \in \mathscr{N}, \tag{2.5}$$

$$\omega([X,Y],Z) + \text{c.p.} \sim 0, \quad \forall X,Y,Z \in \mathscr{N}, \tag{2.6}$$

where $f \sim g$ means $(f-g) \in \operatorname{Im} \partial$, and c.p. stands for "cyclic permutation."

Lemma 2.1: The following are two-cocycles on $\mathscr{g}(\lambda)$:

$$\omega_1(1,2) = \varphi_1 X_2^{(1)} + X_1 \varphi_2^{(1)}, \tag{2.7}$$

$$\omega_2(1,2) = -\varphi_1 X_2^{(2)} + X_1 \varphi_2^{(2)}, \tag{2.8}$$

$$\omega_3(1,2) = \varphi_1 \varphi_2^{(1)}, \tag{2.9}$$

$$\omega_4(1,2) = (a_1,a_2) f_1 f_2^{(1)}, \tag{2.10}$$

$$\omega_5(1,2) = 2(b_1,b_2) g_1 g_2^{(1)} + \lambda(a_1,b_2) f_1 g_2^{(1)} + \lambda(b_1,a_2) g_1 f_2^{(1)}, \tag{2.11}$$

where $\omega(1,2)$ is a shorthand notation for

$$\omega\left(\begin{pmatrix} \varphi_1 \\ \vdots \end{pmatrix}, \begin{pmatrix} \varphi_2 \\ \vdots \end{pmatrix}\right),$$

and $(\ ,\ )$ is the invariant scalar product on $\mathscr{g}$.

Proof: A direct verification omitted due to its length and triviality. ■

Recall (Ref. 9, Chap. VIII) that to the Lie algebra $\mathscr{N}$ one associates the $n \times n$ matrix differential operator $B' = B(\mathscr{N})$, with coefficients in a differential ring $K[v_\alpha^{(j)}]$, $j \in \mathbb{Z}_+$, $1 \leqslant \alpha \leqslant n$, via the formula

$$X' B'(Y) \sim \sum_{\alpha=1}^{n} v_\alpha [X,Y]_\alpha, \quad \forall X,Y \in \mathscr{N}, \tag{2.12}$$

where $[X,Y]_\alpha$ is the component $\#\alpha$ of the commutator $[X,Y]$ in $\mathscr{N}$. Therefore, from (2.3) and (2.12) we get

$$\begin{aligned} & u(X_1 \varphi_2^{(1)} - X_2 \varphi_1^{(1)}) + h(X_1 X_2^{(1)} - X_1^{(1)} X_2) \\ & \quad + \Sigma r_\mu (X_1 f_{2\mu}^{(1)} - X_2 f_{1\mu}^{(1)} + \Sigma t_{\sigma\nu}^{\mu} f_{1\alpha} f_{2\nu}) \\ & \quad + \Sigma s_\mu (X_1 g_{2\mu}^{(1)} - X_2 g_{1\mu}^{(1)} - (\lambda/2) \Sigma t_{\sigma\nu}^{\mu} f_{1\alpha} f_{2\nu}) \\ & \sim \varphi_1 [\partial u(X_2)] + X_1 [u \partial(\varphi_2) \\ & \quad + (h\partial + \partial h)(X_2) + \Sigma r_\mu \partial(f_{2\mu}) + \Sigma s_\mu \partial(g_{2\mu})] \\ & \quad + f_{1\nu} [\partial r_\nu(X_2) + \Sigma t_{\nu\mu}^{\sigma} (r_\sigma - (\lambda/2) s_\sigma)(f_{2\mu})] \\ & \quad + g_{1\nu} [\partial s_\nu(X_2)]. \end{aligned} \tag{2.13}$$

Thus the natural Hamiltonian matrix $B(\mathscr{g}(\lambda))$ associated to the Lie algebra $\mathscr{g}(\lambda)$ is

$$B(\mathscr{g}(\lambda)) = \begin{pmatrix} 0 & \partial u & 0 & 0 \\ u\partial & h\partial + \partial h & \mathbf{r}'\partial & \mathbf{s}'\partial \\ 0 & \partial \mathbf{r} & [-\mathbf{r} + (\lambda/2)\mathbf{s},\ ] & 0 \\ 0 & \partial \mathbf{s} & 0 & 0 \end{pmatrix}. \tag{2.14}$$

For a two-cocycle ω on $\mathscr{N}$, let $\beta = \beta_\omega$ be the corresponding matrix differential operator

$$X' \beta(Y) \sim \omega(X,Y), \quad \forall X,Y \in \mathscr{N}. \tag{2.15}$$

Let β_l be the matrix associated to the two-cocycle ω_l [(2.6) + l]. Define

$$B(\lambda) := B(\mathscr{g}(\lambda)) + \lambda\beta_1 + \beta_2 + 2\beta_3 + \beta_5, \tag{2.16}$$

where $\beta_5 = \beta_5(\lambda)$ depends upon λ [see (2.11)],

$$B(\lambda) = \begin{pmatrix} 2\partial & \partial(u-\partial) + \lambda\partial & 0 & 0 \\ (u+\partial)\partial + \lambda\partial & h\partial + \partial h & \mathbf{r}'\partial & \mathbf{s}'\partial \\ 0 & \partial \mathbf{r} & [-\mathbf{r} + (\lambda/2)\mathbf{s},\ ] & \lambda\partial \mathbf{1} \\ 0 & \partial \mathbf{s} & \lambda\partial \mathbf{1} & 2\partial \mathbf{1} \end{pmatrix}. \tag{2.17}$$

Since the ω_l's are two-cocycles on $\mathscr{g}(\lambda)$, the matrix $B(\lambda)$ (2.16) is Hamiltonian (Ref. 9, Chap. VIII). In particular, the following matrices are Hamiltonian:

$$B^1 := \lambda^{-1} B(\lambda)|_{1/\lambda = 0}, \tag{2.18}$$

$$B^2 := B(\lambda)|_{\lambda = 0}, \tag{2.19}$$

which are precisely the matrices (1.11) and (1.12). This proves Theorem 1.1(i). ■

III. DERIVATION OF THE MATRIX B^3

Writing Eq. (1.15a) in longhand, with B^1 and B^2 given by (1.11) and (1.12), we obtain

$$\partial\left(\frac{\delta H_{m+1}}{\delta h}\right) = \partial\left[2\frac{\delta H_m}{\delta u} + (u-\partial)\left(\frac{\delta H_m}{\delta h}\right)\right], \tag{3.1u}$$

$$\partial\left(\frac{\delta H_{m+1}}{\delta u}\right) = (u+\partial)\partial\left(\frac{\delta H_m}{\delta u}\right) + (h\partial + \partial h)\left(\frac{\delta H_m}{\delta h}\right) + \mathbf{r}'\partial\left(\frac{\delta H_m}{\delta \mathbf{r}}\right) + \mathbf{s}'\partial\left(\frac{\delta H_m}{\delta \mathbf{s}}\right), \tag{3.1h}$$

$$\frac{1}{2}\left[\mathbf{s},\frac{\delta H_{m+1}}{\delta \mathbf{r}}\right]+\partial\left(\frac{\delta H_{m+1}}{\delta \mathbf{s}}\right)$$
$$=\partial \mathbf{r}\left(\frac{\delta H_m}{\delta h}\right)-\left[\mathbf{r},\frac{\delta H_m}{\delta \mathbf{r}}\right], \tag{3.1r}$$
$$\partial\left(\frac{\delta H_{m+1}}{\delta \mathbf{r}}\right)=\partial\left(\mathbf{s}\frac{\delta H_m}{\delta h}+2\frac{\delta H_m}{\delta \mathbf{s}}\right). \tag{3.1s}$$

We would like to conclude from (3.1u) and (3.1s) that

$$\frac{\delta H_{m+1}}{\delta h}=2\frac{\delta H_m}{\delta u}+(u-\partial)\left(\frac{\delta H_m}{\delta h}\right), \tag{3.2}$$
$$\frac{\delta H_{m+1}}{\delta \mathbf{r}}=\mathbf{s}\frac{\delta H_m}{\delta h}+2\frac{\delta H_m}{\delta \mathbf{s}}, \tag{3.3}$$

but we cannot do that unless we take care of the elements in $\operatorname{Ker}\partial|_{C=}\not\subset$. (Exercise: Show that $\operatorname{Ker}\partial$ in $K[v_\alpha^{(j)}]$ is the same as $\operatorname{Ker}\partial$ in K.) To exclude elements of $\not\subset$ from $\delta(H)$ we can follow the usual route by introducing a grading in C in which H_m's will become homogeneous. Such a grading must be consistent with Eqs. (3.1u)–(3.1s), which imply, upon choosing the units where $r\not\subset(\partial)=1$ and $r\not\subset(K)=0$:

$$r\not\subset(u^{(j)})=1+j,\quad r\not\subset(h^{(j)})=2+j,$$
$$r\not\subset(s_\mu^{(j)})=1+j,\quad r\not\subset(r_\mu^{(j)})=2+j, \tag{3.4}$$

$$r\not\subset(H_{m+1})=r\not\subset(H_m)+1, \tag{3.5}$$

provided

$$r\not\subset([\,,\,])=1. \tag{3.6}$$

This last equality forces us to have $r\not\subset(t_{\mu\nu}^\sigma)=1$, which is inconsistent with $r\not\subset(K)=0$ since $\{t_{\mu\nu}^\sigma\}\subset\tilde{\not\subset}\subset\not\subset\subset K$. To avoid a contradiction we need to enlarge our ring C into

$$\overline{C}=K[u^{(j)},h^{(j)},r_\mu^{(j)},s_\mu^{(j)}][t_{\mu\nu}^\sigma], \tag{3.7}$$

where $\{t_{\mu\nu}^\sigma\}$ are treated as formal variables satisfying the cyclic symmetry condition (1.5) and the two other conditions (skew symmetry and the Jacobi identity) guaranteeing that the set $\{t_{\mu\nu}^\sigma\}$ defines a Lie algebra. With that, we simply declare

$$\partial(t_{\mu\nu}^\sigma)=0, \tag{3.8}$$
$$r\not\subset(t_{\mu\nu}^\sigma)=1\quad\text{for}\quad t_{\mu\nu}^\sigma\neq 0. \tag{3.9}$$

However, we meet now another difficulty, namely, that $\{r\not\subset(\operatorname{Ker}\partial|_{\overline{C}})\}=\mathbb{Z}_+$ as (3.8) and (3.9) show, so that again we fall short of deducing (3.2) and (3.3) from (3.1u) and (3.1s). A brief contemplation leads one to conclude that the only way out is to conjecture that

$$\frac{\partial H_m}{\partial t_{\mu\nu}^\sigma}=0,\quad\text{all } m,\sigma,\mu,\nu, \tag{3.10}$$

which restores the needed property $r\not\subset(\operatorname{Ker}\partial|_C)=0$. Formula (3.10) follows from Theorem 1.4 (proven in Sec. V) and formula (1.26). To summarize, we impose the grading (3.4) on C, and look for a sequence $\{H_m\}$ in C satisfying

$$r\not\subset(H_m)=m+1. \tag{3.11}$$

Then (3.2) and (3.3) *do* follow from (3.1u) and (3.1s) for $m>0$, when

$$r\not\subset(\delta H_{m+1}/\delta h)=r\not\subset(\delta H_{m+1}/\delta r_\mu)=m>0,$$

and both (3.2) and (3.3) are satisfied for $m=0$ as can be seen from (1.7) and (1.8). Substituting (3.3) into (3.1r), we get

$$\partial\left(\frac{\delta H_{m+1}}{\delta \mathbf{s}}\right)=\partial \mathbf{r}\left(\frac{\delta H_m}{\delta h}\right)-\left[\mathbf{r},\frac{\delta H_m}{\delta \mathbf{r}}\right]-\left[\mathbf{s},\frac{\delta H_m}{\delta \mathbf{s}}\right]. \tag{3.12}$$

Now notice that, although we do not possess exact formulas for $\delta H_{m+1}/\delta u$ and $\delta H_{m+1}/\delta \mathbf{s}$ but only for their ∂-images, the rhs's of (3.1u), (3.1h), (3.1s) involve these functional derivatives also *only* as images of ∂. Therefore, we can iterate the bi-Hamiltonian definition $B^1\delta(H_{m+1})=B^2\delta(H_m)$, to obtain for $B^2\delta(H_{m+1})$,

$$\partial\left[2\frac{\delta H_{m+1}}{\delta u}+(u-\partial)\left(\frac{\delta H_{m+1}}{\delta h}\right)\right]\quad[\text{by (3.2) and (3.5)}]$$
$$=2\left[(u+\partial)\partial\left(\frac{\delta H_m}{\delta u}\right)+(h\partial+\partial h)\left(\frac{\delta H_m}{\delta u}\right)+\mathbf{r}^t\partial\left(\frac{\delta H_m}{\delta \mathbf{r}}\right)+\mathbf{s}^t\partial\left(\frac{\delta H_m}{\delta \mathbf{s}}\right)\right]$$
$$+\partial(u-\partial)\left[2\frac{\delta H_m}{\delta u}+(u-\partial)\left(\frac{\delta H_m}{\delta h}\right)\right]$$
$$=2(u\partial+\partial u)\left(\frac{\delta H_m}{\delta u}\right)+[\partial(u-\partial)^2+2(h\partial+\partial h)]\left(\frac{\delta H_m}{\delta h}\right)+2\mathbf{r}^t\partial\left(\frac{\delta H_m}{\delta \mathbf{r}}\right)+2\mathbf{s}^t\partial\left(\frac{\delta H_m}{\delta \mathbf{s}}\right); \tag{3.13u}$$

$$(u+\partial)\partial\left(\frac{\delta H_{m+1}}{\delta u}\right)+(h\partial+\partial h)\left(\frac{\delta H_{m+1}}{\delta h}\right)+\mathbf{r}^t\partial\left(\frac{\delta H_{m+1}}{\delta \mathbf{r}}\right)+\mathbf{s}^t\partial\left(\frac{\delta H_{m+1}}{\delta \mathbf{s}}\right)\quad[\text{by (3.1}h\text{), (3.2), (3.3), and (3.12)}]$$
$$=(u+\partial)\left[(u+\partial)\partial\left(\frac{\delta H_m}{\delta u}\right)+(h\partial+\partial h)\left(\frac{\delta H_m}{\delta h}\right)+\mathbf{r}^t\partial\left(\frac{\delta H_m}{\delta \mathbf{r}}\right)+\mathbf{s}^t\partial\left(\frac{\delta H_m}{\delta \mathbf{s}}\right)\right]+(h\partial+\partial h)\left[2\frac{\delta H_m}{\delta u}\right.$$
$$\left.+(u-\partial)\left(\frac{\delta H_m}{\delta h}\right)\right]+\mathbf{r}^t\partial\left(\mathbf{s}\frac{\delta H_m}{\delta h}+2\frac{\delta H_m}{\delta \mathbf{s}}\right)+\mathbf{s}^t\left\{\partial \mathbf{r}\left(\frac{\delta H_m}{\delta h}\right)-\left[\mathbf{r},\frac{\delta H_m}{\delta \mathbf{r}}\right]-\left[\mathbf{s},\frac{\delta H_m}{\delta s}\right]\right\}\quad[\text{by (1.5)}]$$
$$=[(u+\partial)^2\partial+2(h\partial+\partial h)]\left(\frac{\delta H_m}{\delta u}\right)+[(u+\partial)(h\partial+\partial h)+(h\partial+\partial h)(u-\partial)$$
$$+(\mathbf{r}^t\mathbf{s}\partial+\partial\mathbf{r}^t\mathbf{s})]\left(\frac{\delta H_m}{\delta h}\right)+\{(u+\partial)\mathbf{r}^t\partial-[\mathbf{s},\mathbf{r}]^t\}\left(\frac{\delta H_m}{\delta \mathbf{r}}\right)+[(u+\partial)\mathbf{s}^t\partial+2\mathbf{r}^t\partial]\left(\frac{\delta H_m}{\delta \mathbf{s}}\right); \tag{3.13h}$$

$$\partial\mathbf{r}\left(\frac{\delta H_{m+1}}{\delta h}\right)-\left[\mathbf{r},\frac{\delta H_{m+1}}{\delta \mathbf{r}}\right]\quad\text{[by (3.2) and (3.3)]}$$
$$=\partial\mathbf{r}\left[2\frac{\delta H_m}{\delta u}+(u-\partial)\left(\frac{\delta H_m}{\delta h}\right)\right]-[\mathbf{r},\mathbf{s}]\frac{\delta H_m}{\delta h}-2\left[\mathbf{r},\frac{\delta H_m}{\delta \mathbf{s}}\right]$$
$$=2\partial\mathbf{r}\left(\frac{\delta H_m}{\delta u}\right)+\{\partial\mathbf{r}(u-\partial)-[\mathbf{r},\mathbf{s}]\}\left(\frac{\delta H_m}{\delta h}\right)-2\left[\mathbf{r},\frac{\delta H_m}{\delta \mathbf{s}}\right];\tag{3.13r}$$
$$\partial\left(\mathbf{s}\frac{\delta H_{m+1}}{\delta h}+2\frac{\delta H_{m+1}}{\delta \mathbf{s}}\right)\quad\text{[by (3.2) and (3.12)]}$$
$$=\partial\mathbf{s}\left[2\frac{\delta H_m}{\delta u}+(u-\partial)\left(\frac{\delta H_m}{\delta h}\right)\right]+2\left\{\partial\mathbf{r}\left(\frac{\delta H_m}{\delta h}\right)-\left[\mathbf{r},\frac{\delta H_m}{\delta \mathbf{r}}\right]-\left[\mathbf{s},\frac{\delta H_m}{\delta \mathbf{s}}\right]\right\}$$
$$=2\partial\mathbf{s}\left(\frac{\delta H_m}{\delta u}\right)+[\partial\mathbf{s}(u-\partial)+2\partial\mathbf{r}]\left(\frac{\delta H_m}{\delta h}\right)-2\left[\mathbf{r},\left(\frac{\delta H_m}{\delta \mathbf{r}}\right)\right]-2\left[\mathbf{s},\left(\frac{\delta H_m}{\delta \mathbf{s}}\right)\right].\tag{3.13s}$$

The rhs of (3.13) is precisely $B^3\delta H_m$, as (1.13) shows. Thus (1.15b) is deduced. ■

IV. LIE ALGEBRA $\mathscr{L}(\lambda)$ AND ASSOCIATED OBJECTS

On the K-module $K\oplus K\oplus K\oplus K$ we consider the following Lie algebra structure, denoted $\mathscr{L}(\lambda)$:

$$\left[\begin{pmatrix}\varphi_1\\X_1\\Y_1\\Z_1\end{pmatrix},\begin{pmatrix}\varphi_2\\X_2\\Y_2\\Z_2\end{pmatrix}\right]=\begin{pmatrix}X_1\varphi_2^{(1)}-X_2\varphi_1^{(1)}\\X_1X_2^{(1)}-X_1^{(1)}X_2\\(X_1Y_2^{(1)}-X_1^{(1)}Y_2)-(X_2Y_1^{(1)}-X_2^{(1)}Y_1)+(Z_1Z_2^{(1)}-Z_1^{(1)}Z_2)\\(X_1Z_2^{(1)}-X_1^{(1)}Z_2)-(X_2Z_1^{(1)}-X_2^{(1)}Z_1)\end{pmatrix}$$
$$+\begin{pmatrix}0\\0\\\lambda(Z_1Y_2^{(1)}-Z_1^{(1)}Y_2)-\lambda(Z_2Y_1^{(1)}-Z_2^{(1)}Y_1)\\\lambda(Z_1Z_2^{(1)}-Z_1^{(1)}Z_2)\end{pmatrix},\quad\varphi_l,X_l,Y_l,Z_l\in K,\quad l=1,2.\tag{4.1}$$

Apart from the φ-component, the remaining three-dimensional piece of the commutator in (4.1) is a particular case $\mathscr{M}=\mathscr{D}$ of the following simple fact: If $\mathscr{M}$ is an abstract Lie algebra, then on $\mathscr{M}^3$ one has the following Lie algebra structure:

$$\left[\begin{pmatrix}X_1\\Y_1\\Z_1\end{pmatrix},\begin{pmatrix}X_2\\Y_2\\Z_2\end{pmatrix}\right]=\begin{pmatrix}[X_1,X_2]\\[X_1,Y_2]-[X_2,Y_1]+[Z_1,Z_2]+\lambda[Z_1,Y_2]-\lambda[Z_2,Y_1]\\[X_1,Z_2]-[X_2,Z_1]+\lambda[Z_1,Z_2]\end{pmatrix},\quad X_l,Y_l,Z_l\in\mathscr{M},\quad l=1,2.\tag{4.2}$$

Since both $\mathscr{g}(\lambda)$ and $\mathscr{L}(\lambda)$ have a common Lie subalgebra $\mathscr{D}\circledS V_0$ (the first two components), from Lemma 2.1 we immediately obtain the following lemma.

Lemma 4.1: The following are two-cocycles on $\mathscr{L}(\lambda)$:

$$\bar\omega_1(1,2)=\varphi_1X_2^{(1)}+X_1\varphi_2^{(1)},\tag{4.3}$$
$$\bar\omega_2(1,2)=-\varphi_1X_2^{(2)}+X_1\varphi_2^{(2)},\tag{4.4}$$
$$\bar\omega_3(1,2)=\varphi_1\varphi_2^{(1)}.\tag{4.5}$$

Computing the matrix $B(\mathscr{L}(\lambda))$ associated with the Lie algebra $\mathscr{L}(\lambda)$, we get with the help of (2.12) and (4.1):

$$B(\mathscr{L}(\lambda))=\begin{pmatrix}0&\partial u&0&0\\u\partial&h\partial+\partial h&q\partial+\partial q&p\partial+\partial p\\0&q\partial+\partial q&0&0\\0&p\partial+\partial p&0&q\partial+\partial q\end{pmatrix}$$
$$+\lambda\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&0&q\partial+\partial q\\0&0&q\partial+\partial q&p\partial+\partial p\end{pmatrix}.\tag{4.6}$$

Let $\bar\beta_l$ be the matrix corresponding to the two-cocycle $\bar\omega_l$ [(4.2) + l], $l=1,2,3$. Set

$$\bar B(\lambda)=B(\mathscr{L}(\lambda))+\lambda\bar\beta_1+\bar\beta_2+2\bar\beta_3.\tag{4.7}$$

The matrix $\bar B(\lambda)$ is Hamiltonian for any λ. In particular, the following matrices are also Hamiltonian:

$$\bar B^1=\lambda^{-1}\bar B(\lambda)|_{1/\lambda=0},\tag{4.8}$$
$$\bar B^2=\bar B(\lambda)|_{\lambda=0},\tag{4.9}$$

which are precisely the matrices (1.16) and (1.17).

We now embark on the proof of Theorem 1.3. Writing Eq. (1.24) in longhand, we get

$$\partial\left(\frac{\delta\bar H_{m+1}}{\delta h}\right)=\partial\left[2\left(\frac{\delta\bar H_m}{\delta u}\right)+(u-\partial)\left(\frac{\delta\bar H_m}{\delta h}\right)\right],\tag{4.10u}$$

$$\partial\left(\frac{\delta\bar{H}_{m+1}}{\delta u}\right)=(u+\partial)\partial\left(\frac{\delta\bar{H}_m}{\delta u}\right)+(h\partial+\partial h)\left(\frac{\delta\bar{H}_m}{\delta h}\right)+(q\partial+\partial q)\left(\frac{\delta\bar{H}_m}{\delta q}\right)+(p\partial+\partial p)\left(\frac{\delta\bar{H}_m}{\delta p}\right), \tag{4.10h}$$

$$(q\partial+\partial q)\left(\frac{\delta\bar{H}_{m+1}}{\delta p}\right)=(q\partial+\partial q)\left(\frac{\delta\bar{H}_m}{\delta h}\right), \tag{4.10q}$$

$$(p\partial+\partial p)\left(\frac{\delta\bar{H}_{m+1}}{\delta p}\right)+(q\partial+\partial q)\left(\frac{\delta\bar{H}_{m+1}}{\delta q}\right)=(p\partial+\partial p)\left(\frac{\delta\bar{H}_m}{\delta h}\right)+(q\partial+\partial q)\left(\frac{\delta\bar{H}_m}{\delta p}\right). \tag{4.10p}$$

Equations (4.10) are obviously satisfied for $m=0,1,2$, as seen from (1.19)–(1.22). For $m>1$, $r\not{k}(\delta\bar{H}_{m+1}/\delta h)=m-1>0$, so that (4.10$u$) implies (since $\mathrm{Ker}\,\partial|_{C_1}=\not{k}$)

$$\frac{\delta\bar{H}_{m+1}}{\delta h}=2\frac{\delta\bar{H}_m}{\delta u}+(u-\partial)\left(\frac{\delta\bar{H}_m}{\delta h}\right). \tag{4.11}$$

Also, since $q\partial+\partial q=2q^{1/2}\,\partial q^{1/2}$, we conclude that $\mathrm{Ker}(q\partial+\partial q)|_{C_1}=\{0\}$, and analogously for $p\partial+\partial p$. Hence from (4.10q) we obtain

$$\frac{\delta\bar{H}_{m+1}}{\delta p}=\frac{\delta\bar{H}_m}{\delta h}, \tag{4.12}$$

and substituting this into (4.10p) we get

$$\delta\bar{H}_{m+1}/\delta q=\delta\bar{H}_m/\delta p. \tag{4.13}$$

We first prove (1.24) and then deduce (1.25) from it. We prove (1.24) by induction on m. Denote

$$T_m=\delta(\bar{H}_m)=\begin{pmatrix}x_m\\ y_m\\ z_m\\ d_m\end{pmatrix}. \tag{4.14}$$

Suppose we have already found properly homogeneous elements $\bar{H}_0,\ldots,\bar{H}_m$ in C_1' satisfying (1.24), and want to determine $\bar{H}_{m+1}$. We do it in two substeps: (i) First, from (4.10) we determine T_{m+1}; and (ii) then we show that there exists an element in C_1' of correct grading whose vector of functional derivatives is the thus determined vector T_{m+1}. This will conclude the induction step $m\mapsto m+1$.

(i) Formulas (4.11)–(4.13) determine y_{m+1}, d_{m+1}, and z_{m+1}, with the correct grading

$$r\not{k}(y_{m+1})=m,\quad r\not{k}(z_{m+1})=m-2,\quad r\not{k}(d_{m+1})=m-1. \tag{4.15}$$

To show that $x_{m+1}\in C_1'$, we have to show that the rhs of (4.10h) is ~ 0. Since the rhs of (4.10h) is

$$\sim u\partial\left(\frac{\delta\bar{H}_m}{\delta u}\right)+h\partial\left(\frac{\delta\bar{H}_m}{\delta h}\right)+q\partial\left(\frac{\delta\bar{H}_m}{\delta q}\right)+p\partial\left(\frac{\delta\bar{H}_m}{\delta p}\right),$$

it is indeed ~ 0, since, as is well known,

$$\sum_\alpha v_\alpha\partial\left(\frac{\delta H}{\delta v_\alpha}\right)\sim 0,\quad \forall H\in\not{k}[v_\alpha^{(j)}]. \tag{4.16}$$

Also, from (4.10h) we see that we can take x_{m+1} to have

$$r\not{k}(x_{m+1})=m+1. \tag{4.17}$$

Thus we have found a vector T_{m+1} of correct grading satisfying

$$\bar{B}^1(T_{m+1})=\bar{B}^2(T_m). \tag{4.18}$$

(ii) To show that the thus obtained vector T_{m+1} is a vector of functional derivatives of some element in C_1' (which all the components of the vector T_{m+1} belong to), it is necessary and sufficient to show (see Refs. 11 and 12) that the Fréchet derivative $D(T_{m+1})$ of T_{m+1} is symmetric:

$$D(T_{m+1})^\dagger=D(T_{m+1}). \tag{4.19}$$

Denote $G_m:=D(T_m)$. Since $\bar{B}^1$ (1.16) is a nondegenerate matrix, instead of (4.19) we show that $\bar{B}^1D(T_{m+1})\bar{B}^1$ is symmetric. To do that, we apply the Fréchet derivative operator D to the both sides of Eq. (4.18), obtaining

$$\bar{B}^1G_{m+1}+\begin{pmatrix}0&0&0&0\\0&0&0&0\\0&0&d_{m+1}\partial+2d_{m+1}^{(1)}&0\\0&0&z_{m+1}\partial+2z_{m+1}^{(1)}&d_{m+1}\partial+2d_{m+1}^{(1)}\end{pmatrix}=\bar{B}^2G_m+\begin{pmatrix}y_m\partial+y_m^{(1)}&0&0&0\\x_m^{(1)}&y_m\partial+2y_m^{(1)}&z_m\partial+2z_m^{(1)}&d_m\partial+2d_m^{(1)}\\0&0&y_m\partial+2y_m^{(1)}&0\\0&0&d_m\partial+2d_m^{(1)}&y_m\partial+2y_m^{(1)}\end{pmatrix}. \tag{4.20}$$

Using (4.12) and (4.13), we simplify (4.20) into

$$\bar{B}^1G_{m+1}=\bar{B}^2G_m+F_m, \tag{4.21.m}$$

where

$$F_m=\begin{pmatrix}y_m\partial+y_m^{(1)}&0&0&0\\x_m^{(1)}&y_m\partial+2y_m^{(1)}&z_m\partial+2z_m^{(1)}&d_m\partial+2d_m^{(1)}\\0&0&0&0\\0&0&0&0\end{pmatrix}. \tag{4.22}$$

Now multiply $(4.21.m-1)$ from the right by $\overline{B}^1$, and subtract from the result $[(4.21.m-1)]$ multiplied from the right by $\overline{B}^2$:

$$\overline{B}^1 G_{m+1} \overline{B}^1 = (\overline{B}^2 G_m \overline{B}^1 + \overline{B}^1 G_m \overline{B}^2) - \overline{B}^2 G_{m-1} \overline{B}^2 + \tilde{F}_m, \tag{4.23}$$

where

$$\tilde{F}_m := F_m \overline{B}^1 - F_{m-1} \overline{B}^2. \tag{4.24}$$

From (4.23) it follows that, since G_m and G_{m-1} are symmetric due to the inductive assumption, $\overline{B}^1 G_{m+1} \overline{B}^1$ is symmetric iff $\tilde{F}_m$ is. Computing $\tilde{F}_m$ from (4.22), (4.24), (1.16), (1.17), we find that $\tilde{F}_m$ vanishes outside the upper left 2×2 corner, where it equals

$$\left(\begin{array}{l|l} -2\partial y_{m-1}\partial & \partial y_m \partial - \partial y_{m-1}\partial(u-\partial) \\ \hline (y_m\partial + 2y_m^{(1)})\partial & x_m^{(1)}\partial - x_{m-1}^{(1)}\partial(u-\partial) - (y_{m-1}\partial + 2y_{m-1}^{(1)})(h\partial + \partial h) \\ \quad -(y_{m-1}\partial + 2y_{m-1}^{(1)})(u+\partial)\partial & \quad -(z_{m-1}\partial + 2z_{m-1}^{(1)}(q\partial + \partial q) \\ \quad -2x_{m-1}^{(1)}\partial & \quad -(d_{m-1}\partial + 2d_{m-1}^{(1)})(p\partial + \partial p) \end{array}\right). \tag{4.25}$$

Using $(4.10h)$ and (4.11)–(4.13), one checks by a lengthy but straightforward calculation that the matrix (4.25) is indeed symmetric. This finishes the proof of (1.24).

Now we deduce (1.25) from (1.24), using the same arguments which we have employed in the preceding section when deriving B^3. For the lhs of (1.25) we get

$$\begin{aligned} &2\partial\left(\frac{\delta \overline{H}_{m+1}}{\delta u}\right) + \partial(u-\partial)\left(\frac{\delta \overline{H}_{m+1}}{\delta h}\right) \quad [\text{by } (4.10h) \text{ and } (4.11)] \\ &\quad = 2\left[(u+\partial)\partial\left(\frac{\delta \overline{H}_m}{\delta u}\right) + (h\partial+\partial h)\left(\frac{\delta \overline{H}_m}{\delta h}\right) + (q\partial+\partial q)\left(\frac{\delta \overline{H}_m}{\delta q}\right) + (p\partial+\partial p)\left(\frac{\delta \overline{H}_m}{\delta p}\right)\right] \\ &\qquad + \partial(u-\partial)\left[2\left(\frac{\delta \overline{H}_m}{\delta u}\right) + (u-\partial)\left(\frac{\delta \overline{H}_m}{\delta h}\right)\right] \\ &\quad = 2(u\partial+\partial u)\left(\frac{\delta \overline{H}_m}{\delta u}\right) + [\partial(u-\partial)^2 + 2(h\partial+\partial h)]\left(\frac{\delta \overline{H}_m}{\delta h}\right) + 2(q\partial+\partial q)\left(\frac{\delta \overline{H}_m}{\delta q}\right) + 2(p\partial+\partial p)\left(\frac{\delta \overline{H}_m}{\delta p}\right); \end{aligned} \tag{4.26u}$$

$$\begin{aligned} &(u+\partial)\partial\left(\frac{\delta \overline{H}_{m+1}}{\delta u}\right) + (h\partial+\partial h)\left(\frac{\delta \overline{H}_{m+1}}{\delta h}\right) + (q\partial+\partial q)\left(\frac{\delta \overline{H}_{m+1}}{\delta q}\right) \\ &\qquad + (p\partial+\partial p)\left(\frac{\delta \overline{H}_{m+1}}{\delta p}\right) \quad [\text{by } (4.10h), (4.11)\text{–}(4.13)] \\ &\quad = (u+\partial)\left[(u+\partial)\partial\left(\frac{\delta \overline{H}_m}{\delta u}\right) + (h\partial+\partial h)\left(\frac{\delta \overline{H}_m}{\delta h}\right) + (q\partial+\partial q)\left(\frac{\delta \overline{H}_m}{\delta q}\right) + (p\partial+\partial p)\left(\frac{\delta \overline{H}_m}{\delta p}\right)\right] \\ &\qquad + (h\partial+\partial h)\left[2\frac{\delta \overline{H}_m}{\delta u} + (u-\partial)\left(\frac{\delta \overline{H}_m}{\delta h}\right) + (q\partial+\partial q)\left(\frac{\delta \overline{H}_m}{\delta p}\right) + (p\partial+\partial p)\left(\frac{\delta \overline{H}_m}{\delta h}\right)\right] \\ &\quad = [(u+\partial)^2\partial + 2(h\partial+\partial h)]\left(\frac{\delta \overline{H}_m}{\delta u}\right) + [(u+\partial)(h\partial+\partial h) + (h\partial+\partial h)(u-\partial) + (p\partial+\partial p)]\left(\frac{\delta \overline{H}_m}{\delta h}\right) \\ &\qquad + (u+\partial)(q\partial+\partial q)\left(\frac{\delta \overline{H}_m}{\delta q}\right) + [(u+\partial)(p\partial+\partial p) + (q\partial+\partial q)]\left(\frac{\delta \overline{H}_m}{\delta p}\right); \end{aligned} \tag{4.26h}$$

$$\begin{aligned} &(q\partial+\partial q)\left(\frac{\delta \overline{H}_{m+1}}{\delta h}\right) \quad [\text{by } (4.11)] \\ &\quad = (q\partial+\partial q)\left[2\frac{\delta \overline{H}_m}{\delta u} + (u+\partial)\left(\frac{\delta \overline{H}_m}{\delta h}\right)\right] = 2(q\partial+\partial q)\left(\frac{\delta \overline{H}_m}{\delta u}\right) + (q\partial+\partial q)(u-\partial)\left(\frac{\delta \overline{H}_m}{\delta h}\right); \end{aligned} \tag{4.26q}$$

$$\begin{aligned} &(p\partial+\partial p)\left(\frac{\delta \overline{H}_{m+1}}{\delta h}\right) + (q\partial+\partial q)\left(\frac{\delta \overline{H}_{m-1}}{\delta p}\right) \quad [\text{by } (4.11) \text{ and } (4.12)] \\ &\quad = (p\partial+\partial p)\left[2\left(\frac{\delta \overline{H}_m}{\delta u}\right) + (u-\partial)\left(\frac{\delta \overline{H}_m}{\delta h}\right)\right] + (q\partial+\partial q)\left(\frac{\delta \overline{H}_m}{\delta h}\right) \\ &\quad = 2(p\partial+\partial p)\left(\frac{\delta \overline{H}_m}{\delta u}\right) + [(p\partial+\partial p)(u-\partial) + (q\partial+\partial q)]\left(\frac{\delta \overline{H}_m}{\delta h}\right), \end{aligned} \tag{4.26p}$$

which is exactly $\overline{B}^3 \delta(\overline{H}_m)$ for $\overline{B}^3$ given by (1.18). ■

V. PROOF OF THEOREM 1.4

Since $r\not{k}(\bar{H}_m) = m+1$ and Φ (1.26) preserves the gradings (in C_1 and C), we see that $r\not{k}(H_m) = r\not{k}(\bar{H}_{m+1}) = m+1$, as required in (3.11). We have to check Eq. (3.1). For this we need to compute the functional derivatives of the Hamiltonians $\Phi(\bar{H}_m)$. We use the following formula[8,9]:

$$\frac{\delta\Phi(\bar{H})}{\delta(\cdots)} = D(\Phi)^{\dagger}\Phi\left(\frac{\delta\bar{H}}{\delta(\cdots)}\right), \tag{5.1}$$

where $(\cdots)$ denotes the vector of basic variables in the corresponding ring. Since

$$D(\Phi) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 2\mathbf{r} & \mathbf{s} \\ 0 & 0 & 0 & \mathbf{r} \end{pmatrix}, \tag{5.2}$$

Eq. (5.1) yields

$$\frac{\delta H_m}{\delta u} = \Phi\left(\frac{\delta\bar{H}_m}{\delta u}\right), \tag{5.3}$$

$$\frac{\delta H_m}{\delta h} = \Phi\left(\frac{\delta\bar{H}_m}{\delta h}\right), \tag{5.4}$$

$$\frac{\delta H_m}{\delta \mathbf{r}} = 2\mathbf{r}\Phi\left(\frac{\delta\bar{H}_m}{\delta q}\right) + \mathbf{s}\Phi\left(\frac{\delta\bar{H}_m}{\delta p}\right), \tag{5.5}$$

$$\frac{\delta H_m}{\delta \mathbf{s}} = \mathbf{r}\Phi\left(\frac{\delta\bar{H}_m}{\delta p}\right). \tag{5.6}$$

Using (5.3)–(5.6), we check separately each of Eqs. (3.1). For the u-component, we have

$$\begin{aligned} \partial\left(\frac{\delta\bar{H}_{m-1}}{\delta h}\right) \quad [\text{by (5.4)}] &= \Phi\partial\left(\frac{\delta\bar{H}_{m+1}}{\delta h}\right) \quad [\text{by (4.10}u\text{)}] \\ &= \partial\Phi\left[2\frac{\delta\bar{H}_m}{\delta u} + (u-\partial)\left(\frac{\delta\bar{H}_m}{\delta h}\right)\right] \quad [\text{by (5.3) and (5.4)}] \\ &= \partial\left[2\frac{\delta H_m}{\delta u} + (u-\partial)\left(\frac{\delta H_m}{\delta u}\right)\right], \end{aligned} \tag{5.7u}$$

which proves (3.1u). For the h-component, we have

$$\begin{aligned} &\partial\left(\frac{\delta H_{m+1}}{\delta u}\right) \quad [\text{by (5.3)}] \\ &= \Phi\partial\left(\frac{\delta\bar{H}_{m+1}}{\delta u}\right) \quad [\text{by (4.10}h\text{)}] \\ &= \Phi\left\{(u+\partial)\partial\left(\frac{\delta\bar{H}_m}{\delta u}\right) + (h\partial+\partial h)\left(\frac{\delta\bar{H}_m}{\delta h}\right) + (q\partial+\partial q)\left(\frac{\delta\bar{H}_m}{\delta q}\right) + (p\partial+\partial p)\left(\frac{\delta\bar{H}_m}{\delta p}\right)\right\} \quad [\text{by (5.3)–(5.6)}] \\ &= (u+\partial)\partial\left(\frac{\delta H_m}{\delta u}\right) + (h\partial+\partial h)\left(\frac{\delta H_m}{\delta h}\right) + \mathbf{r}^t\partial\left(\frac{\delta H_m}{\delta \mathbf{r}}\right) + \mathbf{s}^t\partial\left(\frac{\delta H_m}{\delta \mathbf{s}}\right), \end{aligned} \tag{5.7h}$$

since

$$\begin{aligned} \mathbf{r}^t\partial\left(\frac{\delta H_m}{\delta \mathbf{r}}\right) + \mathbf{s}^t\partial\left(\frac{\delta H_m}{\delta \mathbf{s}}\right) \quad [\text{by (5.5) and (5.6)}] &= \mathbf{r}^t 2\,\partial\mathbf{r}\,\Phi\left(\frac{\delta\bar{H}_m}{\delta q}\right) + (\mathbf{r}^t\partial\mathbf{s} + \mathbf{s}^t\partial\mathbf{r})\Phi\left(\frac{\delta\bar{H}_m}{\delta p}\right) \quad [\text{by (1.26)}] \\ &= \Phi\left\{(q\partial+\partial q)\left(\frac{\delta\bar{H}_m}{\delta q}\right) + (p\partial+\partial p)\left(\frac{\delta H_m}{\delta p}\right)\right\}; \end{aligned}$$

thus, (3.1h) is verified. For the $\mathbf{r}$-component, we have

$$\frac{1}{2}\left[\mathbf{s},\frac{\delta H_{m+1}}{\delta \mathbf{r}}\right]+\partial\left(\frac{\delta H_{m+1}}{\delta \mathbf{s}}\right) \quad \text{[by (5.5) and (5.6)]}$$

$$=[\mathbf{s},\mathbf{r}]\Phi\left(\frac{\delta \bar{H}_{m+1}}{\delta q}\right)+\partial \mathbf{r}\,\Phi\left(\frac{\delta \bar{H}_{m+1}}{\delta p}\right) \quad \text{[by (4.12) and (4.13)]}$$

$$=[\mathbf{s},\mathbf{r}]\Phi\left(\frac{\delta \bar{H}_{m}}{\delta p}\right)+\partial \mathbf{r}\,\Phi\left(\frac{\delta \bar{H}_{m}}{\delta h}\right) \quad \text{[by (5.4) and (5.5)]}=\left[-\mathbf{r},\frac{\delta H_m}{\delta \mathbf{r}}\right]+\partial \mathbf{r}\left(\frac{\delta H_m}{\delta h}\right), \tag{5.7r}$$

which proves (3.1r). Finally, for the s-component, we have

$$\partial\left(\frac{\delta \bar{H}_{m+1}}{\delta \mathbf{r}}\right) \quad \text{[by (5.5)]}=\partial\, 2\mathbf{r}\Phi\left(\frac{\delta \bar{H}_{m+1}}{\delta q}\right)+\partial \mathbf{s}\Phi\left(\frac{\delta \bar{H}_{m+1}}{\delta p}\right) \quad \text{[by (4.12) and (4.13)]}$$

$$=\partial\, 2\mathbf{r}\Phi\left(\frac{\delta \bar{H}_{m}}{\delta p}\right)+\partial \mathbf{s}\Phi\left(\frac{\delta \bar{H}_{m}}{\delta h}\right) \quad \text{[by (5.4) and (5.6)]}=\partial\, 2\left(\frac{\delta H_m}{\delta \mathbf{s}}\right)+\partial \mathbf{s}\left(\frac{\delta H_m}{\delta h}\right), \tag{5.7s}$$

which proves (3.1s). ∎

VI. LINEARIZATION OF THE MATRIX B^3

Let $C_2=K[u^{(j)},Q^{(j)},r_\mu^{(j)},s_\mu^{(j)}]$, $j\in\mathbb{Z}_+$, $1\leqslant\mu\leqslant N$, be another differential algebra. Consider the following *invertible* differential homomorphism Ψ: $C\to C_2$:

$$\Psi(u)=u,\quad \Psi(h)=Q-\tfrac{1}{4}u^2-\tfrac{1}{2}u^{(1)}-\tfrac{1}{4}\mathbf{s}'\mathbf{s},\quad \Psi(r_\mu)=r_\mu,\quad \Psi(s_\mu)=s_\mu. \tag{6.1}$$

Theorem 6.1: In the ring C_2, the matrix B^3 (1.13) becomes $2\tilde{B}^3$, where $\tilde{B}^3$ is the following affine matrix:

$$\tilde{B}^3=\begin{pmatrix} u\partial+\partial u & Q\partial+\partial Q+\frac{1}{2}\partial^3 & \mathbf{r}'\partial & \mathbf{s}'\partial \\ Q\partial+\partial Q+\frac{1}{2}\partial^3 & 0 & 0 & \mathbf{r}'\partial \\ \partial\mathbf{r} & 0 & 0 & -[\mathbf{r},\,] \\ \partial\mathbf{s} & \partial\mathbf{r} & -[\mathbf{r},\,] & -[\mathbf{s},\,] \end{pmatrix}. \tag{6.2}$$

Proof: We have, using (1.13),

$$D(\Psi)=\begin{pmatrix} 1 & 0 & 0 & 0 \\ -\frac{1}{2}(u+\partial) & 1 & 0 & -\frac{1}{2}\mathbf{s}' \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \end{pmatrix}, \tag{6.3}$$

$$D(\Psi)^\dagger=\begin{pmatrix} 1 & \frac{1}{2}(\partial-u) & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & -\frac{1}{2}\mathbf{s} & 0 & 1 \end{pmatrix}, \tag{6.4}$$

$$D(\Psi)B^3=\left(\begin{array}{c|c|c|c} 2(u\partial+\partial u) & \partial(u-\partial)^2+2(h\partial+\partial h) & 2\mathbf{r}'\partial & 2\mathbf{s}'\partial \\ -(u+\partial)(u\partial+\partial u) & -\frac{1}{2}(u+\partial)\partial(u-\partial)^2 & 0 & 2\mathbf{r}'\partial \\ \quad +(u+\partial)^2\partial & \quad +(h\partial+\partial h)(u-\partial) & & \\ \quad +2(h\partial+\partial h) & \quad +\mathbf{r}'\partial\mathbf{s}-\frac{1}{2}\mathbf{s}'\partial\mathbf{s}(u-\partial) & & \\ \quad -\mathbf{s}'\partial\mathbf{s} & & & \\ \hline 2\partial\mathbf{r} & \partial\mathbf{r}(u-\partial)-[\mathbf{r},\mathbf{s}] & 0 & -2[\mathbf{r},\,] \\ \hline 2\partial\mathbf{s} & \partial\mathbf{s}(u-\partial)+2\,\partial\mathbf{r} & -2[\mathbf{r},\,] & -2[\mathbf{s},\,] \end{array}\right), \tag{6.5}$$

and then $D(\Psi)B^3D(\Psi)^\dagger=2\Psi(\tilde{B}^3)$, in accordance with formula (1.27). ∎

Thus we have found coordinates in which our nonlinear matrix B^3 becomes linear. (In new coordinates the linear matrices B^1 and B^2 remain linear, but we will not need this fact here.) To prove Theorem 1.1 (ii), we will show that the matrix $\tilde{B}^3$ (6.2) is Hamiltonian. Since the matrix $\tilde{B}^3$ is affine, the Lie-algebraic route is available. On the space $K^2\oplus(K\otimes\mathcal{G})\oplus(K\otimes\mathcal{G})$, consider the following structure of a Lie algebra:

$$\left[\begin{pmatrix} X_1 \\ \overline{X}_1 \\ \bar{f}_1\otimes\bar{a}_1 \\ f_1\otimes a_1 \end{pmatrix},\begin{pmatrix} X_2 \\ \overline{X}_2 \\ \bar{f}_2\otimes\bar{a}_2 \\ f_2\otimes a_2 \end{pmatrix}\right]=\begin{pmatrix} X_1X_2^{(1)}-X_1^{(1)}X_2 \\ (X_1\overline{X}_2^{(1)}-X_1^{(1)}\overline{X}_2)-(X_2\overline{X}_1^{(1)}-X_2^{(1)}\overline{X}_1) \\ X_1\bar{f}_2^{(1)}\otimes\bar{a}_2-X_2\bar{f}_1^{(1)}\otimes\bar{a}_1+\overline{X}_1f_2^{(1)}\otimes a_2-\overline{X}_2f_1^{(1)}\otimes a_1 \\ X_1f_2^{(1)}\otimes a_2-X_2f_1^{(1)}\otimes a_1+f_1f_2\otimes[a_1,a_2] \end{pmatrix}$$
$$+\begin{pmatrix} 0 \\ 0 \\ f_1\bar{f}_2\otimes[a_1,\bar{a}_2]-f_2\bar{f}_1\otimes[a_2,\bar{a}_1] \\ 0 \end{pmatrix},\quad X_l,\overline{X}_l,\bar{f}_l,f_l\in K,\quad \bar{a}_l,a_l\in\mathscr{g},\quad l=1,2. \tag{6.6}$$

To avoid checking the Jacobi identity for the commutator (6.6), we can notice that this commutator is a particular case of the commutator in the semidirect product Lie algebra

$$\tilde{\mathscr{M}}:=\mathscr{M}\underset{\text{ad}}{\circledS}V(\mathscr{M}) \tag{6.7}$$

for an abstract Lie algebra $\mathscr{M}$, in the case

$$\mathscr{M}=\mathscr{g}_{\text{aff}}, \tag{6.8}$$

where $V(\mathscr{M})$ is $\mathscr{M}$ as a $\mathscr{M}$-module, but considered as an Abelian Lie algebra by itself, and $\mathscr{g}_{\text{aff}}$ is the differential analog of the affine Lie algebra produced by $\mathscr{g}$, with the commutator

$$\left[\begin{pmatrix} X_1 \\ f_1\otimes a_1 \end{pmatrix},\begin{pmatrix} X_2 \\ f_2\otimes a_2 \end{pmatrix}\right]=\begin{pmatrix} X_1X_2^{(1)}-X_1^{(1)}X_2 \\ X_1f_2^{(1)}\otimes a_2-X_2f_1^{(1)}\otimes a_1+f_1f_2\otimes[a_1,a_2] \end{pmatrix},$$
$$X_l,f_l\in K,\quad a_l\in\mathscr{g},\quad l=1,2. \tag{6.9}$$

On the Lie algebra $\tilde{\mathscr{g}}_{\text{aff}}$, we have the following two-cocycle:

$$\tilde{\omega}(1,2)=X_1\overline{X}_2^{(3)}+\overline{X}_1X_2^{(3)}. \tag{6.10}$$

If $\tilde{\beta}$ is the matrix corresponding to the two-cocycle $\tilde{\omega}$, and $B(\tilde{\mathscr{g}}_{\text{aff}})$ is the matrix associated with the Lie algebra $\tilde{\mathscr{g}}_{\text{aff}}$, then the matrix

$$\tilde{B}^3=B(\tilde{\mathscr{g}}_{\text{aff}})+\tfrac{1}{2}\tilde{\beta} \tag{6.11}$$

is Hamiltonian. ■

ACKNOWLEDGMENT

This work was supported in part by the National Science Foundation.

[1]D. J. Benney, "Some properties of long nonlinear waves," Stud. Appl. Math. L **11**, 45 (1973).

[2]B. A. Kupershmidt and Yu. I. Manin, "Long-wave equation with free boundaries. I. Conservation laws," Funct. Anal. Appl. **11**, 188 (1977).

[3]B. A. Kupershmidt and Yu. I. Manin, "Equations of long waves with a free surface. II. Hamiltonian structure and higher equations," Funct. Anal. Appl. **12**, 20 (1978).

[4]B. A. Kupershmidt, "Normal and universal forms in integrable hydrodynamical systems," in *Proceedings of NASA Ames–Berkeley Conference on Nonlinear Problems in Optimal Control and Hydrodynamics*, edited by L. R. Hunt and C. F. Martin (Math. Sci., Brookline, MA, 1984), pp. 357–378.

[5]L. J. F. Broer, "Approximate equations for long water waves," Appl. Sci. Res. **31**, 377 (1975).

[6]D. J. Kaup, "A higher-order water-wave equation and the method for solving it," Prog. Theor. Phys. **54**, 396 (1975).

[7]B. A. Kupershmidt, "Mathematics of dispersive water waves," Commun. Math. Phys. **99**, 51 (1985).

[8]Yu. I. Manin, "Algebraic aspects of non-linear differential equations," Itogi Nauki–Serija Mat. **11**, 5 (1978); J. Sov. Math. **11**, 1(1979).

[9]B. A. Kupershmidt, *Discrete Lax Equations and Differential-Difference Calculus* (Asterisque, Paris, 1985).

[10]B. A. Kupershmidt and G. Wilson, "Modifying Lax equations and the second Hamiltonian structure," Invent. Math. **62**, 403 (1981).

[11]B. A. Kupershmidt, "Geometry of jet bundles and the structures of Lagrangian and Hamiltonian formalisms," in *Lecture Notes in Mathematics*, Vol. 755 (Springer, Berlin, 1980), pp. 162–218.

[12]B. A. Kupershmidt, "Noncommutative Calculus of Variations", in *Lecture Notes in Pure Applied Mathematics*, Vol. 100 (Dekker, New York, 1985), pp. 297–303.

Tensor formulation of Hamilton's equations

Paul H. Lim
Department of Physics, University of Victoria, Victoria, British Columbia V8W 2Y2, Canada

(Received 14 July 1986; accepted for publication 27 April 1988)

Hamilton's equations are presented in manifestly covariant form. The resulting equations of motion are solved via a covariant Hamilton–Jacobi scheme. A covariant correspondence principle is introduced, and it is employed to quantize the equations.

I. INTRODUCTION

Kalman's equations[1,2] describe in manifestly covariant form the motion of a classical particle in an external field in flat space-time. Recently, the author of the present paper examined these equations and showed that they reduce to Lagrange's equations in a specific Lorentz frame.[3] Thus Kalman's equations are the tensorial expression of Lagrange's equations. This work considers the concomitant problem of casting Hamilton's equations into tensor form in flat space-time. The essential equations (called G equations) were derived in Ref. 3 and are presented in Sec. II below. Their solutions are constrained to lie in a seven-dimensional subspace of the space of coordinates and four-momenta. The origin of this constraint is shown to be the Minkowski relation between four-velocity components. In Appendix A, it is shown that the G equations do indeed reduce to Hamilton's equations in a specific Lorentz frame.

Although the development of the tensorial G equations is a logical continuation of the work of Refs. 1 and 3, this approach to particle dynamics is somewhat unconventional. In the standard approach,[4–6] a Hamiltonian formalism is developed from Lagrange's equations in a specific Lorentz frame; the approach is taken in this paper to develop the G equations from Kalman's manifestly covariant equations of motion. It should be noted that the Kalman equations arise from a proper-time parametrization of the action integral, as shown explicitly in Ref. 3. Thus the G equations are equivalent to the standard formalism of Refs. 4–6 in the case of proper-time parametrization.

In Sec. III, Kalman's equations and the G equations are applied to several simple examples, and the corresponding Lagrangian and Hamiltonian formulations are presented for comparison.

In the remainder of the paper, the G equations are developed along two lines. First, in the classical domain, conventional Hamilton–Jacobi theory is extended to apply to the manifestly covariant G equations. This begins in Sec. IV with a derivation of the G equations from a variational principle. Using this method, the G equations arise independently of the Kalman formulation. In Sec. V, the variational principle generates canonical transformations of the G equations. Although the tensor formulation closely parallels its conventional Hamiltonian counterpart, there are important differences arising from the above-mentioned geometrical constraint.[7] Finally, the covariant Hamilton–Jacobi scheme is presented in Sec. VI, and it is used to derive solutions in several simple cases. The second line of development of the G equations is in the quantum-mechanical domain. Section VII introduces a covariant correspondence principle which, for the examples of Sec. III, leads to scalar wave equations. These are shown to be physically realistic, and all reduce in the nonrelativistic limit to the appropriate Schrödinger equation.

II. GENERALIZED HAMILTONIAN FORMULATION

In Refs. 1 and 3, the action functional for a particle in an external field takes the form (limits of integration are removed for simplicity)

$$S=\int f(x,U)ds, \tag{2.1}$$

where x, U, and s are the particle's space-time coordinates, four-velocity, and proper time, respectively, and $f(x,U)$ is a scalar function. The relativistic variational principle for (2.1) in a Lorentz frame yields Kalman's manifestly covariant equations of motion:

$$\frac{\partial f}{\partial x^{\alpha}}=-\frac{dP_{\alpha}}{ds}, \tag{2.2a}$$

$$P_{\alpha}=-\left[\frac{\partial f}{\partial U^{\alpha}}+U_{\alpha}\left(f-U^{\beta}\frac{\partial f}{\partial U^{\beta}}\right)\right]. \tag{2.2b}$$

As shown in Ref. 3, (2.2) have the following properties: (i) Kalman's equations are well defined, despite the relativistic constraint on the four-velocity[8]

$$U\cdot U\equiv\eta_{\mu\nu}U^{\mu}U^{\nu}=1; \tag{2.3}$$

(ii) the P_{α} of (2.2b) are covariant components of four-momentum, i.e.,

$$(P^{\alpha})\equiv(P^{0},\mathbf{P})=(H,\mathbf{P}), \tag{2.4}$$

where H is the Hamiltonian and $\mathbf{P}$ is the canonical three-momentum; (iii) Eqs. (2.2) reduce to Lagrange's equations in a specific Lorentz frame (thus Kalman's equations are the tensorial expression of Lagrange's equations); and (iv) Eqs. (2.2) may be recast into a generalized Hamiltonian form. Since the resulting equations are the subject of this paper, their derivation is described below in some detail.

The first step involved is recognition that (2.2b) with (2.3) imply that x and P are interdependent; hence there exists a scalar equation of the form

$$\phi(x,P)=0. \tag{2.5}$$

Equations (2.2) may be expressed in a Hamiltonian form provided that the constraint of (2.5) is taken into account. The resulting manifestly covariant equations are

 0022-2488/88/092001-09$02.50

$$\frac{dx^\alpha}{ds}=\frac{\partial G}{\partial P_\alpha}-\mu\frac{\partial \phi}{\partial P_\alpha},$$
$$\frac{-dP_\alpha}{ds}=\frac{\partial G}{\partial x^\alpha}-\mu\frac{\partial \phi}{\partial x^\alpha}, \tag{2.6}$$
$$G(x,P)\equiv f+P\cdot U\approx 0,$$

where μ is a scalar Lagrange multiplier arising from (2.5). In the last of (2.6), the symbol $\approx$ indicates that G vanishes[9] weakly[4]; thus the partial derivatives occurring in (2.6) are found before imposition of $G\approx 0$. Equations (2.6), shown in their original form of Ref. 3, may be simplified as follows. For a particle with no external constraints, there is only one relationship of the form (2.5); hence $\phi(x,P)$ may be taken to be identical with $G(x,P)$ in the last of (2.6). With the scalar λ defined as

$$\lambda=1-\mu,$$

(2.6) then take on the compact form

$$\frac{dx^\alpha}{ds}=\lambda\frac{\partial G}{\partial P_\alpha}, \tag{2.7a}$$
$$\frac{-dP_\alpha}{ds}=\lambda\frac{\partial G}{\partial x^\alpha}, \tag{2.7b}$$
$$G(x,P)\equiv f+P\cdot U, \tag{2.7c}$$
$$G(x,P)\approx 0. \tag{2.7d}$$

Equations (2.7) are henceforth designated G equations in order to distinguish them from the conventional Hamiltonian formulation for several reasons: First, (2.7) are manifestly covariant while Hamilton's equations are not. Second, the function $G(x,P)$ of (2.7) is a scalar while the Hamiltonian H, of (2.4), is the time component of the four-momentum. Third, the existence of λ in (2.7a) and (2.7b) means that G equations are not exactly analogous to Hamilton's equations; indeed, they cannot be since setting $\lambda=1$ in (2.7) destroys their self-consistency.[1,3]

Now consider the functional dependence of λ in (2.7). Equations (2.3) and (2.7a) yield[10]

$$\lambda^2=\left(\frac{\partial G}{\partial P}\cdot\frac{\partial G}{\partial P}\right)^{-1}. \tag{2.8}$$

The covariance of the G equations allows only one of the roots of (2.8) for λ. In the particle rest frame, (2.7a) reduce to

$$\lambda=\left(\frac{\partial G}{\partial P_0}\right)^{-1},\quad \left(\frac{\partial G}{\partial P_k}\right)=0.$$

Thus, since (2.8) is valid in all frames, only the positive root is allowed. The scalar λ satisfies

$$\lambda(x,P)=\left(\frac{\partial G}{\partial P}\cdot\frac{\partial G}{\partial P}\right)^{-1/2} \tag{2.9}$$

in any frame. The exclusion of one of the roots in (2.8) is a necessary feature of the G equations, since a degeneracy in λ would imply nonunique particle trajectories in the "phase space" (x,P). From (2.7d), particle trajectories lie on the seven-dimensional surface $G\approx 0$ in (x,P). This surface is henceforth referred to as the G surface.

In Appendix A, it is shown that (2.7) reduce precisely to Hamilton's equations in a specific Lorentz frame. Thus the G equations express Hamilton's equations in tensor form. The relationship between Kalman's equations (2.2) and the G equations (2.7) is analogous to that between Lagrangian and Hamiltonian dynamics: First, (2.2) and (2.7) are the tensorial expression of Lagrange's and Hamilton's equations, respectively. Further, (2.2) are cast in the space (x,U) while the G equations are cast in the phase space (x,P).

III. APPLICATIONS OF KALMAN'S EQUATIONS AND THE *G* EQUATIONS

To illustrate the manifestly covariant formulations of (2.2) and (2.7), consider their application to the following cases of a particle of mass m interacting with various external fields: (a) free particle; (b) particle in a scalar field $\phi(x)$; and (c) particle, charge e, in an electromagnetic field. Although the resulting equations of motion are well known in the literature,[11] the application of the manifestly covariant formulations to these cases serves to demonstrate their relationship to the Lagrange and Hamiltonian formulations.

Kalman's formulation is presented in Table I. In the first column, $f(x,U)$ is the scalar integrand of (2.1). The

TABLE I. Kalman and Lagrange formulations.

	Kalman formulation			Lagrange formulation (in Lorentz frame)		
System	$f(x,U)$	P^μ	Equations of motion	$L(\mathbf{r},\mathbf{v},t)$	P^μ	Equations of motion (3.5)
(a) Free particle	$-m$	mU^μ	$\frac{d}{ds}(mU_\mu)=0$	$\frac{-m}{\gamma}$	$P^0=\gamma m$	$\frac{d}{dt}(\gamma m)=0$
					$\mathbf{P}=\gamma m\mathbf{v}$	$\frac{d}{dt}(\gamma m\mathbf{v})=0$
(b) Particle in scalar field $\phi(x)$	$-m-\phi(x)$	$(m+\phi)U^\mu$	$\frac{d}{ds}[(m+\phi)U_\mu]=\frac{\partial\phi}{\partial x^\mu}$	$\frac{-(m+\phi)}{\gamma}$	$P^0=\gamma(m+\phi)$	$\frac{d}{dt}[\gamma(m+\phi)]=\frac{1}{\gamma}\frac{\partial\phi}{\partial t}$
					$\mathbf{P}=\gamma(m+\phi)\mathbf{v}$	$\frac{d}{dt}[\gamma(m+\phi)\mathbf{v}]=-\frac{1}{\gamma}\nabla\phi$
(c) Particle in electromagnetic field	$-m-eA(x)\cdot U$	$mU^\mu+eA^\mu$	$\frac{d}{ds}(mU_\mu)=eF_{\mu\alpha}U^\alpha$	$\frac{-m}{\gamma}-eA_0+e\mathbf{A}\cdot\mathbf{v}$	$P^0=\gamma m+eA^0$	$\frac{d}{dt}(\gamma m)=-e\left[\frac{\partial\mathbf{A}}{\partial t}\cdot\mathbf{v}+\mathbf{v}\cdot\nabla A^0\right]$
			$F_{\mu\alpha}\equiv\partial_\mu A_\alpha-\partial_\alpha A_\mu$		$\mathbf{P}=\gamma m\mathbf{v}+e\mathbf{A}$	$\frac{d}{dt}(\gamma m\mathbf{v})=-e\nabla A^0-e\frac{\partial\mathbf{A}}{\partial t}+e\mathbf{v}\wedge\operatorname{curl}\mathbf{A}$

second and third columns show P^μ from (2.2b) and the covariant equations of motion (2.2a).

As mentioned in Sec. II, Lagrange's equations are equivalent to Kalman's equations in a specific Lorentz frame. It is interesting to compare (2.2) with Lagrange's equations for cases (a)–(c). The relationship between $f(x,U)$ and the corresponding Lagrangian[12] $L(\mathbf{r},\mathbf{v},t)$ is obtained by equating the action of (2.1) with the expression

$$S=\int L\,dt, \tag{3.1}$$

for a specific frame. Equations (2.1) and (3.1) imply

$$L=f/U_0\equiv f/\gamma, \tag{3.2}$$

where

$$U^0=\gamma=(1-\mathbf{v}\cdot\mathbf{v})^{-1/2} \tag{3.3}$$

in a Lorentz frame. In the Lagrangian formalism, the four-momentum components are

$$P^0=\sum_k \mathbf{v}^k\frac{\partial L}{\partial \mathbf{v}^k}-L,\quad P^k=\frac{\partial L}{\partial \mathbf{v}^k}, \tag{3.4}$$

and the equations of motion are

$$\frac{dP^0}{dt}=\frac{-\partial L}{\partial t},\quad \frac{dP^k}{dt}=\frac{\partial L}{\partial x^k}. \tag{3.5}$$

In the remaining columns of Table I, Eqs. (3.2), (3.4), and (3.5) are presented for comparison with the Kalman formulation. [Note that (3.3) is assumed; hence the Lagrange formulation is developed in a Lorentz frame.] Although the corresponding equations of motion (2.2a) and (3.5) are equivalent, it is clear that in each case the Kalman formulation displays a structural simplicity lacking in the Lagrange formulation.

Table II presents the G formulation of (2.7) for the three cases of interest. In the first column,[8] the scalar $G(x,P)$ is obtained from (2.7c). Next, λ is found from (2.9) with (2.7d), and it turns out that $\lambda=\frac{1}{2}$ in each case. In the third column, (2.7a), (2.7b), and (2.7d) are the covariant G equations of motion which are, of course, equivalent to their counterparts of Table I.

In the remaining columns of Table II, the Hamiltonian formulation is displayed for comparison with the G formulation. First, $H(\mathbf{r},\mathbf{p},t)$ is found[13] from (2.4) with (3.4). Hamilton's equations

$$\frac{dx^k}{dt}=\frac{\partial H}{\partial p^k},\quad \frac{dH}{dt}=\frac{\partial H}{\partial t},\quad -\frac{dp^k}{dt}=\frac{\partial H}{\partial x^k}, \tag{3.6}$$

are written in the final column.

IV. DERIVATION OF THE G EQUATIONS FROM A VARIATIONAL PRINCIPLE

This section aims to derive the G equations directly from a variational principle without reference to the Kalman formulation. (The variational principle is thus analogous to the modified Hamilton's principle,[14] which establishes Hamilton's equations independently of the Lagrange formulation.) Using this method, the G equations are given an independent status and the variational principle is then used in the next section to generate canonical transformations of the G equations.

The action from which Kalman's equations were originally derived satisfies (2.1). From (2.7c), S then takes the form

$$S=\int_a^b\left[G(x,P)-\frac{dx^\mu}{ds}P_\mu\right]ds. \tag{4.1}$$

In (2.7), x^μ are the space-time coordinates and P_μ are the corresponding components of canonical four-momentum. It is convenient to define such an eight-dimensional set (x,P) as a coordinate set. Thus each of the cases in Table II is described in a coordinate set. Since the variational principle in this section is to generate canonical transformations, however, it is important that the (x,P) appearing in (4.1) do not necessarily constitute a coordinate set; the eight variables (x^μ,P_μ) are to be treated here as having no particular physical connotation. Similarly, s in (4.1) is to be treated as an

TABLE II. G and Hamiltonian formulations.

System	$G(x,P)$	G formulation λ	G equations (2.7)	Hamiltonian formulation (in Lorentz frame) $H(\mathbf{r},\mathbf{P},t)$	Hamilton's equations (3.6)
(a) Free particle	$-m+\frac{P\cdot P}{m}$	$\frac{1}{2}$	$P^\mu=mU^\mu$ $\frac{dP_\mu}{ds}=0$	$[m^2+\mathbf{P}\cdot\mathbf{P}]^{1/2}$	$\mathbf{v}=\mathbf{P}/H$ $\frac{dH}{dt}=0$ $\frac{d\mathbf{P}}{dt}=0$
(b) Particle in scalar field $\phi(x)$	$-(m+\phi)+\frac{P\cdot P}{(m+\phi)}$	$\frac{1}{2}$	$P^\mu=(m+\phi)U^\mu$ $\frac{dP_\mu}{ds}=\frac{\partial\phi}{\partial x^\mu}$	$[(m+\phi)^2+\mathbf{P}\cdot\mathbf{P}]^{1/2}$	$\mathbf{v}=\frac{\mathbf{P}}{H}$ $\frac{dH}{dt}=\frac{(m+\phi)}{H}\frac{\partial\phi}{\partial t}$ $\frac{d\mathbf{P}}{dt}=\frac{-(m+\phi)}{H}\nabla\phi$
(c) Particle in electromagnetic field	$-m+\frac{(P-eA)\cdot(P-eA)}{m}$	$\frac{1}{2}$	$P^\mu=mU^\mu+eA^\mu$ $\frac{d}{ds}(mU_\mu)=eF_{\mu\nu}U^\nu$ $F_{\mu\nu}\equiv\partial_\mu A_\nu-\partial_\nu A_\mu$	$eA^0+[m^2+(\mathbf{P}-e\mathbf{A})^2)]^{1/2}$	$\mathbf{v}=\frac{\mathbf{P}-e\mathbf{A}}{(H-eA^0)}$ $\frac{d}{dt}(\gamma m)=-e\left[\frac{\partial\mathbf{A}}{\partial t}\cdot\mathbf{v}+\mathbf{v}\cdot\nabla A^0\right]$ $\frac{d}{dt}(\gamma m\mathbf{v})=-e\nabla A^0-e\frac{\partial\mathbf{A}}{\partial t}+e\mathbf{v}\times\operatorname{curl}\mathbf{A}$

arbitrary scalar parameter; it is not necessarily particle proper time. In the integrand of (4.1), s is merely the means of parametrizing curves in the space (x,P).

Consider now the variational principle itself. In (4.1), the eight variables (x^μ,P_μ) undergo infinitesimal variations $(\delta x^\mu,\delta P_\mu)$ that vanish at the end points of integration but are otherwise arbitrary. These variations induce the first-order variation δS in the action. In the variational principle, δS is set to zero,

$$\delta S = 0, \tag{4.2}$$

and then the weak condition

$$G(x,P) \approx 0 \tag{4.3}$$

is imposed. The weak constraint (4.3) thus reduces to seven the number of independent variations in $(\delta x^\mu,\delta P_\mu)$.

Let the integral of (4.1) be stationary, i.e., satisfy (4.2), on the trajectory $x(s), P(s)$. Then, as in Ref. 3, δS is obtained by comparing (4.1) with a similar integral evaluated over a neighboring curve $x'(s'), P'(s')$. Thus

$$S+\delta S = \int_a^b \left[G(x',P') - \frac{dx'^\mu}{ds'} P'_\mu \right] ds', \tag{4.4}$$

where

$$x'^\mu = x^\mu + \delta x^\mu, \quad P'_\mu = P_\mu + \delta P_\mu, \quad s' = s + \delta s. \tag{4.5}$$

From (4.5), the first-order variation in dx^μ/ds satisfies[15]

$$\delta\left(\frac{dx^\mu}{ds}\right) = \frac{d}{ds}(\delta x^\mu) - \frac{dx^\mu}{ds}\frac{d}{ds}(\delta s). \tag{4.6}$$

Equations (4.1) and (4.4)–(4.6) yield, after a partial integration

$$\delta S = \int_a^b \left[\left(\frac{\partial G}{\partial P_\mu} - \frac{dx^\mu}{ds} \right) \delta P_\mu + \left(\frac{\partial G}{\partial x^\mu} + \frac{dP_\mu}{ds} \right) \delta x^\mu \right] ds + \int_a^b G(x,P) \frac{d(\delta s)}{ds} ds - P_\mu \, \delta x^\mu \Big|_a^b. \tag{4.7}$$

Finally, (4.2), (4.3), and (4.7), with the condition of vanishing δx^μ at the end points, yield the essential G equations

$$\frac{dx^\mu}{ds} = \lambda \frac{\partial G}{\partial P_\mu}, \quad \frac{-dP_\mu}{ds} = \lambda \frac{\partial G}{\partial x^\mu}, \quad G(x,P) \approx 0. \tag{4.8}$$

[Note that λ in (4.8) arises from (4.7) as a Lagrange multiplier.]

Equations (4.8) have the *form* of Eq. (2.7a), (2.7b), and (2.7d). For the choice that s be particle proper time and (x,P) be a coordinate set, (4.8) obviously have the same *content* as the original G equations (2.7) of Sec. II. In that case,

$$ds^2 = \eta_{\mu\nu}\, dx^\mu\, dx^\nu, \tag{4.9}$$

from which (2.3) follows, hence λ in (4.8) also satisfies (2.9). In the general case of noncoordinate sets (x,P), λ does not satisfy (2.9) and the G equations derived here comprise (4.8) only.

In the above variational principle, it is taken that both δx and δP vanish at the end points. In the application of the principle, however, only the condition of vanishing δx is employed [see (4.7)]. The condition on δP is maintained for the following reasons: First, the variables x and P are to be treated in the variational principle in a completely symmetric manner. This is an extension of the conventional treatment, in the modified Hamilton's principle, which places the spatial coordinates and three-momenta on the same footing. Second, the condition that δP vanish at the end points implies that the functional

$$S_1 \equiv S + \int_a^b \frac{dF}{ds}(x,P) ds \tag{4.10}$$

also generates, via the above variational principle, Eqs. (4.8). Thus S_1 and S are equivalent. This result is employed in the following section to generate canonical transformations of the G equations (4.8).

V. CANONICAL TRANSFORMATIONS OF THE G EQUATIONS

Suppose the G equations (4.8) hold for the variables (x,P). The transformations

$$x' = x'(x,P), \quad P' = P'(x,P), \tag{5.1}$$

are defined here to be canonical transformations (CT's) if, in the variables (x', P'), the equations

$$\frac{dx'^\mu}{ds} = \lambda' \frac{\partial G'}{\partial P'_\mu}, \quad \frac{-dP'_\mu}{ds} = \lambda' \frac{\partial G'}{\partial x'^\mu}, \quad G'(x',P') \approx 0, \tag{5.2}$$

hold. Equations (5.2) are designated the G' equations, and the definition of CT here is an obvious generalization of the conventional CT's of Hamilton's equations.

Consider the variational principle of Sec. IV with the following action functional:

$$S = \int_a^b \left[G(x,P) - \frac{dx^\mu}{ds} P_\mu \right] ds = \int_a^b \left[G'(x',P') - \frac{dx'^\mu}{ds} P'_\mu \right] ds - \int_a^b \frac{dF}{ds}(x,P) ds. \tag{5.3}$$

Equations (4.1) and (4.10) show that the action of (5.3) generates both the G equations (4.8) and the G' equations (5.2) provided the weak conditions

$$G(x,P) \approx 0, \quad G'(x',P') \approx 0 \tag{5.4}$$

are imposed. The theory of CT's of the G equations can now be developed by considering (5.3) before imposition of (5.4). Then, the eight variables (x,P) are treated as being independent. Once the transformations of the form (5.1) are found, the constraints (5.4) are imposed and the transformations are then automatically canonical. Geometrically, this means that the transformations for independent (x,P) and (x',P') are first found, and then they are constrained to lie on the G and G' surfaces defined in (5.4).

From (5.3), the integrands satisfy

$$\frac{dx}{ds} \cdot P - G = \frac{dx'}{ds} \cdot P' - G' + \frac{dF}{ds}. \tag{5.5}$$

The analysis now proceeds as in the standard treatment[14] of CT's of Hamilton's equations. One may choose which eight of the 16 variables (x, P, x', P') appearing in (5.1) are the

independent set. First, suppose x and x' are the independent variables, then F in (5.5) may be considered as a function of x and x', and is the generator of the required transformations.

A. Generator $F=F_1(x,x')$

Equation (5.5) yields

$$P_\mu = \frac{\partial F_1}{\partial x^\mu}, \quad P'_\mu = \frac{-\partial F_1}{\partial x'^\mu}, \quad G = G'. \tag{5.6}$$

The final step in obtaining CT's of the G equations is to impose (5.4) (a condition that is, of course, absent from the conventional CT's of Hamilton's equations). The required CT's are then the first and second of (5.6) evaluated on the G surface (i.e., $G \approx 0$) which, from the last of (5.6), is identical with the G' surface.

It is interesting to note that CT's prohibit explicit s dependence in the transformations (5.1). Suppose the generator F in (5.3) were s dependent, i.e., $F = F(x,P,s)$. In that case,[16] the first and second of (5.6) remain unchanged, and the last of (5.6) is modified to

$$G' = G + \frac{\partial F_1}{\partial s}.$$

The constraint (5.4), however, then implies

$$\frac{\partial F_1}{\partial s} = 0;$$

hence explicit s dependence is excluded from CT's.

In Secs. V B–V D below, further archetypes of CT's are obtained by considering, before imposition of (5.4), the independent-variable sets to be (x,P'), (P,x'), and (P,P'), respectively.

B. Generator $F=F_2(x,P')-x'\cdot P'$

Equation (5.5) yields

$$P_\mu = \frac{\partial F_2}{\partial x^\mu}, \quad x'^\mu = \frac{\partial F_2}{\partial P'_\mu}, \quad G = G'. \tag{5.7}$$

C. Generator $F=F_3(P,x')+x\cdot P$

Equation (5.5) yields

$$x^\mu = \frac{-\partial F_3}{\partial P_\mu}, \quad P'_\mu = \frac{-\partial F_3}{\partial x'^\mu}, \quad G = G'. \tag{5.8}$$

D. Generator $F=F_4(P,P')+x\cdot P-x'\cdot P'$

Equation (5.5) yields

$$x^\mu = \frac{-\partial F_4}{\partial P_\mu}, \quad x'^\mu = \frac{\partial F_4}{\partial P'_\mu}, \quad G = G'. \tag{5.9}$$

As in case A, the transformation equations of (5.7)–(5.9) become CT's on imposition of (5.4). Thus, in each of (5.6)–(5.9), all partial derivatives are evaluated on the G surface (which is identical with the G' surface). Since the generator F cannot have explicit s dependence, the CT's necessarily have the form of (5.1).

For an example of a covariant CT, consider a four-dimensional point transformation, i.e., a coordinate transformation. Let (x,P) be a coordinate set and consider the CT of type B with

$$F_2(x,P') = f^\mu(x)\, P'_\mu. \tag{5.10}$$

Equations (5.7) yield for the CT

$$\begin{aligned} x'^\mu &= f^\mu(x) \equiv x'^\mu(x), \\ P_\mu &= \frac{\partial f^\alpha}{\partial x^\mu} P'_\alpha = \frac{\partial x'^\alpha}{\partial x^\mu} P'_\alpha. \end{aligned} \tag{5.11}$$

Clearly, (5.11) display a space-time coordinate transformation and the corresponding transformation of covariant components of four-momentum.

In Appendix B, it is shown that λ, the multiplier in the G equations (4.8), is invariant under CT's, i.e.,

$$\lambda = \lambda'. \tag{5.12}$$

This result is consistent with the statement in Sec. II that λ is a scalar under coordinate transformations; the latter, with generator of the form (5.10), are a subgroup of the group of all CT's.

In Appendix C, it is shown that if G is a function of P only, it is possible to generate a CT such that $G' = P'_0$. (This result is used in the next section.)

VI. DEVELOPMENT OF A MANIFESTLY COVARIANT HAMILTON–JACOBI SOLUTION TO THE G EQUATIONS

Analogous to the Hamiltonian formalism, the Hamilton–Jacobi method of solution, based on canonical transformations, is now developed for the G formalism. The aim is to generate solutions $x(s)$, $P(s)$ of the G equations (4.8) as CT's of the form

$$x^\mu(s) = x^\mu(x',P'), \quad P_\mu(s) = P_\mu(x',P'), \tag{6.1}$$

where (x',P') are functions of s and the initial values of (x,P).

It is to be noted that CT's of the G equations cannot be explicitly s dependent, as shown in Sec. V. Thus, since s dependence on the right-hand side of (6.1) arises only from x' and P', it follows that one cannot choose all the x'^μ and P'_μ to be constant. Instead, a transformation is constructed for which seven of the eight variables (x'^μ, P'_μ) are constant.

To begin, suppose there exists a CT of the form

$$(x,P) \to (x'',P'')$$

such that G'' is a function of P'' only, i.e., P'' is constant. In that case, as shown in Appendix C, there exists a further CT

$$(x'',P'') \to (x',P')$$

such that $G' = P'_0$. Consider now the direct CT

$$(x,P) \to (x',P')$$

with generator type B of Sec. V. Equations (5.7) then yield (after dropping the subscript from F_2)

$$F = F(x,P'), \tag{6.2a}$$

$$P_\mu = \frac{\partial F}{\partial x^\mu}, \tag{6.2b}$$

$$x'^\mu = \frac{\partial F}{\partial P'_\mu}, \tag{6.2c}$$

$$G(x,P) = G' = P'_0 . \tag{6.2d}$$

From (6.2d), the G' equations [i.e., (5.2)] integrate to

$$x'^{\mu} = \delta^{\mu 0} \int \lambda \, ds + \alpha^{\mu} , \quad P'_{\mu} = \beta_{\mu} , \tag{6.3}$$

where α^{μ}, β_{μ} are constants (and $\lambda = \lambda'$, as shown in Appendix B).

Equations (6.3) show that s dependence of the right-hand side of the required CT's (6.1) arises entirely from x'^0. The existence of the transformations is guaranteed by the existence of the generator $F(x,P)$ in (6.2). The generator itself is the solution of (6.2d) with (6.2b) and the last of (6.3):

$$G\left(x, \frac{\partial F}{\partial x}\right) = \beta_0 . \tag{6.4}$$

Equation (6.4) is the manifestly covariant version of the Hamilton–Jacobi equation.

The solutions to the G equations are contained in (6.2b) and (6.2c) which are, of course, evaluated on the G surface

$$G = \beta_0 \approx 0 . \tag{6.5}$$

Inversion of (6.2c) yield the first of (6.1) which in turn, with (6.2b), yield the remaining solutions of (6.1).

To illustrate the Hamilton–Jacobi method, the G equations are solved for several simple systems with one spatial dimension. In the following examples, the G equations are cast in a coordinate set with

$$x^0 = t, \quad x^1 = x . \tag{6.6}$$

To begin, a study of the free particle reveals the essential details of the covariant Hamilton–Jacobi method.

A. Free particle

From Table II,

$$G(x,P) = -m + (P \cdot P)/m , \quad \lambda = \tfrac{1}{2} , \tag{6.7}$$

and the covariant Hamilton–Jacobi equation (6.4) takes the form

$$\left(\frac{\partial F}{\partial t}\right)^2 - \left(\frac{\partial F}{\partial x}\right)^2 = m(m + \beta_0) . \tag{6.8}$$

Thus the generator F [satisfying (6.2a) with the second of (6.3)] is given by

$$F = \beta_1 t \pm x[(\beta_1)^2 - m(m + \beta_0)]^{1/2} . \tag{6.9}$$

The solutions to the G equations are finally obtained from (6.9) with the CT's of (6.2b) and (6.2c), evaluated on the G surface (6.5). Equations (6.2b) yield

$$P_0 = \beta_1, \quad P_1 = \pm [(\beta_1)^2 - m^2]^{1/2} . \tag{6.10}$$

Equations (6.2c) with (6.3) (and $\lambda = \frac{1}{2}$) yield

$$\begin{aligned} \frac{dx}{ds} &= U^1 = \mp \frac{[(\beta_1)^2 - m^2]^{1/2}}{m} , \\ \frac{dx}{dt} &= v = \mp \frac{[(\beta_1)^2 - m^2]^{1/2}}{\beta_1} . \end{aligned} \tag{6.11}$$

[Note that signs $\pm$ in the generator equation (6.9) lead to $\mp$ velocities.]

From (6.11), U^0 of (3.3) satisfies

$$U^0 = \gamma = \beta_1 / m , \tag{6.12}$$

hence (6.10) take on the familiar free-particle form

$$\begin{aligned} P_0 &= m U_0 = \beta_1 = \text{const} , \\ P_1 &= m U_1 = -\gamma m v = \text{const} . \end{aligned} \tag{6.13}$$

B. Particle in scalar field $\phi(x)$

In this case, Table II shows that

$$G(x,P) = -(m + \phi) + (P \cdot P)/(m + \phi) , \quad \lambda = \tfrac{1}{2} . \tag{6.14}$$

Equation (6.4) for the one-dimensional system is

$$\left(\frac{\partial F}{\partial t}\right)^2 - \left(\frac{\partial F}{\partial x}\right)^2 = (m + \phi)(m + \phi + \beta_0) . \tag{6.15}$$

Clearly, the solution for F in (6.15) depends on the form of the potential $\phi(x,t)$.

C. Special case of B: Time-independent potential

Suppose the scalar field ϕ, of example B above, depends on $x^1 = x$ only, in some frame Σ:

$$\phi = \phi(x) \quad \text{in } \Sigma . \tag{6.16}$$

In that case, a solution (in Σ) for F of (6.15) is the separable form

$$\begin{aligned} &F(t,x,\beta_0,\beta_1) \\ &\quad = \beta_1 t \pm \int dx [(\beta_1)^2 - (m + \phi)(m + \phi + \beta_0)]^{1/2} . \end{aligned} \tag{6.17}$$

The solutions to the G equations, given by (6.17) with (6.2b), (6.2c), and (6.3), are

$$\begin{aligned} &P_0 = H = \beta_1 , \quad P_1 = -\gamma v (m + \phi) , \\ &\gamma = U^0 = \beta_1 / (m + \phi) , \\ &v = \frac{dx}{dt} = \mp \frac{[(\beta_1)^2 - (m + \phi)^2]^{1/2}}{\beta_1} . \end{aligned} \tag{6.18}$$

It is to be noted that the first three of (6.18) yield components P_{μ} identical with those of Table II, as required. Furthermore, direct substitution into the appropriate G equations of Table II shows that (6.18) are indeed the required solutions.

In principle, the trajectory $x(t)$ may be found by integration of the last of (6.18) subject to the auxiliary condition on the sign of H:

$$H = \beta_1 \geqslant 0 . \tag{6.19}$$

In the next example another special case of scalar potential is considered, for which the trajectory has a sinusoidal form.

D. Special case of C: Linear potential

Suppose ϕ in (6.16) is the simple linear form[17]

$$\phi = ax \quad \text{in } \Sigma , \tag{6.20}$$

where a is a positive constant. The particle trajectory in Σ, found from the first and last of (6.18), is

$$x(t) = [\beta_1 \sin \omega (t - t_0) - m]/a , \tag{6.21}$$

where

$$\omega = a/\beta_1 , \tag{6.22}$$

and t_0 is a constant. Thus the linear potential of (6.20) gives rise to the simple-harmonic trajectory (6.21).[18]

The particle velocity, given by differentiation of (6.21) [or by the last of (6.18) with (6.21)], satisfies

$$\frac{dx}{dt} = \cos\omega(t - t_0) . \tag{6.23}$$

This result is rather surprising; the particle attains precisely luminal velocities ($|v| = 1$) at times

$$(t - t_0) = 0, \pm (T/2), T, \pm (3/2)T, \pm 2T,...,$$
$$T = 2\pi/\omega = 2\pi\beta_1/a. \tag{6.24}$$

Note that this does not indicate any singular behavior of the system. The total energy of the system is, from the first of (6.18), constant throughout the motion.

VII. FIRST QUANTIZATION OF THE *G* EQUATIONS

This section initiates a heuristic approach to the problem of quantizing the G equations of motion for a particle. In a space-time coordinate representation, the (scalar) wave function ψ depends on (x), and four-momentum components are represented by operators

$$P^\alpha = i\hbar\partial^\alpha \equiv i\hbar\eta^{\alpha\beta}\partial_\beta . \tag{7.1}$$

The central assumption here is that the classical equation (2.7d) goes over in the quantum-mechanical case to

$$G(x,P)\psi(x) = 0, \tag{7.2}$$

where the four-momentum P satisfies (7.1).

In the following calculations, (7.2) is considered for the three cases of Table II; the operator $G(x,P)$ is obtained from the classical G function by the correspondence principle. It is shown that in each case, (7.2) yields a physically realistic model. Thus Eq. (7.2) may represent a viable starting point for the first quantization of the G equations.

A. Free particle

From Table II, the classical G function is

$$G = -m + (P\cdot P)/m, \tag{7.3}$$

hence (7.2) with (7.1) yield the Klein–Gordon equation

$$(-m^2 + \hbar^2\Box)\psi = 0, \tag{7.4}$$

where $\Box$ is the Dalembertian

$$\Box \equiv -\eta^{\alpha\beta}\partial^2_{\alpha\beta} = \nabla^2 - \frac{\partial^2}{\partial t^2} . \tag{7.5}$$

It is to be noted that (7.4) arises from (7.2) and (7.3) *without ambiguities*. In the conventional Hamiltonian approach, the square root in the form of H (see Table II) gives rise to a formally ill-defined operator

$$H = [m^2 - \hbar^2\nabla^2]^{1/2}.$$

Then, (7.4) is usually obtained in the form $[H^2 + \hbar^2\nabla^2 - m^2]\Psi = 0$. Thus ambiguities arising from the interpretation of the square root in the form of the Hamiltonian are completely avoided in the corresponding G formalism.[19]

B. Particle in scalar field $\phi(x)$

From Table II, G satisfies

$$G = -(m + \phi) + (P\cdot P)/(m + \phi), \tag{7.6}$$

hence (7.2) with (7.1) and (7.5) yield

$$[-(m + \phi)^2 + \hbar^2\Box]\psi = 0. \tag{7.7}$$

It is instructive here to consider the nonrelativistic (NR) limit of (7.7). This is obtained by first defining[20]

$$\psi \equiv \psi_s\, e^{(i/\hbar)mt}, \tag{7.8}$$

in terms of which (7.7) take the exact form

$$(2m\phi + \phi^2)\psi_s - \hbar^2\nabla^2\psi_s$$
$$- 2mi\hbar\frac{\partial\psi_s}{\partial t} + \hbar^2\frac{\partial^2\psi_s}{\partial t^2} = 0. \tag{7.9}$$

In the NR limit, ϕ is small and $\partial\psi_s/\partial t$ slowly varying, i.e.,

$$|\phi| \ll m, \quad \hbar\left|\frac{\partial^2\psi_s}{\partial t^2}\right| \ll m\left|\frac{\partial\psi_s}{\partial t}\right| . \tag{7.10}$$

Finally, (7.9) with the NR approximations of (7.10) yield the NR Schrödinger equation:

$$\frac{-\hbar^2}{2m}\nabla^2\psi_s + \phi\psi_s \simeq i\hbar\frac{\partial\psi_s}{\partial t} . \tag{7.11}$$

C. Particle in electromagnetic field

From Table II, G satisfies

$$G = -m + [(P - eA)\cdot(P - eA)]/m; \tag{7.12}$$

hence (7.2) yields the Klein–Gordon equation for a particle in an electromagnetic field[21]:

$$[-m^2 + (i\hbar\partial^\mu - eA^\mu)(i\hbar\partial_\mu - eA_\mu)]\psi = 0. \tag{7.13}$$

This equation is reduced by (7.1) with (7.8) to the exact form

$$\left[(i\hbar\nabla - e\mathbf{A})^2 + i\hbar e\frac{\partial A^0}{\partial t} + 2emA^0 - (eA^0)^2\right]\psi_s$$
$$+ (2i\hbar eA^0 - 2mi\hbar)\frac{\partial\psi_s}{\partial t} + \hbar^2\frac{\partial^2\psi_s}{\partial t^2} = 0. \tag{7.14}$$

The NR case of (7.14) is obtained as the limit of small and slowly varying A^0 with slowly varying $\partial\psi_s/\partial t$, i.e.,

$$|eA^0| \ll m|A^0|, \quad \hbar\left|\frac{\partial A^0}{\partial t}\right| \ll m|A^0|,$$
$$\hbar\left|\frac{\partial^2\psi_s}{\partial t^2}\right| \ll m\left|\frac{\partial\psi_s}{\partial t}\right| . \tag{7.15}$$

Finally, (7.15) reduces (7.14) to the NR Schrödinger equation for a particle in an electromagnetic field:

$$\left[\frac{1}{2m}(-i\hbar\nabla - e\mathbf{A})^2 + eA^0\right]\psi_s \simeq i\hbar\frac{\partial\psi_s}{\partial t} . \tag{7.16}$$

In each case of examples A–C above, (7.2) leads to physically realistic wave equations. It is now conjectured that (7.2) provides a viable foundation for the first quantization of the G equations. The formalism presented in this section is incomplete, since a full theory of the quantized G equations would have to contend with several nontrivial problems: (i) Derivation of equations (2.7a), (2.7b), and (2.7d) as a clas-

sical limit via a covariant Ehrenfest theorem; and (ii) construction of self-consistent inner products of states. For the problem of a Klein–Gordon particle in an external electromagnetic field, the conventional nonrelativistic inner product

$$\langle\psi,\psi\rangle = \int |\psi|^2 \, d^3\mathbf{r}$$

leads to a violation of the condition of orthogonality of eigenstates of H with differing energies.[22]

VIII. CONCLUSIONS

The G equations (2.7) and Kalman's equations (2.2) are the tensorial formulation of Hamilton's and Lagrange's equations, respectively. This is shown explicitly by casting the equations in a specific Lorentz frame. Manifestly covariant versions of canonical transformation theory and Hamilton–Jacobi theory are developed for the G equations. Finally, an outline of a covariant first quantization of the G equations is given in Sec. VII.

ACKNOWLEDGMENTS

For helpful discussion and criticism, the author thanks Professor K. Dunn (Mathematics Department, Dalhousie University), Professor G. Miller (Mathematics Department, University of Victoria), and the anonymous referee.

APPENDIX A: REDUCTION OF THE G EQUATIONS (2.7) TO HAMILTON'S EQUATIONS IN A SPECIFIC LORENTZ FRAME

Since (2.7) are defined in a coordinate set (see Sec. III), Eq. (3.3) for U^0 casts (2.7a) and (2.7b) into the form

$$\begin{aligned} &\gamma = \lambda \frac{\partial G}{\partial P_0}, \quad \gamma \frac{dx^k}{dt} = \lambda \frac{\partial G}{\partial P_k}, \\ &-\gamma \frac{dP_0}{dt} = \lambda \frac{\partial G}{\partial t}, \quad -\gamma \frac{dP_k}{dt} = \lambda \frac{\partial G}{\partial x^k}. \end{aligned} \tag{A1}$$

The first of (A1) with (2.4) reduces the remaining equations of (A1) to

$$\begin{aligned} &\frac{dx^k}{dt} = \frac{\partial G}{\partial P_k}\left(\frac{\partial G}{\partial P_0}\right)^{-1}, \\ &\frac{dP_0}{dt} = \frac{dH}{dt} = -\frac{\partial G}{\partial t}\left(\frac{\partial G}{\partial P_0}\right)^{-1}, \\ &\frac{dP^k}{dt} = \frac{\partial G}{\partial x^k}\left(\frac{\partial G}{\partial P_0}\right)^{-1}. \end{aligned} \tag{A2}$$

Since the constraint (2.7d) is weak, the partial derivatives in (A2) are found for independent (x,P) before evaluation on the G surface. Thus, before imposition of (2.7d), G is treated as a function of eight independent variables (x,P).

A consideration of the differential,[23] dG, of a function G of N independent variables (ξ_n) $[n = 0,1,\ldots,(N-1)]$ yields the relations

$$\frac{\partial G}{\partial \xi_n} = \left(\frac{\partial \xi_n}{\partial G}\right)^{-1}, \quad \frac{\partial G}{\partial \xi_m}\frac{\partial \xi_m}{\partial \xi_n}\frac{\partial \xi_n}{\partial G} = -1, \tag{A3}$$

for all fixed m,n where $m \neq n$. Thus (A3) with $n = 0$ result in

$$\frac{\partial G}{\partial \xi_m}\left(\frac{\partial G}{\partial \xi_0}\right)^{-1} = -\frac{\partial \xi_0}{\partial \xi_m}. \tag{A4}$$

Finally, application of (A4) [with $G = G(x,P)$] and (2.4) to (A2) yields Hamilton's equations (3.6).

APPENDIX B: INVARIANCE OF λ UNDER CANONICAL TRANSFORMATIONS

It is shown below that λ of (4.8) remains invariant under a CT of type A, Sec. V. The relevant transformations are given by (5.6) (and written here without the subscript on F_1):

$$\begin{aligned} &P_\mu = \frac{\partial F}{\partial x^\mu} \equiv F_\mu, \quad P'_\mu = \frac{-\partial F}{\partial x'^\mu} \equiv -F_{\mu'}, \\ &G = G'. \end{aligned} \tag{B1}$$

Equations (4.8), (5.2), and the last of (B1) yield

$$\begin{aligned} \frac{dP_\mu}{ds} &= -\lambda \frac{\partial G}{\partial x^\mu} = -\lambda \frac{\partial G'}{\partial x^\mu} \\ &= \frac{\lambda}{\lambda'}\left[\frac{dP'_\alpha}{ds}\left(\frac{\partial x'^\alpha}{\partial x^\mu}\right)_p - \frac{dx'^\alpha}{ds}\left(\frac{\partial P'_\alpha}{\partial x^\mu}\right)_p\right]. \end{aligned} \tag{B2}$$

On the right-hand side of (B2), dP'_α/ds is given by the second of (B1):

$$\frac{dP'_\alpha}{ds} = -F_{\alpha'\beta}\frac{dx^\beta}{ds} - F_{\alpha'\beta'}\frac{dx'^\beta}{ds}. \tag{B3}$$

Further, (B2) for fixed P yields

$$\begin{aligned} &\left(\frac{\partial P'_\alpha}{\partial x^\mu}\right)_p = -F_{\alpha'\mu} - F_{\alpha'\beta'}\left(\frac{\partial x'^\beta}{\partial x^\mu}\right)_p, \\ &F_{\mu\alpha} + F_{\mu\beta'}\left(\frac{\partial x'^\beta}{\partial x^\alpha}\right)_p = 0. \end{aligned} \tag{B4}$$

Finally, (B3) and (B4) reduce (B2) to

$$\frac{dP_\mu}{ds} = \frac{\lambda}{\lambda'}\left[F_{\mu\alpha}\frac{dx^\alpha}{ds} + F_{\mu\alpha'}\frac{dx'^\alpha}{ds}\right]. \tag{B5}$$

However, the first of (B1) yields

$$\frac{dP_\mu}{ds} = F_{\mu\alpha}\frac{dx^\alpha}{ds} + F_{\mu\alpha'}\frac{dx'^\alpha}{ds}. \tag{B6}$$

A comparison of (B5) with (B6) shows that

$$\lambda = \lambda', \tag{B7}$$

under the CT of (B1). Since any CT can, in principle, be reduced to the form of (5.6), Eq. (B7) holds for all CT's.

APPENDIX C: SPECIAL CASE OF THE x-INDEPENDENT G FUNCTION

Suppose G is a function of P only. Consider then a CT of type B, Sec. V, with generator of the form

$$F_2(x,P') = \Pi_\alpha(P')x^\alpha. \tag{C1}$$

Equation (5.7) yield

$$P_\mu = \Pi_\mu(P'), \quad x'^\mu = \frac{\partial \Pi_\alpha}{\partial P'_\mu}x^\alpha, \quad G(P) = G'. \tag{C2}$$

Inversion of the first of (C2) yields

$$P'_\mu = \rho_\mu(P). \tag{C3}$$

Thus the choice

$$\rho_0(P) = G(P), \tag{C4}$$

reduces the last of (C2) to

$$G(P) = G' = P'_0 . \tag{C5}$$

[1]G. Kalman, Phys. Rev. **123**, 384 (1961).

[2]A. Barut, *Electrodynamics and Classical Theory of Fields and Particles* (MacMillan, New York, 1964); H. Arzeliès, *Relativistic Point Dynamics* (Pergamon, Oxford, 1972).

[3]P. H. Lim, J. Math. Phys. **23**, 1641 (1982).

[4]P. A. M. Dirac, Can. J. Math. **2**, 129 (1950); Proc. R. Soc. London Ser. A **246**, 326 (1958); *Lectures on Quantum Mechanics* (Belfer Graduate School, Yeshiva University, New York, 1964).

[5]J. L. Anderson and P. G. Bergmann, Phys. Rev. **83**, 1018 (1951); P. G. Bergmann, in *Encyclopedia of Physics*, edited by S. Flügge (Springer, Berlin, 1962), Vol. IV.

[6]For a review, see K. Sundermeyer, *Constrained Dynamics, Lecture Notes in Physics*, Vol. 169 (Springer, Berlin, 1982); D. Dominici, G. Longhi, J. Gomis, and J. M. Pons, J. Math. Phys. **25**, 2439 (1984); W. R. Davis, *Classical Fields, Particles and the Theory of Relativity* (Gordon and Breach, New York, 1970).

[7]The particle equations of motion are therefore constrained by the space-time geometry. For a general-relativistic variational principle in which the reverse occurs, i.e., geometry is constrained by the dynamics of matter, see P. H. Lim, Phys. Rev. D **27**, 719 (1983).

[8]The notation employed is that of Ref. 3, with the sole exception of indices. Here, Greek (space-time) indices run from 0 to 3 while Latin (spatial) indices run from 1 to 3. The Minkowski metric η has signature $(+---)$, and scalar products $A_\mu B^\mu$ are henceforth written in the form $A\cdot B$.

[9]$G(x,P)$ vanishes by virtue of (2.2b) with (2.3).

[10]In (2.8), $(\partial G/\partial P)\cdot(\partial G/\partial P) \equiv \eta_{\alpha\beta}(\partial G/\partial P_\alpha)(\partial G/\partial P_\beta)$.

[11]Case (b) is considered in Barut, Ref. 2, and (c) is studied in I. T. Todorov, *Relativistic Action at a Distance, Lecture Notes in Physics*, Vol. 162, edited by J. Llosa, (Springer, Berlin, 1982), pp. 213–263; L. D. Landau and E. M. Lifshitz, *The Classical Theory of Fields* (Pergamon, Oxford, 1975), 4th ed.

[12]Conventional three-vector notation is employed; $\mathbf{r} = (x^k)$, $\mathbf{v} = (dx^k/dt)$.

[13]Alternatively, H could be found directly from the Kalman formalism [(2.4) with (2.2b)] or from the G formalism [(2.4) with (2.7d)].

[14]H. Goldstein, *Classical Mechanics* (Addison–Wesley, Reading, MA, 1980).

[15]As in Ref. 3 [Eq. (4)], this relation is independent of any specific physical interpretation of ds in (4.1).

[16]It is assumed that S and S_1 of (4.10) remain equivalent for the modified function $F(x,P,s)$. Otherwise, the transformations generated by F are no longer canonical since the action of (5.3) would generate the G equations but not the G' equations. This situation arises if the scalar s is path dependent in the space (x,P).

[17]An explicit example of (6.20) is $\phi = A\cdot x$, where A is a space-like vector.

[18]Note that (6.21) yields the required nonrelativistic motion [i.e., accelerations $\ddot{x} = -a/m$ for $(t-t_0) = \pm T/4, \pm 5T/4,\ldots$ and $\ddot{x} = a/m$ for $(t-t_0) = \pm 3T/4, \pm 7T/4,\ldots$] for small velocities.

[19]This remark applies to each of the cases in Table II.

[20]J. J. Sakurai, *Advanced Quantum Mechanics* (Addison–Wesley, Reading, MA, 1978).

[21]H. Feschbach, and F. Villars, Rev. Mod. Phys. **30**, 24 (1958).

[22]A review of the history of this problem is given in V. A. Rizov, H. Sazdjian, and T. T. Todorov, Ann. Phys. (NY) **165**, 59 (1985).

[23]M. H. Protter, and C. B. Morrey, *Modern Mathematical Analysis* (Addison–Wesley, Reading, MA, 1964).

The quadratic-Hamiltonian theorem in infinite dimensions

Rudolf Schmid
Department of Mathematics and Computer Science, Emory University, Atlanta, Georgia 30322

(Received 18 February 1988; accepted for publication 27 April 1988)

It is shown that a smooth diffeomorphism on a symplectic Banach space is canonical if it is canonoid with respect to all dynamical systems whose Hamiltonian functions are quadratic.

I. INTRODUCTION

In classical mechanics a dynamical system on phase space is called Hamiltonian if there exists a coordinate system $(q^1,q^2,...,q^N,p_1,p_2,...,p_N)$ in which its evolution equations can be written in the form of Hamilton's canonical equations, i.e., if there exists a Hamiltonian function $H(q^i,p_i)$ such that $dq^i/dt = \partial H/\partial p_i$, $dp_i/dt = -\partial H/\partial q^i$, $i = 1,2,...,N$. A transformation on phase space is then said to be canonical if it maps every Hamiltonian system into a Hamiltonian one; if it maps only some Hamiltonian systems into Hamiltonian ones the transformation is called canonoid. It has been shown in Currie and Saletan[1] that it is sufficient for a transformation to be canonical to map Hamiltonian systems with quadratic Hamiltonian functions into Hamiltonian ones. In Marmo, Saletan, Schmid, and Simoni[2] we gave an intrinsic proof of this finite-dimensional quadratic-Hamiltonian theorem and found a similar result for infinite-dimensional linear Hamiltonian systems. In this paper these results are generalized to nonlinear infinite-dimensional Hamiltonian systems on Banach spaces.

II. THEORY

Let V be a real Banach space and Ω: $V\times V\to\mathbb{R}$ be a continuous bilinear map. Then Ω is said to be *weakly nondegenerate* if $\Omega(v_1,v_2) = 0$ for all $v_2\in V$ implies $v_1 = 0$. In terms of the induced continuous linear mapping Ω^b: $V\to V^*$, defined by $\Omega^b(v_1)(v_2) = \Omega(v_1,v_2)$, $v_1,v_2\in V$, weak nondegeneracy of Ω is equivalent to injectivity of Ω^b. The map Ω is said to be *nondegenerate* if Ω^b is an isomorphism. In most of the infinite-dimensional examples Ω will be only weakly nondegenerate.

A *symplectic form* Ω on a Banach space V is a weakly nondegenerate skew symmetric bilinear map; and (V,Ω) is called a *symplectic Banach space*. Let (V,Ω) and (W,Ξ) be symplectic Banach spaces. A smooth map f: $(V,\Omega)\to(W,\Xi)$ is *symplectic* (or a *canonical transformation*) if $f^*\Xi = \Omega$, that is if $\Omega(v_1,v_2) = \Xi(Df(x)v_1, Df(x)v_2)$ for all $x,v_1,v_2\in V$. A smooth vector field X: $(V,\Omega)\to(V,\Omega)$, is *Hamiltonian* if there exists a C^1 function H: $V\to\mathbb{R}$ such that $\Omega^b(X(v)) = dH(v)$, in which case we write $X = X_H$. Equivalently, using the inner product notation [which is defined the same way as in finite dimensions, namely $i_X\Omega(Y)\equiv\Omega(X,Y)$, for all vector fields X, Y, Abraham, Marsden, and Ratiu[3] and Schmid[4]] the condition is $i_{X_H}\Omega = dH$. Using the identity $dH(tv)\cdot v = \Omega(X_H(tv),v)$, $v\in V$, $t\in\mathbb{R}$, we can express H in terms of X_H and Ω by

$$H(v) - H(0) = \int_0^1 \frac{dH(tv)}{dt}\,dt = \int_0^1 dH(tv)\cdot v\,dt = \int_0^1 \Omega(X_H(tv),v)dt.$$

Note that unlike the situation in finite dimensions, in infinite dimensions not every Hamiltonian function H will have a Hamiltonian vector field X_H.

Let X: $(V,\Omega)\to(V,\Omega)$ be a linear vector filed. Then X is Hamiltonian if and only if X is Ω skew; that is, if $\Omega(Xv_1,v_2) = -\Omega(v_1,Xv_2)$ for all $v_1,v_2\in V$. Furthermore, in that case one can take $H(v) = \frac{1}{2}\Omega(Xv,v)$. Similarly for the nonlinear case; if X: $V\to V$ is a smooth vector field, then $X = X_H$ for some H: $V\to\mathbb{R}$ if and only if $DX(x)$ is Ω skew for all x; in which case

$$H(v) = \int_0^1 \Omega(X(tv),v)dt + \text{const.}$$

The important property of canonical diffeomorphisms is that they map Hamiltonian vector fields into Hamiltonian ones, i.e., a diffeomorphism f: $(V,\Omega)\to(V,\Omega)$ is a canonical transformation if for all H: $V\to\mathbb{R}$, such that X_H exists, we have $f^*X_H = X_{H\circ f}$. Indeed if X_H exists we have $f^*(dH) = f^*i_{X_H}\Omega = i_{f^*X_H}f^*\Omega$ and $f^*dH = d(f^*H) = i_{X_{f^*H}}\Omega$. On the other hand, there exist transformations that map only some Hamiltonian vector fields into Hamiltonian ones; such transformations have been called *canonoid* (Saletan and Cromer[5]) with respect to the vector fields it maps this way. The *canonical* transformations are then those canonoid with respect to *all* vector fields. For finite-dimensional vector spaces it has been shown in Currie and Saletan[1] that this definition is stronger than need be; that is, that there exists a subset (called a *sufficiency subset*) of Hamiltonian vector fields such that if a transformation is canonoid with respect to the vector fields in this sufficiency subset, then it is canonoid with respect to all Hamiltonian vector fields, i.e., is canonical. The sufficiency subset found in Currie and Saletan[1] consists of Hamiltonian vector fields whose Hamiltonian functions are quadratic in a set of coordinates on phase space. We call this result the *finite-dimensional quadratic-Hamiltonian theorem*. In Marmo, Saletan, Schmid, and Simoni[2] we proved the analog of the finite-dimensional quadratic-Hamiltonian theorem for linear diffeomorphisms in infinite dimensions. Here we generalize this theorem to nonlinear diffeomorphisms on infinite-dimensional vector spaces.

Theorem: Let f: $(V,\Omega)\to(V,\Omega)$ be a smooth diffeomorphism such that $Df(x)$ is Ω skew for all x. If $f^*X_H = X_{f^*H}$

 0022-2488/88/092010-02$02.50

for all functions H: $V \to \mathbf{R}$ of the form $H(v) = \Omega(Av,v)$, where A: $V \to V$ is a smooth map such that $DA(x)$ is Ω skew for all x, then f is symplectic, i.e., $f^*\Omega = \Omega$.

Proof: (1) We compute X_H:

$$\Omega(X_H(x),u) = dH(x)u = \frac{d}{dt}_{|t=0} \Omega(A(x+tu),x+tu)$$
$$= \Omega(DA(x)\cdot x + Ax,u).$$

Hence

$$X_H(x) = DA(x)\cdot x + A(x).$$

Note in the special case where A is linear, we have $DA(x)\cdot x = A(x)$ and hence $X_H(x) = 2A(x)$.

(2) We compute X_{f^*H}:

$$\Omega(X_{f^*H}(x),u) = d(f^*H)(x)\cdot u = dH(f(x))\cdot Df(x)u.$$

For any $w \in V$ we compute

$$dH(f(x))\cdot w = \frac{d}{dt}_{|t=0} \Omega(A(f(x)+tw), f(x)+tw)$$
$$= \Omega(DA(f(x))\cdot w, f(x)) + \Omega(Af(x),w)$$
$$= \Omega(DA(f(x))\cdot f(x) + A(f(x)),w).$$

In particular for $w = Df(x)u$ we get

$$dH(f(x))\cdot Df(x)u$$
$$= -\Omega(Df(x)\cdot[DA(f(x))\cdot f(x) + A(f(x))],u),$$

and hence

$$X_{f^*H}(x) = -Df(x)[DA(f(x))\cdot f(x) + A(f(x))].$$

(3) On the other hand we have

$$f^*X_H(x) = Df^{-1}\circ X_H \circ f(x)$$

and

$$X_H(f(x)) = DA(f(x))\cdot f(x) + A(f(x)).$$

Therefore if $f^*X_H(x) = X_{f^*H}(x)$, then $Df^{-1}(x) = -Df(x)$; hence $-Df(x)^2 = \mathrm{id}$ and we have $\Omega(Df(x)v_1, Df(x)v_2) = -\Omega(v_1, Df(x)^2v_2) = \Omega(v_1,v_2)$, which means that f is symplectic. ■

ACKNOWLEDGMENTS

We wish to thank the Dipartimento di Fisica of the University of Naples, Italy, for its hospitality.

This work was supported in part by the INFN and the GNFM del CNR.

[1] D. G. Currie and E. J. Saletan, "Canonical transformations and quadratic Hamiltonians," Nuovo Cimento B **9**, 143 (1972).

[2] G. Marmo, E. J. Saletan, R. Schmid, and A. Simoni, "Bi-Hamiltonian dynamical systems and the quadratic-Hamiltonian theorem," Nuovo Cimento B **100**, 297 (1987).

[3] R. Abraham, J. E. Marsden, and T. S. Ratiu, *Manifolds, Tensor Analysis, and Applications* (Addison-Wesley, Reading, MA, 1983).

[4] R. Schmid, *Infinite Dimensional Hamiltonian Systems* (Biblionopolis, Naples, 1987).

[5] E. J. Saletan and A. H. Cromer, *Theoretical Mechanics* (Wiley, New York, 1971).

Spectral transform and solitons for generalized coupled Bloch systems

Jérôme J.-P. Leon[a)]
Laboratoire de Physique Mathématique, Université des Sciences et Techniques du Languedoc, F-34060 Montpellier Cedex, France

(Received 23 September 1987; accepted for publication 11 May 1988)

The spectral transform and Bäcklund transformation for a generalization of the coupled Bloch system arising in nonlinear optics are studied. The equation is related to a singular dispersion relation and is treated as a representative example of a general method for studying such equations. The spectral transform is developed through the $\bar{\partial}$ formalism. The Bäcklund transformation is derived in a very general way and solved to obtain the one-soliton and breatherlike solutions. The nonlinear superposition formula is also constructed.

I. INTRODUCTION

Recent progress in nonlinear optics[1,2] has shown the importance of the coupled Bloch system

$$\mathbf{S}_t^+ = (1/\lambda_0)\mathbf{S}^+ \times \mathbf{S}^-, \quad \mathbf{S}_x^- = (1/\lambda_0)\mathbf{S}^+ \times \mathbf{S}^-, \tag{1.1}$$

where the three-vectors $\mathbf{S}^\pm = \mathbf{S}^\pm(x,t)$ are the Stokes vectors representing the polarizations of the electric field, and λ_0^{-1} is the strength of the nonlinearity.[1] The interesting feature of (1.1) is that it possesses soliton solutions called "polarization solitons"[1] or "polarization domains."[2]

Equation (1.1) is a special case of the more general one,

$$\mathbf{S}_t^+ = \mathbf{S}^+ \times \langle (1/\lambda)\mathbf{S}^- \rangle, \quad \mathbf{S}_x^- = \mathbf{S}^+ \times (1/\lambda)\mathbf{S}^-, \tag{1.2}$$

in which $\mathbf{S}^+ = \mathbf{S}^+(x,t)$ and $\mathbf{S}^- = \mathbf{S}^-(\lambda,x,t)$, λ being a complex parameter in $D = \{\lambda \in \mathbb{C}, 0 < |\lambda| < \infty\}$. The average is defined as

$$\langle (1/\lambda)\,\mathbf{S}^- \rangle = -\frac{1}{2\pi} \int\int_D d\lambda \wedge d\bar{\lambda}\, \mathbf{S}^-(\lambda,x,t)\, \frac{a(\lambda,t)}{\lambda}, \tag{1.3}$$

where the scalar distribution $a(\lambda,t)$ vanishes as $|\lambda| \to \infty$, is bounded in $\lambda = 0$, but is otherwise arbitrary.

We refer to Eq. (1.2) as the *generalized coupled Bloch system*, but we emphasize the fact that the question of its physical meaning is open.

Equation (1.1) is obtained for $a(\lambda,t) = -i\pi\delta(\lambda - \lambda_0)$. The averaging appearing in (1.2) has mathematically the same origin as the inhomogeneous broadening in the reduced Maxwell–Bloch system[3] or the caviton equation.[4] The common relevant feature of these equations is the fact that they possess, in the spectral transform scheme, *singular dispersion relations.*

We have constructed a general method to investigate nonlinear evolution equations possessing singular dispersion relations with the help of the $\bar{\partial}$ approach.[5] The system (1.2) has been shown to be integrable in this formalism.[6] Here we study (1.2) in more detail as a simple representative example of the method sketched in Ref. 5, which we reformulate for completeness in Sec. II.

In Sec. III we establish the matrix local $\bar{\partial}$ problem from which we obtain the integrability of (1.2) by using the general method of Sec. II. The inverse spectral transform method to solve (1.2) is therefore constructed.

In Sec. IV we obtain the general formulas defining the Bäcklund transformations for evolutions with singular dispersion relations. Only very recently and independently[7] has this method been constructed. In Ref. 7 the (Darboux) transformation is solved for singular dispersion relations for the first time, in the only known example—the one corresponding to the Zakharov–Shabat spectral problem (which leads to the reduced Maxwell–Bloch system,[3] the self-induced transparency equation,[3] or the caviton equation,[4] depending on the chosen reduction). The method that we use here proceeds through the Boiti–Tu method[8] and formally applies for any starting spectral (or $\bar{\partial}$) problem. In some sense the "t part" of the Bäcklund transformation will appear to be *universal* for singular dispersion relations.

These results are used in Sec. V to derive the one-soliton solution to Eq. (1.2), which will explicitly depend on the arbitrary weight $a(\lambda,t)$ appearing in (1.3). Although this step consists "simply" in solving first-order (2×2)-matrix ordinary differential equations, the way to obtain the result is not so simple, and we shall provide the computations in some detail. We also obtain, as a special case, a breatherlike solution.

It is well known[8] that, as soon as one has a Bäcklund gauge, one can construct the nonlinear superposition formula (Bianchi theorem). This is done for our case in Sec. VI, where we obtain a way to construct a new solution from three other known solutions by algebraic steps. This provides a means to build recursively an N-soliton solution to Eq. (1.2).

We examine finally in Sec. VII the time evolution of the spectral transform corresponding to the evolution (1.2). It possesses the usual characteristic features common to evolutions with singular dispersion relations: while the evolution of the discrete spectrum is standard, that of the radiation component may present exponential growth or damping. This fact has been noticed as an irreversible evolution in the case of the self-induced transparency equation.[3] It means that the medium, while being transparent to the solitons, may produce an exponential damping (or growth) of the radiation in the t direction. In our case the t evolution depends on the weight $a(\lambda,t)$. It is then possible to obtain a bounded evolution of the radiation component.

[a)] Unité de Recherche Associée au Centre National de la Recherche Scientifique No. UA 040768.

 0022-2488/88/092012-08$02.50

II. SINGULAR DISPERSION RELATIONS

The first instance of the solution by the inverse spectral transform of an evolution having a singular dispersion relation was given in 1974[3] for the Zakharov–Shabat spectral problem. Then we had to wait until 1987 to have another example using instead the Schrödinger spectral problem.[4]

In the meantime a number of different spectral problems have been studied, but always in association with polynomial (or, at most, rational) dispersion relations.

We have recently given[5] a very simple recipe to construct, almost by inspection, the nonlinear evolution equations associated with a given spectral problem for any singular dispersion relation vanishing for large values of the modulus of the spectral parameter. We recall the method hereafter.

The starting point is a $\bar{\partial}$ problem,[9,10] which we choose to be matrix and local, having the form

$$\frac{\partial}{\partial \bar{k}} \Psi = \Psi R, \quad k \in D \subset \mathbb{C}, \tag{2.1}$$

with $R(k)$ given in D, and

$$\Psi = \Psi_0, \quad k \in \partial D, \tag{2.2}$$

with $\Psi_0(k)$ given and continuous on the boundary ∂D of D.

By extension of the vocabulary used in the scattering theory, $\Psi(k)$ is called the "eigenfunction" (to be determined), and $R(k)$ the "spectral transform."

We shall always assume that $R(k)$ is chosen in such a way that the solution of the $\bar{\partial}$ problem (2.1), (2.2) is given by the *unique* solution of the following integral equation (Cauchy–Green formula):

$$\Psi(k) = \frac{1}{2i\pi} \int_{\partial D} \frac{d\lambda}{\lambda - k} \Psi_0(\lambda) E^{-1}(\lambda) E(k) + \frac{1}{2i\pi} \int\int_D \frac{d\lambda \wedge d\bar{\lambda}}{\lambda - k} \Psi(\lambda) R(\lambda) E^{-1}(\lambda) E(k). \tag{2.3}$$

The term $E(k)$ is a matrix factor (in general, an exponential) chosen to make $\Psi_0(\lambda) E^{-1}(\lambda)$ bounded on the boundary ∂D.

Next the distribution R (and consequently Ψ) is supposed to depend on a real parameter t according to

$$\frac{\partial}{\partial t} R = [R, \Omega], \tag{2.4}$$

where the matrix-valued function $\Omega(k,t)$ is the *dispersion relation*. Here Ω is assumed to be nonanalytic in D (more precisely, it will be assumed to have a set Σ of poles or lines of discontinuity) and to vanish on the boundary ∂D. (A straightforward extension to k-polynomial behavior on ∂D is possible.[6]) In short,

$$\frac{\partial}{\partial \bar{k}} \Omega(k,t) \neq 0, \quad k \in \Sigma \subset D; \qquad \Omega(k,t) \sim 0, \quad k \in \partial D. \tag{2.5}$$

The space variable x is now introduced as an external real parameter in Ψ only. The problem is to account for this dependence by using (2.1), (2.4), and the hypothesis

$$\frac{\partial}{\partial x} R = 0. \tag{2.6}$$

To that end we define the matrices $U(k,x,t)$ and $V(k,x,t)$ by (subscripts x, t, and $\bar{k}$ denote differentiation)

$$U \doteq (\Psi_x)\Psi^{-1}, \quad V \doteq (\Psi_t - \Psi\Omega)\Psi^{-1}. \tag{2.7}$$

Equations (2.4) and (2.6) then imply

$$U_{\bar{k}} = 0, \quad V_{\bar{k}} = -\Psi \Omega_{\bar{k}} \Psi^{-1}, \quad k \in D. \tag{2.8}$$

The method for constructing the analytic function U in D is given in Ref. 10. In short, it is done by computing Ψ_x from (2.3), by using the fact that (2.6) implies that Ψ_x solves (2.1) (with a different bound), and by applying the hypothesis that the solution of (2.3) is unique. An explicit example is given in Sec. III, and we suppose here that $U(k,x,t)$ has been obtained or, in other words, that we have linked the $\bar{\partial}$ problem to the corresponding spectral problem $\Psi_x = U\Psi$.

To construct the function V we apply the Cauchy–Green formula (forgetting for a while the x and t dependences):

$$V(k) = \frac{1}{2i\pi} \int_{\partial D} \frac{d\lambda}{\lambda - k} V(\lambda) + \frac{1}{2i\pi} \int\int_D \frac{d\lambda \wedge d\bar{\lambda}}{\lambda - k} \frac{\partial}{\partial \bar{\lambda}} V(\lambda). \tag{2.9}$$

In general, as Ω has been chosen to vanish on ∂D, the bound (2.2) and the definition (2.7) allow us to prove that V vanishes on ∂D, or at least that the first term of (2.9) vanishes (see Sec. III). The result is then

$$V(k) = -\frac{1}{2i\pi} \int\int_D \frac{d\lambda \wedge d\bar{\lambda}}{\lambda - k} \Phi(\lambda), \tag{2.10}$$

where the distribution Φ stands for

$$\Phi(\lambda,x,t) = \Psi(\lambda,x,t) \left[\frac{\partial}{\partial \bar{\lambda}} \Omega(\lambda,t) \right] \Psi^{-1}(\lambda,x,t). \tag{2.11}$$

It remains to write the evolution of U, obtained by computing U_t and V_x out of (2.7) and checking directly that

$$\Omega_x = 0 \Rightarrow U_t - V_x + [U,V] = 0. \tag{2.12}$$

Inserting in the above equation the expression (2.10) for V and noting that

$$\Phi_x(k,x,t) = [U(k,x,t), \Phi(k,x,t)], \tag{2.13}$$

we obtain

$$U_t(k,x,t) = -\frac{1}{2i\pi} \int\int_D \frac{d\lambda \wedge d\bar{\lambda}}{\lambda - k} \times [U(\lambda,x,t) - U(k,x,t), \Phi(\lambda,x,t)]. \tag{2.14}$$

The above two coupled differential equations constitute the general form of solvable evolutions associated with a $\bar{\partial}$ problem and a singular dispersion relation.

III. GENERALIZED COUPLED BLOCH SYSTEM

We apply the general formalism of Sec. II to an explicit example that will lead to the system (1.2) of solvable evolution.

The starting $\bar{\partial}$ problem on the domain

$$D=\{k\in\mathbb{C},\ 0<|k|<\infty\} \tag{3.1}$$

is chosen to be

$$\frac{\partial}{\partial\bar{k}}\Psi(k,x,t)=\Psi(k,x,t)R(k,t), \tag{3.2}$$

with the bounds

$$\Psi(k,x,t)\sim(1/k)[A(x,t)+O(k)]\times\exp[(i/k)\sigma_3 x],\quad \text{as } |k|\to 0, \tag{3.3}$$

$$\Psi(k,x,t)\sim O(1/k)\exp[(i/k)\sigma_3 x],\quad \text{as } |k|\to\infty. \tag{3.4}$$

The boundary ∂D is naturally made out with the circles centered in $k=0$, with radii ρ and ϵ in the limits $\rho\to\infty$, $\epsilon\to 0$, oriented as shown in Fig. 1.

To compute the matrix $U(k,x,t)$ (the "potential") we follow Ref. 10: U being analytic in D [see (2.8)], $U\Psi$ also solves (2.1) and we compare its behavior as $|k|\to 0$ to that of Ψ_x [which from (2.6) also solves (2.1)]:

$$\Psi_x\sim(i/k^2)A\sigma_3A^{-1}[A+O(k)]\exp[(i/k)\sigma_3 x]. \tag{3.5}$$

Now we use the fundamental assumption that the solution of the $\bar{\partial}$ problem is uniquely given by the solution of the integral equation (2.3), which from (3.3) and (3.4) reads

$$\Psi(k,x,t)e^{-(i/k)\sigma_3 x}=\frac{1}{k}A(x,t)+\frac{1}{2i\pi}\iint_D\frac{d\lambda\wedge d\bar{\lambda}}{\lambda-k}\times\Psi(\lambda,x,t)R(\lambda,t)e^{-(i/\lambda)\sigma_3 x}. \tag{3.6}$$

The solution is then uniquely determined by its behavior on the boundary, and comparison of (3.5) with (3.3) implies

$$\Psi_x(k,x,t)=(i/k)S(x,t)\Psi(k,x,t),\quad S=A\sigma_3A^{-1}. \tag{3.7}$$

The above spectral problem is that of Takhtajan,[11] used to solve the hierarchy of "Heisenberg spin-chain" equations when the dispersion relation is chosen polynomial in k^{-1}.

The relation (3.7) has been obtained by analyzing the behavior of Ψ in the neighborhood of $k=0$. We have to prove that this result is consistent with the behavior (3.4), which we write

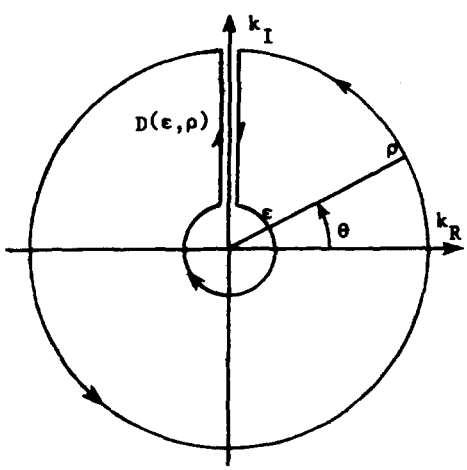


FIG. 1. Boundary of the domain $D(\epsilon,\rho)=\{k\in\mathbb{C},\ \epsilon<|k|<\rho\}$. The domain D itself is given by

$$D=\lim_{\epsilon\to 0}\lim_{\rho\to\infty}D(\epsilon,\rho).$$

$$\Psi e^{-(i/k)\sigma_3 x}\doteq\varphi(k,x,t)\sim[(1/k)B+O(1/k^2)],\quad |k|\to\infty. \tag{3.8}$$

By differentiation of the above behavior with respect to x and comparison with (3.7) we must have

$$B_x=0. \tag{3.9}$$

The matrix B may be evaluated by looking at the large k asymptotics directly in the integral equation (3.6) in order to obtain

$$B=A-\frac{1}{2i\pi}\iint_D d\lambda\wedge d\bar{\lambda}\,\Psi(\lambda)R(\lambda)e^{-(i/\lambda)\sigma_3 x}. \tag{3.10}$$

Making use of (3.7) allows us to derive

$$B_x=A_x-i[A\sigma_3A^{-1}\varphi_1-\varphi_1\sigma_3], \tag{3.11}$$

with

$$\varphi_1(x,t)=\frac{1}{2i\pi}\iint_D\frac{d\lambda\wedge d\bar{\lambda}}{\lambda}\Psi(\lambda,x,t)R(\lambda,t)e^{-(i/\lambda)\sigma_3 x}. \tag{3.12}$$

The quantity $\varphi_1(x,t)$ also appears in the expansion around $k=0$:

$$\varphi(k,x,t)\sim(1/k)A(x,t)+\varphi_1(x,t)+O(k),\quad k\to 0, \tag{3.13}$$

which, by using (3.7), gives

$$A_x=i[A\sigma_3A^{-1}\varphi_1-\varphi_1\sigma_3], \tag{3.14}$$

and hence (3.9) is proved.

The evolution equation related to (3.7) for a singular dispersion relation is obtained by replacing $U(k,x,t)$ in (2.14) by $(i/k)S(x,t)$. We obtain

$$S_t(x,t)=\left[S(x,t),\frac{1}{2i\pi}\iint_D\frac{d\lambda\wedge d\bar{\lambda}}{\lambda}\varphi(\lambda,x,t)\right]. \tag{3.15}$$

It is important to check here that, for the boundary ∂D of Fig. 1 and the bounds (3.3) and (3.4) of $\Psi(k)$, the first term in the expression (2.9) for V does indeed vanish. Using the definition (2.7) of V, it is immediately seen that

$$\Omega(k,t)\sim O(1/|k|),\quad \text{as } |k|\to\infty, \tag{3.16}$$

is a sufficient condition to ensure that the contribution on the large circle of Fig. 1 vanishes. We are left with

$$\int_{\partial D}\frac{d\lambda}{\lambda-k}V(\lambda)=-i\lim_{\epsilon\to 0}\int_0^{2\pi}\frac{\epsilon e^{i\theta}\,d\theta}{\epsilon e^{i\theta}-k}\times\left\{A_tA^{-1}-A\exp\left[\frac{i}{\epsilon}e^{-i\theta}\sigma_3 x\right]\Omega(\epsilon e^{i\theta},t)\times\exp\left[-\frac{i}{\epsilon}e^{-i\theta}\sigma_3 x\right]A^{-1}\right\}. \tag{3.17}$$

A necessary and sufficient condition for this integral to vanish for any value of x is that Ω be diagonal and bounded for $k=0$. The part proportional to $\mathbf{1}$ in Ω being irrelevant in the commutator (2.7), we write

$$\Omega(k,t)=\omega(k,t)\sigma_3,\quad \omega(k)<\infty \text{ for } |k|\to 0. \tag{3.18}$$

Equation (3.15) and the related equation (2.13) become

$$S_t(x,t) = [S(x,t),\langle (i/\lambda)F(\lambda,x,t)\rangle], \tag{3.19}$$

$$F_x(k,x,t) = [S(x,t),(i/k)F(k,x,t)], \tag{3.20}$$

where we have defined

$$F(k,x,t) \doteqdot \Psi(k,x,t)\sigma_3\Psi^{-1}(k,x,t), \tag{3.21}$$

$$\langle (i/\lambda)F(\lambda,x,t)\rangle \doteqdot -\frac{1}{2\pi}\int\int_D d\lambda\wedge d\bar{\lambda} \times\frac{i}{\lambda}F(\lambda,x,t)\frac{\partial}{\partial\bar{\lambda}}\omega(\lambda,t). \tag{3.22}$$

We call the above system (3.19), (3.20) the generalized coupled Bloch system, and we recover Eq. (1.2) for the notation $\boldsymbol{\sigma} = (\sigma_1,\sigma_2,\sigma_3)$ and

$$S = -\tfrac{1}{2}\mathbf{S}^+\cdot\boldsymbol{\sigma}, \quad F = -\tfrac{1}{2}\mathbf{S}^-\cdot\boldsymbol{\sigma}, \quad \omega_{\bar{\lambda}} = a. \tag{3.23}$$

This system is therefore solvable by the inverse spectral transform method. More precisely, its solution can be reduced to that of an integral equation in the following way: for a given spectral transform $R(k,0)$ and the dispersion relation $\Omega(k,t)$, we first solve the evolution (2.4) to obtain [for the choice (3.18)]

$$R^d(k,t) = R^d(k,0), \tag{3.24}$$

$$R^a(k,t) = R^a(k,0)\exp\left[2\sigma_3\int_0^t dt'\,\omega(k,t')\right]. \tag{3.25}$$

(The superscripts d and a denote, respectively, the diagonal and antidiagonal parts of the matrix R.) Note that we may always choose $R^d = 0$.[5] Then the solution $\Psi(k,x,t)$ of the integral equation (3.6) provides $S(x,t)$ with the help of (3.7), and $F(k,x,t)$ with the help of (3.21). To solve the integral equation (3.6) we define $G = A^{-1}\Psi\exp[-(i/k)\sigma_3 x]$, which is the solution of

$$G(k,x,t) = \frac{1}{k} + \frac{1}{2i\pi}\int\int_D \frac{d\lambda\wedge d\bar{\lambda}}{\lambda - k}G(\lambda,x,t) \times R(\lambda,t)\exp\left[-\frac{2i}{\lambda}\sigma_3 x\right]. \tag{3.26}$$

The matrix potential $A(x,t)$ is then obtained by solving

$$A^{-1}A_x = ((i/k)[\sigma_3,G] - G_x)G^{-1}, \quad A\to\sigma_3 \text{ as } x\to-\infty. \tag{3.27}$$

If instead of $R(k,0)$ we had the initial datum $S(x,0)$, we would first have to compute $\Psi(k,x,0)$ and $R(k,0)$ by solving the direct spectral problem associated with (3.7). Details on this question can be found in Refs. 11 and 12.

IV. BÄCKLUND TRANSFORMATION FOR SINGULAR DISPERSION RELATIONS

We derive here the general basic formulas for the Bäcklund transformation without specifying the starting $\bar{\partial}$ problem, that is, without specifying the structure in k of the potential U in the spectral problem ($k\in D$)

$$\Psi_x = U\Psi. \tag{4.1}$$

The dispersion relation Ω is of course chosen singular, and the auxiliary spectral problem written in (2.7) reads

$$\Psi_t = V\Psi + \Psi\Omega, \tag{4.2}$$

with

$$V(k,x,t) = -\frac{1}{2i\pi}\int\int_D \frac{d\lambda\wedge d\bar{\lambda}}{\lambda - k}\Phi(\lambda,x,t), \tag{4.3}$$

where Φ was given in (2.11).

The method used is that of Boiti–Tu[8] and consists in making the gauge transformation from Ψ to

$$\Psi'(k,x,t) = B(k,x,t)\Psi(k,x,t), \tag{4.4}$$

with the following requirements:

$$\Psi'_x = U'\Psi', \quad \Psi'_t = V'\Psi' + \Psi'\Omega, \tag{4.5}$$

in which U' has the same structure in k as U, and V' is given by (4.3), where Ψ is changed with Ψ'. The consequence of this is that B has to solve

$$B_x = U'B - BU, \tag{4.6}$$

$$B_t = V'B - BV. \tag{4.7}$$

It is then immediate to check by cross differentiation of the above system that if (U,V) solves the nonlinear evolution equation (2.12), then (U',V') does also and B defines an auto-Bäcklund transformation.

We are interested here in an elementary Bäcklund transformation corresponding to the simplest nontrivial choice:

$$B(k,x,t) = ikB_0(x,t) + B_1(x,t). \tag{4.8}$$

The first equation (4.6), the so-called x part, holds for any evolution associated with (4.1). Therefore we shall suppose here that it is solved, which will be done explicitly in Sec. V for the spectral problem (3.7).

So we examine the t part (4.7). We first insert in (4.7) the expression (4.3) for V and the analogous one for V'. Then we make use of the following trivial identities:

$$B(k,x,t) \equiv B(\lambda,x,t) + i(k-\lambda)B_0(x,t), \tag{4.9}$$

obtained from (4.8), and

$$\Phi'(\lambda,x,t)B(\lambda,x,t) - B(\lambda,x,t)\Phi(\lambda,x,t) \equiv 0, \tag{4.10}$$

obtained from (4.4) and (2.11). Finally, by identifying the coefficients of k^0 and k^1 we get

$$\frac{\partial}{\partial t}B_0(x,t) = 0, \tag{4.11}$$

$$\frac{\partial}{\partial t}B_1(x,t) = \left\{\frac{1}{2\pi}\int\int_D d\lambda\wedge d\bar{\lambda}\,\Phi'(\lambda,x,t)\right\}B_0(x,t) - B_0(x,t)\left\{\frac{1}{2\pi}\int\int_D d\lambda\wedge d\bar{\lambda}\,\Phi(\lambda,x,t)\right\}. \tag{4.12}$$

Any higher-order Bäcklund transformation, corresponding to higher-order k polynomials in (4.8), can be solved in the same way.

The above two equations constitute the general form of the t part of the elementary Bäcklund transformation related to the singular dispersion relation.

It is worth remarking that the structure of the t part does not depend on the choice of the starting principal spectral problem. In that sense it is *universal*.

V. SOLITON SOLUTION

We solve the Bäcklund transformation for the Takhtajan spectral problem (3.7) when the related evolution equation is the pair (3.19), (3.20), or, equivalently, (1.2).

We shall first solve the x part, which will lead to the usual structure of the soliton solution for the Heisenberg ferromagnet.[11] The t dependence of the "constants" of integration will be determined while solving the t part. Although there is no special difficulty, the computations are not straightforward and we give below the method of solution with some necessary details.

By using (4.6) with both the expression (4.8) of B and the choice $(i/k)S(x,t)$ as potential $U(k,x,t)$, we obtain the set of equations (coefficients of k^1, k^0, and k^{-1})

$$B_{0,x} = 0, \tag{5.1}$$

$$B_{1,x} = B_0 S - S' B_0, \tag{5.2}$$

$$B_1 S - S' B_1 = 0, \tag{5.3}$$

which defines the x part of the Bäcklund transformation relating the two potentials S' and S.

First of all we remark that we may always choose

$$B_0(x,t) = \mathbf{1}. \tag{5.4}$$

Indeed (5.1) implies that any solution S'' obtained with a gauge B made out with a full matrix B_0'' (x independent) can be equivalently obtained from the solution S', corresponding to a gauge with $B_0 = \mathbf{1}$, via

$$S''(x,t) = B_0'' S(x,t) B_0''^{-1}. \tag{5.5}$$

This is also valid for the t dependence because of (4.11).

The question now is that of the choice of the starting solution (Ψ,S) from which we shall construct the new solution (Ψ',S'). Looking for stationary solutions of (3.2) we note that the pair

$$S_0(x,t) = s_0(x) + s_3(x)\sigma_3, \tag{5.6}$$

$$F_0(k,x,t) = M_0(k,t)\sigma_3 M_0^{-1}(k,t), \quad M_0^a(k,t) = 0, \tag{5.7}$$

solves the generalized coupled Bloch system (3.19), (3.20) with the corresponding eigenfunction

$$\Psi_0(k,x,t) = \exp\left[\frac{i}{k}\int_0^x S_0\right] M_0(k,t). \tag{5.8}$$

If we now use (5.3) where S' decomposes as in (3.6),

$$S'(x,t) = A'(x,t)\sigma_3 A'^{-1}(x,t), \tag{5.9}$$

we obtain the following equation for the matrix $N = B_1^{-1}A'$:

$$[N,\sigma_3] = \{(s_3 - 1)\sigma_3 + s_0\}N. \tag{5.10}$$

The two sets of solutions for this equation are obtained either for $\{s_3 = 1,\ s_0 = 0,\ N^a = 0\}$ or for $\{s_3 = -1,\ s_0 = 0,\ N^d = 0\}$. No x dependence is then allowed in (5.6), and we choose

$$S = \sigma_3 \tag{5.11}$$

as starting potential in the set of equations (5.1)–(5.3).

For S' we follow Ref. 11 and seek it in the form

$$S' = \sigma_1 \sin\theta\cos\varphi + \sigma_2 \sin\theta\sin\varphi + \sigma_3\cos\theta, \tag{5.12}$$

where $\theta(x,t)$ and $\varphi(x,t)$ are real valued. [Note that A' in (5.9) is obtained from S' by replacing θ by $\frac{1}{2}\theta$, and that $A'^{-1} \equiv A'$.] The general solution of (5.3) can be written by using (5.9):

$$B_1 = A' \begin{pmatrix} \mu & 0 \\ 0 & \nu \end{pmatrix}. \tag{5.13}$$

We insert this expression in (5.2), together with (5.11) and (5.12), and multiply the resulting equation by A'. The diagonal part reveals that $|\mu|^2$ and $|\nu|^2$ do not depend on x, and that $\partial_x \log\mu = -\partial_x \log\nu$. We then define a new unknown real $\sigma(x,t)$ by

$$\begin{aligned} \mu(x,t) &= \mu_0(t)\exp[i\{\sigma(x,t) - \varphi(x,t)\}], \\ \nu(x,t) &= \nu_0(t)\exp[-i\{\sigma(x,t) - \varphi(x,t)\}]. \end{aligned} \tag{5.14}$$

The remaining equation resulting from the diagonal part of (5.2) is

$$\sigma_x = \varphi_x \cos^2(\theta/2). \tag{5.15}$$

After having used (5.14) and the fact that φ and θ are real valued, the antidiagonal part of (5.2) leads to

$$\mu_0(t) = \nu_0^*(t) \tag{5.16}$$

(* stands for complex conjugation), and, after having used (5.15), to

$$\left(e^{i\sigma}\sin\frac{\theta}{2}\right)_x = -\frac{2}{\mu_0}\sin\frac{\theta}{2}\cos\frac{\theta}{2}e^{i\varphi}. \tag{5.17}$$

Equations (5.15) and (5.17) are the basic equations to be solved. This is done in three steps: (i) proof that the function $\sigma(x,t)$ introduced in (5.14) is a linear function of x; (ii) computation of the solution $\theta(x,t)$; and (iii) computation of the solution $\varphi(x,t)$.

To prove that σ_x is a constant, we first rewrite (5.15) as

$$\varphi_x - \sigma_x = \varphi_x \sin^2(\theta/2). \tag{5.18}$$

Second, we use (5.17) to express $\exp[i(\varphi - \sigma)]$, which is then differentiated. Together with (5.18) we obtain

$$\left(iu^2\varphi_x - \frac{\partial}{\partial x}\right)[i\varphi_x(1-u^2)^{1/2} + u_x(1-u^2)^{-1/2}] = 0, \tag{5.19}$$

in which $u(x,t) = \sin\frac{1}{2}\theta(x,t)$. Upon integration of the imaginary part of (5.19) we obtain the desired result that $\varphi_x(1-u^2)$ is a constant. We write

$$\sigma(x,t) = a(t)x + \sigma_0(t). \tag{5.20}$$

To compute $\theta(x,t)$ we insert the above expression in that of $\exp[i(\varphi - \sigma)]$ obtained from (5.17), and we simply write that its modulus is 1. After some simple algebra we get

$$\sin\tfrac{1}{2}\theta(x,t) = \alpha/\cosh[\alpha(|2/\mu_0|x + x_0)], \tag{5.21}$$

in which $x_0(t)$ is arbitrary and $\alpha(t)$ is defined by

$$\alpha^2 = 1 - \tfrac{1}{2}a^2|\mu_0|^2. \tag{5.22}$$

The quantities $a(t)$ and $\mu_0(t)$ have to be chosen such that $\alpha^2 < 1$, the condition of existence of a real-valued solution θ.

To compute $\varphi(x,t)$ we insert (5.21) into (5.15) and evaluate $\varphi_x - a(t)$. After some algebra it can be written as an exact derivative and, finally,

$$\begin{aligned} \varphi(x,t) = \varphi_0(t) + a(t)x \\ + \arctan\{(\alpha/\sqrt{1-\alpha^2}) \\ \times\tanh[\alpha(|2/\mu_0|x + x_0)]\}. \end{aligned} \tag{5.23}$$

To complete the solution of the x part of the Bäcklund transformation, it remains to find the relationship between φ_0 and σ_0. This is done by computing explicitly $\exp[i(\varphi - \sigma)]$ with the help of (5.20) and (5.23), and by

comparing the result with (5.17). We obtain

$$\exp[i(\varphi_0-\sigma_0)]=-i\mu_0/|\mu_0|. \tag{5.24}$$

Therefore the soliton (θ,φ) depends on four independent functions of t to be determined (a,x_0,φ_0,μ_0). To simplify the notation we fix

$$\mu_0(t)=i\gamma(t),\quad \gamma\in\mathbb{R},\quad \sigma_0(t)=\varphi_0(t), \tag{5.25}$$

$$\xi(x,t)=\alpha(t)[2x/\gamma(t)+x_0(t)]. \tag{5.26}$$

The resulting explicit form of $B_1(x,t)$ is obtained from (5.13) and (5.14), and reads

$$B_1=i\gamma\sqrt{1-\alpha^2}+(\alpha\gamma/\cosh\xi)\sigma_2 \times\exp[i\sigma_3(ax+\sigma_0)]+\alpha\gamma\sigma_3\tanh\xi. \tag{5.27}$$

Note that (5.2) gives $S'=\sigma_3-B_{1,x}$. The eigenfunction $\Psi'(k,x,t)$ reads

$$\Psi'=(ik+B_1)\exp[(i/k)\sigma_3 x]N, \tag{5.28}$$

where N is a normalization constant matrix, and allows us to compute, for future use,

$$\Psi'\sigma_3\Psi'^{-1}-\sigma_3=\rho(k)\alpha^2\gamma^2\begin{pmatrix}\dfrac{-1}{\cosh^2\xi} & \dfrac{e^{-i(ax+\sigma_0)}}{\cosh\xi}\left[\dfrac{k+\gamma\sqrt{1-\alpha^2}}{\alpha\gamma}-i\tanh\xi\right]\\ \dfrac{e^{i(ax+\sigma_0)}}{\cosh\xi}\left[\dfrac{k+\gamma\sqrt{1-\alpha^2}}{\alpha\gamma}+i\tanh\xi\right] & \dfrac{1}{\cosh^2\xi}\end{pmatrix}, \tag{5.29}$$

in which

$$\rho(k)=2/(k^2+2\gamma k\sqrt{1-\alpha^2}+\gamma^2). \tag{5.30}$$

The above expression enters into $B_{1,t}$ in (4.12) when the dispersion relation is given in (3.18) (and $B_0=\mathbf{1}$):

$$\frac{\partial}{\partial t}B_1(x,t)=\frac{1}{2\pi}\int\!\!\int_D d\lambda\wedge d\bar\lambda\,\omega_{\bar\lambda}(\lambda,t) \times[\Psi'(\lambda,x,t)\sigma_3\Psi'^{-1}(\lambda,x,t)-\sigma_3]. \tag{5.31}$$

This equation, together with (5.27) and (5.29), allows us to determine the time dependence of the set $\{a,x_0,\varphi_0,\gamma\}$.

We remark first of all that the quantity (5.29) vanishes as $|x|\to\infty$. By taking the limit of (5.27) and (5.31) this gives

$$\alpha_t=0,\quad \gamma_t=0, \tag{5.32}$$

and hence [from (5.22) with (5.25)]

$$a_t=0. \tag{5.33}$$

Finally, by using (5.31) explicitly we obtain

$$\frac{\partial}{\partial t}x_0(t)=-\frac{\gamma}{2\pi}\int\!\!\int_D d\lambda\wedge d\bar\lambda\,\omega_{\bar\lambda}(\lambda,t)\rho(\lambda), \tag{5.34}$$

$$\frac{\partial}{\partial t}\sigma_0(t)=-\frac{1}{2\pi}\int\!\!\int_D d\lambda\wedge d\bar\lambda \times\omega_{\bar\lambda}(\lambda,t)(\lambda+\gamma\sqrt{1-\alpha^2})\rho(\lambda), \tag{5.35}$$

which achieve the derivation of the soliton solution.

The soliton depends on four real parameters $(a,\gamma,x_0(0),\sigma_0(0))$ and the weight $\omega_{\bar\lambda}(\lambda,t)$, which has to be chosen such that $x_0(t)$ and $\sigma_0(t)$ be real valued. As an example we may set $(\lambda=\lambda_R+i\lambda_I)$

$$\omega_{\bar\lambda}(\lambda,t)=-i\pi g(\lambda,t)\delta(\lambda_I), \tag{5.36}$$

where $g(\lambda)$ is $L^2(\mathbb{R})$. Then we have

$$x_0(t)=\gamma\int_{-\infty}^{+\infty}d\lambda\left[\rho(\lambda)\int_0^t g(\lambda,t')dt'\right]+x_0(0), \tag{5.37}$$

and similarly for $\sigma_0(t)$. One can play with different explicit choices of g as a function of t to make the soliton oscillate back and forth, make it move significantly only during a finite time interval, etc.

Another interesting simple case is obtained for

$$\omega_{\bar\lambda}(\lambda,t)=-i\pi g_0(t)\delta(\lambda_R-\lambda_0)\delta(\lambda_I). \tag{5.38}$$

The related evolution [see (2.13), (2.14), and (3.23)] becomes

$$\mathbf{S}_t^+=(1/\lambda_0)g_0(t)\mathbf{S}^+\times\mathbf{S}^-,\quad \mathbf{S}_x^-=(1/k)\mathbf{S}^+\times\mathbf{S}^-. \tag{5.39}$$

The above system is the simplest generalization of the equation of polarization domains (1.1) obtained here for $g_0=1$ and $k=\lambda_0$.

There exists another interesting solution if we choose

$$a=0\Rightarrow\alpha=1, \tag{5.40}$$

for which, restricting ourselves for simplicity to the case (5.36), we have

$$x_{0,t}=\gamma\int_{-\infty}^{+\infty}\frac{2\,d\lambda}{\lambda^2+\gamma^2}g(\lambda,t), \tag{5.41}$$

$$\sigma_{0,t}=\int_{-\infty}^{+\infty}\frac{2\lambda\,d\lambda}{\lambda^2+\gamma^2}g(\lambda,t). \tag{5.42}$$

These particular forms allow us to obtain a stationary solution by selecting for $g(\lambda,t)$ an odd function of λ. Centering the x axis in $x_0(0)$ the solution reads

$$S(x,t)=\left[\frac{\sinh(2x/\gamma)}{\cosh^2(2x/\gamma)}\sin\sigma_0(t)\right]\sigma_1 -\left[\frac{\sinh(2x/\gamma)}{\cosh^2(2x/\gamma)}\cos\sigma_0(t)\right]\sigma_2 +\left(1-\frac{2}{\cosh^2(2x/\gamma)}\right)\sigma_3. \tag{5.43}$$

This solution represents a breatherlike solution. Such a solution cannot be obtained for Eq. (1.1), which corresponds to a delta function for $g(\lambda,t)$.

We finally mention that the two poles $(\alpha\gamma>0)$

$$k^{\pm}=-\gamma\sqrt{1-\alpha^2}\pm i\alpha\gamma \tag{5.44}$$

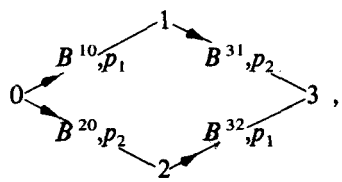
of $\rho(k)$ correspond to the bound states. Indeed we check in (5.28) that we have

$$\Psi'(k,x,t)\sim i\begin{pmatrix}(k-k^{+})e^{(i/k)x} & 0\\ 0 & (k-k^{-})e^{-(i/k)x}\end{pmatrix}N,\quad x\to\pm\infty. \tag{5.45}$$

Therefore, as $\alpha\gamma>0$, Ψ' does vanish at both ends of the x axis for $k=k^{+}$ or $k=k^{-}$.

VI. NONLINEAR SUPERPOSITION FORMULA

As soon as we are able to build a Bäcklund transformation by means of the Boiti–Tu gauge, we also know how to establish the nonlinear superposition formula.[8] It proceeds through the diagram



(6.1)

where p_n denotes a given set of parameters defining the solution $S(i)$ obtained from a solution $S(j)$ with gauge B^{ij}. This gauge is defined in the preceding section by the equations (a parameter p_n is understood)

$$B^{ij}=ik+B_1^{ij}, \tag{6.2}$$

$$B_{1,x}^{ij}=S(j)-S(i), \tag{6.3}$$

$$B_1^{ij}S(j)-S(i)B_1^{ij}=0. \tag{6.4}$$

The above diagram is equivalent to the equation [with the correct parameter as indicated in (6.1)]

$$B^{31}B^{10}=B^{32}B^{20}. \tag{6.5}$$

Inserting (6.2) into (6.5) we obtain the basic equations

$$B_1^{31}+B_1^{10}=B_1^{32}+B_1^{20}, \tag{6.6}$$

$$B_1^{31}B_1^{10}=B_1^{32}B_1^{20}. \tag{6.7}$$

The Bianchi theorem [i.e., the same solution $S(3)$ is indeed obtained by following the two ways indicated in the diagram (6.1)] is easily proved by differentiating (6.6) with respect to x and then using (6.3).

The nonlinear superposition formula is obtained by integrating (6.3) and using the result in (6.7) to get $S(3)$:

$$S(3)=S(0)-B_{1,x}^{30}, \tag{6.8}$$

$$B_1^{30}=\left[(B_1^{10})^2-(B_1^{20})^2\right]\left[B_1^{10}-B_1^{20}\right]^{-1}. \tag{6.9}$$

The matrices B_1^{j0} are given in (5.27) when $S(0)=\sigma_3$, with parameters $p_j=\{a_j,\gamma_j,x_{0j}(0),\sigma_{0j}(0)\}$, $j=1,2$, and with the expressions (5.34) and (5.35) of the velocities.

The above formulas then provide the two-soliton solution for Eq. (1.2). Depending on the choice of the function of t, $\omega_{\bar\lambda}(\lambda,t)$, the solitons may not separate as t increases.

The superposition formula holds for any evolution associated with the $\bar\partial$ problem (3.1), (3.4). It holds, in particular, for the hierarchy of Heisenberg spin-chain equations.[11]

VII. EVOLUTION OF THE SPECTRAL TRANSFORM

The evolution of the spectral transform $R(k,t)$ obeying Eqs. (3.24) and (3.25) presents interesting features in the case of singular dispersion relations as shown in Ref. 3.

We have proved[5] that for a matrix local $\bar\partial$ problem such as (3.2) one may always choose, via an appropriate gauge transformation, an off-diagonal matrix $R(k,t)$.

Moreover, the standard structure of R (in the case of "good" potentials) is given in Ref. 10 for the Zakharov–Shabat system and is equivalent to that of our problem.[13] It reads ($k_n^{\pm}$ are the $N^{\pm}$ bound states in $\pm\operatorname{Im}k^{\pm}>0$)

$$\begin{aligned}R(k,t)=&\begin{pmatrix}0 & r^{+}(k,t)\delta^{+}(k_I)\\ r^{-}(k,t)\delta^{-}(k_I) & 0\end{pmatrix}\\ &+\sum_{n=1}^{N^{+}}\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}c_n^{+}(t)\delta(k-k_n^{+})\\ &+\sum_{n=1}^{N^{-}}\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}c_n^{-}(t)\delta(k-k_n^{-}),\end{aligned} \tag{7.1}$$

where the distributions $\delta^{\pm}$ are defined by

$$f(k)\delta^{\pm}(k_I)=\lim_{\epsilon\to 0^{+}}f(k_R\pm i\epsilon)=f^{\pm}(k_R).$$

Its time dependence is given by (3.25):

$$R(k,t)=R(k,0)\exp\left[2\sigma_3\int_0^t dt\,\omega(k,t)\right]. \tag{7.2}$$

Differentiating the above two equations with respect to t we obtain

$$\begin{aligned}&(\partial_t+2\omega(k,t)\sigma_3)\begin{pmatrix}0 & r^{+}(k,t)\delta^{+}(k_I)\\ r^{-}(k,t)\delta^{-}(k_I) & 0\end{pmatrix}\\ &+\sum_{n=1}^{N^{+}}\delta(k-k_n^{+})\left\{(\partial_t+2\omega(k,t)\sigma_3)\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}c_n^{+}(t)\right\}\\ &+\sum_{n=1}^{N^{-}}\delta(k-k_n^{-})\left\{(\partial_t+2\omega(k,t)\sigma_3)\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}c_n^{-}(t)\right\}\\ &-\sum_{n=1}^{N^{+}}\delta'(k-k_n^{+})k_{n,t}^{+}\begin{pmatrix}0 & 1\\ 0 & 0\end{pmatrix}c_n^{+}(t)\\ &-\sum_{n=1}^{N^{-}}\delta'(k-k_n^{-})k_{n,t}^{-}\begin{pmatrix}0 & 0\\ 1 & 0\end{pmatrix}c_n^{-}(t)=0.\end{aligned} \tag{7.3}$$

Then, by identification of the coefficients of the different δ and δ' functions, we get the evolutions

$$k_{n,t}^{\pm}=0, \tag{7.4}$$

$$r_t^{\pm}(k,t)=\pm 2\omega^{\pm}(k,t)r^{\pm}(k,t),\qquad \omega^{\pm}=\omega(k_R\pm i0,t),\quad k\in R, \tag{7.5}$$

$$(\partial_t\pm 2\omega(k_n^{\pm},t))c_n^{\pm}(t)=0,\quad \pm\operatorname{Im}k_n^{\pm}>0. \tag{7.6}$$

Up to this point everything holds for any kind of dispersion relation $\Omega=\omega\sigma_3$ in the evolution (2.4). Let us now examine the case already mentioned in Sec. V: instead of $\omega(k,t)$ we give $g(k,t)$ in

$$\frac{\partial}{\partial\bar k}\omega(k,t)=-i\pi g(k,t)\delta(k_I). \tag{7.7}$$

The dispersion relation is computed by using the Cauchy–Green formula [see for instance (2.9)]. Since $g(k,t)$ vanishes as $|k|\to\infty$ and is bounded in $k=0$, the integral along the

boundary ∂D (Fig. 1) vanishes, and the result is

$$\omega^{\pm}(k,t) = \mp \pi g(k,t) + i \not\int_{-\infty}^{+\infty} \frac{d\lambda}{\lambda - k} g(\lambda,t), \quad k \in \mathbf{R}, \tag{7.8}$$

where the slashed integral denotes the Cauchy principal-value integral. We have also

$$\omega(k_n,t) = +i \int_{-\infty}^{+\infty} \frac{d\lambda}{\lambda - k_n^{\pm}} g(\lambda,t). \tag{7.9}$$

Note that the one-soliton solution of Sec. V corresponds to $N^+ = N^- = 1$, $k_1^{\pm} = k^{\pm}$ given in (5.44), and $r^{\pm}(k,t) = 0$.

Using (7.8) in (7.5) shows that the evolution of the radiation component may present exponential growth or damping. For physical interpretation of such a phenomenon [when $g(k,t)$ is real and time independent], we refer to Ref. 3 when exponential damping (attenuator) occurs, and to Ref. 14 when exponential growth (amplifier) occurs.

Having here the freedom of choice of $g(k,t)$, one may think of a bounded evolution even for the radiation component. For instance, an oscillatory g provides not only solitons oscillating back and forth [see (5.34)] but also a radiation component that periodically gives (receives) energy to (from) the medium.

In the case corresponding to $\omega_{\bar{k}}$ given in (5.38) and to the evolution (5.39), the dispersion relation is given by

$$\omega(k,t) = i\pi[g_0(t)/(\lambda_0 - k)], \quad \lambda_0 \in \mathbf{R}. \tag{7.10}$$

The evolution of the radiation component implies that we have to impose

$$r^{\pm}(\lambda_0,t) = 0, \tag{7.11}$$

and reads

$$r^{\pm}(k,t) = r^{\pm}(k,0)\exp\left[\pm \frac{2i}{\lambda_0 - k} \int_0^t dt\, g_0(t) \right], \quad \lambda \neq \lambda_0. \tag{7.12}$$

Another way to obtain the soliton solution is to use the nonlinear superposition of elementary Darboux transformations obtained for $B_0 = \frac{1}{2}(1 \pm \sigma_3)$ in (6.1)–(6.3). This has been done in the review paper.[15]

The procedure developed here in 1 + 1 dimensions can be extended to evolutions in 2 + 1 dimensions.[16]

We note also that discontinuous solutions ("step-solitons") can be obtained for dispersion relations with moving singularities. The corresponding spectral transform consists of a rational reflection coefficient.[17]

[1] M. V. Tratnik and J. E. Sipe, Phys. Rev. Lett. **58**, 1104 (1987).
[2] V. E. Zakharov and A. V. Mikhailov, "Polarization domains in nonlinear optics," Pis'ma Zh. Eksp. Teor. Fiz. **6**, 279 (1987) (in Russian).
[3] G. L. Lamb, Jr., Phys. Rev. A **9**, 422 (1974); M. J. Ablowitz, D. J. Kaup, and A. C. Newell, J. Math. Phys. **15**, 1852 (1974). See also D. J. Kaup and A. C. Newell, Adv. Math. **31**, 67 (1979).
[4] D. J. Kaup, Phys. Rev. Lett. **59**, 2063 (1987).
[5] J. J.-P. Leon, Phys. Lett. A **123**, 65 (1987).
[6] J. J.-P. Leon, "Spectral transform and Bäcklund transformation for evolutions with singular dispersion relations," in *Nonlinear Evolutions*, edited by J. J.-P. Leon (World Scientific, Singapore, 1988).
[7] V. B. Matveev, M. A. Salle, and A. V. Rybin, "Darboux transformations and coherent interaction of the light pulse with two-level media," Hamburg preprint DESY 87-007, 1987.
[8] M. Boiti and G. Z. Tu, Nuovo Cimento B **71**, 253 (1982); M. Boiti, F. Pempinelli, and G. Z. Tu, Phys. Lett. A **93**, 107 (1983).
[9] The $\bar{\partial}$ approach has been used to solve nonlinear evolution equations in 2 + 1 dimensions: S. V. Manakov, Physica D **3**, 420 (1981); R. Beals and R. R. Coiffman, Commun. Pure Appl. Math. **37**, 39 (1984); Physica D **18**, 242 (1986); A. S. Fokas and M. J. Ablowitz, J. Math. Phys. **25**, 2494 (1984); M. Boiti, J. J.-P. Leon, M. Manna, and F. Pempinelli, Inv. Problems **3**, 25 (1987).
[10] The complete solution of the $\bar{\partial}$ approach ignoring totally the spectral problem is given by M. Jaulent and M. Manna, Phys. Lett. A **117**, 62 (1986); J. Math. Phys. **28**, 2338 (1987).
[11] L. A. Takhtajan, Phys. Lett. A **64**, 235 (1977).
[12] V. E. Zakharov and L. A. Takhtajan, Teor. Mat. Fiz. **38**, 26 (1979); V. S. Gerdjikov and A. B. Yanovski, Phys. Lett. A **103**, 232 (1984); A. V. Mikhailov, "Integrable magnetic models," in *Solitons*, edited by S. E. Trullinger, V. E. Zakharov, and V. L. Pokrovsky (Elsevier, B.V., 1986).
[13] See the second reference in Ref. 12.
[14] G. L. Lamb, Jr., Phys. Rev. A **12**, 2052 (1975); S. V. Manakov, Sov. Phys. JETP **56**, 37 (1982) [Zh. Eksp. Teor. Fiz. **83**, 68 (1982)].
[15] J. J.-P. Leon, "On the nonlinear evolution equations having nonanalytic dispersion relations," to appear in *Some Topics on Inverse Problems*, edited by P. C. Sabatier (World Scientific, Singapore, 1988).
[16] M. Boiti, J. J.-P. Leon, L. Martina, and F. Pempinelli, "Integrable nonlinear evolutions in 2 + 1 dimensions with nonanalytic dispersion relations," J. Phys. A (in press).
[17] J. J.-P. Leon, "Discontinuous soliton-like solution to the self-induced-transparency equations," Phys. Lett. A (in press).

A theory of amplitudes

Stanley Gudder
Department of Mathematics and Computer Science, University of Denver, Denver, Colorado 80208

(Received 3 April 1987; accepted for publication 27 April 1988)

A stochastic framework for quantum mechanics is presented in which the elements are measurements and amplitudes. The resulting structure provides three stochastic levels. At the bottom level are the sample points. These may be unobservable, in general. At the next level are the outcomes of measurements. The amplitude of an outcome is computed by summing the amplitudes of the sample points resulting in that outcome upon a measurement. At the highest level are the events of a measurement. The probability of an event is computed by summing the moduli squared of the amplitudes for the outcomes comprising that event. Using these guidelines, various probabilities, conditional probabilities, and expectations are defined. Superpositions of amplitude functions are investigated and superselection sectors are shown to occur in a natural way. The framework is illustrated in various mathematical models such as spin-$\frac{1}{2}$ models, the two-slit experiment, and phase space quantum mechanics. Finally, a theory of discrete Feynman amplitudes is presented

I. INTRODUCTION

In this work we present a framework for the foundations of quantum mechanics based upon the concept of amplitude functions. In the study of traditional quantum mechanics, two distinct pictures seem to emerge. One picture stems from what people say and the other picture stems from what they actually do. People say that the (pure) states of a quantum system are described by unit vectors in a Hilbert space H and the observables are described by the self-adjoint operators on H. If the state is given by $\psi \in H$ and the observable by T, then the probability that T has a value in a Borel set $A \in B(\mathbb{R})$ is $\langle E^T(A)\psi,\psi\rangle$ where E^T is the spectral measure for T. Moreover, the evolution of the state is given by Schrödinger's equation or its relativistic counterpart. However, in the modern theory of high energy physics, this is not what people actually do. Roughly speaking, they begin with an amplitude function (or amplitude density) f that is not known precisely but is given heuristically by a Feynman path integral. Using perturbation techniques involving Feynam diagrams and renormalization rules, f is found approximately. Using this method, the amplitude of a quantum outcome y is computed by integrating f over the alternatives that result in y. The absolute value squared of this quantity gives the probability of y. Using these probabilities, various physical quantities such as scattering cross sections, decay modes, and particle lifetimes can be computed.

There are also basic philosophical questions that have been present since the inception of quantum mechanics. These questions involve such concepts as completeness, reality, and hidden variables.[1] In our framework we have an underlying objective reality consisting of the set of all possible configurations of a physical system. This objective reality X depends upon the physical system being described. For example, X might be the set of trajectories in configuration space or the set of points in a phase space. An outcome y of a physical measurement F consists of a set of interfering alternatives represented by points in X. The amplitude $\psi(y)$ of y is found by summing (or integrating) an amplitude density over the alternatives resulting in y. Then $\psi(y)$ gives the "wave function" corresponding to F. We show that ψ is a unit vector in a Hilbert space H_F and that F can be represented by a self-adjoint operator on H_F. Using this method the traditional Hilbert space structure is obtained, but in the process information about the underlying reality X is lost. The space X gives a complete description of the physical system, while H_F only describes results that can be obtained from the measurement F. On X we may consider various measurements of the system simultaneously, while H_F only gives a "projection" of reality due to one measurement. As we shall illustrate, for a nonrelativistic system in which X is a phase space, quantum mechanics can be "derived" as an amplitude average of classical mechanics.

As pointed out by Pitowski,[2] there have been several attempts at accounting for the nonclassical deviant probabilities of quantum mechanics. One can take the nonrealist view of Bohr which denies that an object can be ascribed properties independent of observation. One can take the nonlocal view that explains quantum interference in terms of certain unknown physical mechanisms. There is the possibility that classical logic should be abandoned. Finally, one might abandon classical probability theory. We shall follow Feynman[3] and *choose the last of these alternatives*. Our guiding principle is that quantum probabilities are computed by summing (or integrating) amplitudes and then taking the modulus square.

II. MEASUREMENTS, AMPLITUDES, AND PROBABILITIES

Let X by a nonempty set. We call X a *sample space* and the elements of X are called *sample points*. A map F with domain $D(F) \subseteq X$ and range $Y_F = F(D)$ is a *measurement* (or *measure bundle*) on X if the following conditions hold;

(M1) Y_F is the base space of a measure space (Y_F, Σ_F, ν_F).

(M2) For every $y \in Y_F$, $F^{-1}(y)$ is the base space of a measure space $(F^{-1}(y), \Sigma_y, \mu_y)$.

We call $F^{-1}(y)$ the *fiber over* y, the elements of Y_F are called *F outcomes* and the sets in Σ_F are *F events*. Notice that

 0022-2488/88/092020-16$02.50

$$\mathscr{E}(F) \equiv \{F^{-1}(B):\ B \in \Sigma_F\}$$

is a σ algebra of subsets of $D(F) \subseteq X$.

The sample points correspond to the possible configurations of a physical system S. A measurement F corresponds to a laboratory procedure or experiment that can be performed on S and its domain $D(F)$ gives the sample points that are relevant to the procedure F. For every $x \in D(F)$, $F(x)$ denotes the outcome resulting from executing F when S has configuration x. For $y \in Y_F$, the fiber $F^{-1}(y)$ is the set of sample points that give the outcome y and for $B \in \Sigma_F$, $F^{-1}(B)$ is the set of sample points that give the event B. The measure ν_F is an *a priori* weight for the F events that is independent of the state of the system. In the case of total ignorance, ν_F is a uniform measure such as Lebesgue measure, Haar measure, or the counting measure in the discrete case. Similarly, μ_y is an *a priori* weight for the sample points in the fiber $F^{-1}(y)$.

A function $f: X \to \mathbb{C}$ is an *amplitude density* for the measurement F if the following conditions hold:

$$\text{(A1)}\quad f|F^{-1}(y) \in L^1(F^{-1}(y), \Sigma_y, \mu_y) \text{ for every } y \in Y_F;$$

$$\text{(A2)}\quad F(f)(y) \equiv \int_{F^{-1}(y)} f\,d\mu_y \in L^2(Y_F, \Sigma_F, \nu_F) \equiv H_F;$$

$$\text{(A3)}\quad \|F(f)\| = \int |F(f)|^2\,d\nu_F = 1.$$

We interpret $f(x)$ as the probability amplitude density of the configuration x. Since $F^{-1}(y)$ is interpreted as the set of alternatives that result in the outcome y upon execution of F, $F(f)(y)$ corresponds to "summing" the amplitudes over these alternatives. Then $F(f)(y)$ gives the amplitude density at y and the resulting probability density is $|F(f)(y)|^2$.

Let F be a measurement and let f be an amplitude density for F. A set $A \subseteq X$ is a *generalized* (F,f) *event* if the following conditions hold:

$$\text{(E1)}\quad A \cap F^{-1}(y) \in \Sigma_y \text{ for every } y \in Y_F;$$

$$\text{(E2)}\quad f_F(A)(y) \equiv \int_{A \cap F^{-1}(y)} f\,d\mu_y \in H_F.$$

We denote the set of generalized (F,f) events by $\mathscr{E}(F,f)$. For $A \in \mathscr{E}(F,f)$, we call $f_F(A)$ the (F,f) *amplitude density* of A. Notice that

$$f_F(X) = f_F(D(F)) = F(f).$$

The generalized events are the subsets of X for which a reasonable amplitude density can be defined. In fact, $f_F(A)(y)$ is the "sum" of the amplitudes over the alternatives in A that result in the outcome y upon execution of F. Interpreting $|f_F(A)|^2$ as the probability density of $A \in \mathscr{E}(F,f)$ it becomes reasonable to define the (F,f) *probability* of A as

$$P_{Ff}(A) = \int |f_F(A)|^2\,d\nu_F = \|f_F(A)\|^2.$$

We also define the (F,f) *pseudoprobability* of $A \in \mathscr{E}(F,f)$ as

$$P'_{Ff}(A) = \int f_F(A)\bar{f}_F(X)\,d\nu_F = \langle f_F(A), f_F(X)\rangle = \langle f_F(A), F(f)\rangle.$$

If $A \cap D(F) = \emptyset$, then clearly $f_F(A) = 0$. Hence there is no loss of generality in assuming that $A \subseteq D(F)$ for all $A \in \mathscr{E}(F,f)$. Thus we can identify X and $D(F)$. A nonempty collection of subsets of X is an *additive class* if $\mathscr{S}$ is closed under the formation of complements and finite disjoint unions. Moreover, if $\mathscr{S}$ is closed under the formation of countable disjoint unions, then $\mathscr{S}$ is a *σ-additive class*. We denote the complement of a set A by A^c and its characteristic function by χ_A.

Lemma 2.1: (a) $\mathscr{E}(F,f)$ is an additive class containing $\mathscr{E}(F)$.

(b) For every $B \in \Sigma_F$, $f_F[F^{-1}(B)] = \chi_B F(f)$ and

$$P_{Ff}[F^{-1}(B)] = P'_{Ff}[F^{-1}(B)] = \int_B |F(f)|^2\,d\nu_F.$$

(c) P'_{Ff} is an additive complex-valued set function on $\mathscr{E}(F,f)$ with $P'_{Ff}(X) = 1$.

(d) f_F is an additive vector-valued set function on $\mathscr{E}(F,f)$.

Proof: Clearly, $X \in \mathscr{E}(F,f)$ and

$$P'_{Ff}(X) = \|F(f)\|^2 = 1.$$

If $A \in \mathscr{E}(F,f)$, then for every $y \in Y_F$ we have

$$A^c \cap F^{-1}(y) = F^{-1}(y) \setminus A \cap F^{-1}(y) \in \Sigma_y.$$

Moreover,

$$\begin{aligned} f_F(A^c)(y) &= \int_{A^c \cap F^{-1}(y)} f\,d\mu_y \\ &= \int_{F^{-1}(y)} f\,d\mu_y - \int_{A \cap F^{-1}(y)} f\,d\mu_y \\ &= F(f)(y) - f_F(A)(y) \in H_F. \end{aligned}$$

Hence $\mathscr{E}(F,f)$ is closed under complementation. Let $A_i \in \mathscr{E}(F,f)$, $i = 1,\ldots,n$, be mutually disjoint. Then for every $y \in Y_F$ we have

$$(\cup A_i) \cap F^{-1}(y) = \cup (A_i \cap F^{-1}(y)) \in \Sigma_y.$$

Since the A_i are mutually disjoint

$$f_F(\cup A_i)(y) = \Sigma_i \int_{A_i \cap F^{-1}(y)} f\,d\mu_y = \Sigma_i f_F(A_i)(y) \in H_F.$$

Hence $\mathscr{E}(F,f)$ is an additive class. Also, it is now clear that f_F and P'_{Ff} are additive on $\mathscr{E}(F,f)$. If $B \in \Sigma_F$, then

$$F^{-1}(B) \cap F^{-1}(y) = \begin{cases} F^{-1}(y), & \text{if } y \in B, \\ \emptyset, & \text{if } y \notin B. \end{cases}$$

Hence $F^{-1}(B) \cap F^{-1}(y) \in \Sigma_y$ and

$$f_F[F^{-1}(B)] = \chi_B F(f) \in H_F.$$

Therefore $\mathscr{E}(F) \subseteq \mathscr{E}(F,f)$ and

$$\begin{aligned} P_{Ff}[F^{-1}(B)] = P'_{Ff}[F^{-1}(B)] &= \langle \chi_B F(f), F(f)\rangle \\ &= \int_B |F(f)|^2\,d\nu_F. \end{aligned}$$

□

We use the notation $F_{Ff}(B) \equiv P_{Ff}[F^{-1}(B)]$ for $B \in \Sigma_F$. It is clear that P_{Ff} is a probability measure on Σ_F and we call it the f *distribution* of F. Although P'_{Ff} is additive on $\mathscr{E}(F,f)$, it has the disadvantage of being complex valued so it cannot be interpreted as a probability for an arbitrary $A \in \mathscr{E}(F,f)$. On the other hand, P_{Ff} is non-negative, but it is not necessarily additive and it may attain values larger than

1 on $\mathscr{E}(F,f)$. Hence P_{Ff} is not a probability measure on $\mathscr{E}(F,f)$ in general. We now show that these difficulties can be overcome under certain conditions.

Let $\mathscr{S}\subseteq\mathscr{E}(F,f)$ be an additive class. We say that an amplitude density f is *F-orthogonally scattered over* $\mathscr{S}$ if for $A,B\in\mathscr{S}$ with $A\cap B=\emptyset$ we have $f_F(A)\perp f_F(B)$ [i.e., $\langle f_F(A),f_F(B)\rangle=0$]. We say that f is *F orthogonal* at $A\in\mathscr{E}(F,f)$ if $f_F(A)\perp f_F(A^c)$. Notice that if f is F-orthogonally scattered over $\mathscr{S}$, then f is F orthogonal at every $A\in\mathscr{S}$. Also, f is F orthogonal at A if and only if f is F-orthogonally scattered over $\{X,\emptyset,A,A^c\}$. When we say that μ is a probability measure on an additive class $\mathscr{S}$ we mean that μ is σ additive. That is, if $A_i\in\mathscr{S}$, $i=1,2,\ldots$, are mutually disjoint and $\cup A_i\in\mathscr{S}$, then $\mu(\cup A_i)=\Sigma_i\mu(A_i)$.

Theorem 2.2: Let f be an amplitude density for the measurement F.

(a) f is F-orthogonally scattered over $\mathscr{E}(F)$.

(b) For $A\in\mathscr{E}(F,f)$, $P_{Ff}(A)=P'_{Ff}(A)$ if and only if f is F orthogonal at A.

(c) If $\mathscr{S}\subseteq\mathscr{E}(F,f)$ is an additive class, the following statements are equivalent: (1) f is F orthogonal at every $A\in\mathscr{S}$; (2) $P_{Ff}(A)=P'_{Ff}(A)$ for every $A\in\mathscr{S}$; (3) P_{Ff} and P'_{Ff} are probability measures on $\mathscr{S}$.

(d) If f is F orthogonal at every $A\in\mathscr{E}(F,f)$, then $\mathscr{E}(F,f)$ is a σ-additive class and $P_{Ff}=P'_{Ff}$ is a probability measure on $\mathscr{E}(F,f)$.

Proof: (a) For $A,B\in\Sigma_F$, it follows for Lemma 2.1(b) that $f_F[F^{-1}(A)]=\chi_A F(f)$ and $f_F[F^{-1}(B)]=\chi_B F(f)$. Hence $A\cap B=\emptyset$ implies $f_F[F^{-1}(A)]\perp f_F[F^{-1}(B)]$,

(b) $P_{Ff}(A)=P'_{Ff}(A)$ if and only if

$$\begin{aligned}\|f_F(A)\|^2&=\langle f_F(A),f_F(X)\rangle\\&=\langle f_F(A),f_F(A)+f_F(A^c)\rangle\\&=\|f_F(A)\|^2+\langle f_F(A),f_F(A^c)\rangle.\end{aligned}$$

This last equation holds if and only if $f_F(A)\perp f_F(A^c)$.

(c) (1)$\Rightarrow$(2) follows from (b); (2)$\Rightarrow$(3). If (2) holds, then P_{Ff} is non-negative, additive, and $P_{Ff}(X)=1$. To show that P_{Ff} is σ additive, suppose $A_i\in\mathscr{S}$, $i=1,2,\ldots$, are mutually disjoint and $\cup A_i\in\mathscr{S}$. Since $f|F^{-1}(y)\in L^1(F^{-1}(y),\Sigma_y,\mu_y)$, for every $y\in Y_F$ we have

$$\begin{aligned}f_F(\cup A_i)(y)&=\int_{\cup\{A_i\cap F^{-1}(y)\}}f\,d\mu_y=\Sigma_i\int_{A_i\cap F^{-1}(y)}f\,d\mu_y\\&=\Sigma_i f_F(A_i)(y).\end{aligned}\tag{2.1}$$

Since P_{Ff} is additive, we have

$$\sum_{i=1}^{n}P_{Ff}(A_i)=P_{Ff}(\cup_{i=1}^{n}A_i)\leqslant 1.$$

Hence $\Sigma_{i=1}^{\infty}P_{Ff}(A_i)$ converges to a number that is not greater than 1. Moreover, $f_n\equiv\Sigma_{i=1}^{n}f_F(A_i)$ is Cauchy in H_F since

$$\begin{aligned}\left\|\sum_{i=m}^{n}f_F(A_i)\right\|^2&=\|f_F(\cup_{i=m}^{n}A_i)\|^2\\&=P_{Ff}(\cup_{i=m}^{n}A_i)=\sum_{i=m}^{n}P_{Ff}(A_i)\end{aligned}$$

and the last summation approaches 0 as $m,n\to\infty$. Thus f_n converges to an element $g\in H_F$ in the L^2 norm. It follows[4] that there exists a subsequence of f_n that converges to g a.e. $[\nu_F]$. Now from Eq. (2.1) we have $f_n\to f_F(\cup A_i)$ a.e. $[\nu_F]$ so $g=f_F(\cup A_i)$. Hence

$$\lim_{n\to\infty}\sum_{i=1}^{n}f_F(A_i)=f_F(\cup A_i)$$

in the L^2 norm. Therefore

$$\begin{aligned}\sum_{i=1}^{\infty}P_{Ff}(A_i)&=\lim_{n\to\infty}\sum_{i=1}^{n}P_{Ff}(A_i)=\lim_{n\to\infty}\left\|\sum_{i=1}^{n}f_F(A_i)\right\|^2\\&=\|f_F(\cup A_i)\|^2=P_{Ff}(\cup A_i).\end{aligned}$$

(3)$\Rightarrow$(1). Suppose (3) holds and $A\in\mathscr{S}$. Then

$$\begin{aligned}1=P_{Ff}(A\cup A^c)&=\|f_F(A)+f_F(A^c)\|^2\\&=\|f_F(A)\|^2+\|f_F(A^c)\|^2+2\,\mathrm{Re}\langle f_F(A),f_F(A^c)\rangle\\&=1+2\,\mathrm{Re}\langle f_F(A),f_F(A^c)\rangle.\end{aligned}$$

Hence

$$\mathrm{Re}\langle f_F(A),f_F(A^c)\rangle=0.$$

Moreover,

$$\begin{aligned}P'_{Ff}(A)&=\langle f_F(A),f_F(A)+f_F(A^c)\rangle\\&=\|f_F(A)\|^2+\langle f_F(A),f_F(A^c)\rangle.\end{aligned}$$

Thus $\langle f_F(A),f_F(A^c)\rangle$ is real so $f_F(A)\perp f_F(A^c)$.

(d) Suppose that f is F orthogonal at every $A\in\mathscr{E}(F,f)$. We know from Lemma 2.1(a) that $\mathscr{E}(F,f)$ is an additive class and from (c) that P_{Ff} is a probability measure on $\mathscr{E}(F,f)$. To show that $\mathscr{E}(F,f)$ is a σ-additive class, let $A_i\in\mathscr{E}(F,f)$, $i=1,2,\ldots$, be mutually disjoint. Now for every $y\in Y_F$ we have

$$(\cup A_i)\cap F^{-1}(y)=\cup\{A_i\cap F^{-1}(y)\}\in\Sigma_y.$$

Moreover Eq. (2.1) holds. Applying the same argument as in (c) we conclude that $f_F(\cup A_i)\in H_F$. Hence $\cup A_i\in\mathscr{E}(F,f)$. □

Corollary 2.3: Let $\mathscr{S}\in\mathscr{E}(F,f)$ be an additive class. If f is F orthogonal at every $A\in\mathscr{S}$, then f_F is a σ-additive vector-valued set function on $\mathscr{S}$.

If $A\in\mathscr{E}(F,f)$ with $f_F(A)\neq 0$ and $B\in\Sigma_F$ we define the *conditional* (F,f) *probability of B given A* by

$$P_{Ff}(B|A)\equiv\int_B|f_F(A)|^2\frac{d\nu_F}{P_{Ff}(A)}=\frac{\|\chi_B f_F(A)\|^2}{\|f_F(A)\|^2}.$$

Notice that $P_{Ff}(\cdot|A)$ is a probability measure on Σ_F. Although $P_{Ff}(A)$ cannot always be interpreted as a probability, we see that $P_{Ff}(B|A)$ can be so interpreted. We now show that $P_{Ff}(B|A)$ has the usual properties of a conditional probability.

Lemma 2.4: (a) $P_{Ff}(B|X)=P_{Ff}(B)$ for all $B\in\Sigma_F$.

(b) If $B\in\Sigma_F$, $A\in\mathscr{E}(F,f)$, then $F^{-1}(B)\cap A\in\mathscr{E}(F,f)$,

$$f_F[F^{-1}(B)\cap A]=\chi_B f_F(A)$$

and

$$P_{Ff}(B|A)=P_{Ff}[F^{-1}(B)\cap A]/P_{Ff}(A).$$

(c) If $B,C\in\Sigma_F$, then

$$P_{Ff}(B|F^{-1}(C))=P_{Ff}(B\cap C)/P_{Ff}(C).$$

Proof: (a) Applying Lemma 2.1(b) we have

$$P_{F_f}(B\,|X) = \|\chi_B F(f)\|^2/\|F(f)\|^2$$
$$= P_{F_f}[F^{-1}(B)] = P_{F_f}(B).$$

(b) Since

$$f_F[F^{-1}(B)\cap A\,](y) = \int_{F^{-1}(B)\cap A\cap F^{-1}(y)} f\,d\mu_y$$
$$= \chi_B(y)\int_{A\cap F^{-1}(y)} f\,d\mu_y$$
$$= \chi_B f_F(A)(y)\,,$$

we have

$$P_{F_f}[F^{-1}(B)\cap A\,] = \|\chi_B f_F(A)\|^2 = P_{F_f}(A)P_{F_f}(B\,|A).$$

(c) This follows from (b). □

We denote by $L^2(F,f)$ the set of functions $g\colon D(F)\to\mathbb{C}$ satisfying the following conditions:

(F1) $g^{-1}(B)\in\mathscr{E}(F,f)$ for every $B\in B(\mathbb{C})$;

(F2) $gf\,|F^{-1}(y)\in L^1(F^{-1}(y),\ \Sigma_y,\mu_y)$ for every $y\in Y_F$;

(F3) $f_F(g)(y)\equiv\int_{F^{-1}(y)} gf\,d\mu_F\in H_F$.

We call $f_F(g)$ the *amplitude density* of g. Notice that for $A\in\mathscr{E}(F,f)$ we have $\chi_A\in L^2(F,f)$ and $f_F(\chi_A)=f_F(A)$ so $f_F(g)$ generalizes the amplitude density of generalized events. We denote by $L^2_w(F,f)$ the set of functions $g\colon D(F)\to\mathbb{C}$ satisfying (F2) and (F3). It is clear that $L^2_w(F,f)$ is a linear space. However, $L^2(F,f)$ need not be linear since it is not necessarily closed under summation.[5]

For $g\in L^2_w(F,f)$ we define the (F,f) *pseudoexpectation* of g by

$$E_{F_f}(g) = \langle f_F(g), F(f)\rangle = \langle f_F(g), f_F(1)\rangle,$$

where $1=\chi_x$. It is clear that $E_{F_f}(g)$ is linear on $L^2_w(F,f)$ and $E_{F_f}(1)=1$. Since $E_{F_f}(\chi_A)=P'_{F_f}(A)$ for all $A\in\mathscr{E}(F,f)$, we see that E_{F_f} is the natural linear extension of P'_{F_f}. Unfortunately, $E_{F_f}(g)$ need not be real when g is real valued and $E_{F_f}(g)$ need not be non-negative when g is non-negative. We now show that these difficulties do not occur under certain conditions. Suppose $g\in L^2(F,f)$. Then

$$\mathscr{E}_g(F,f) = \{g^{-1}(B):\ B\in B(\mathbb{C})\}$$

is a σ subalgebra of $\mathscr{E}(F,f)$. If f is F orthogonal at every $A\in\mathscr{E}_g(F,f)$, then by Theorem 2.2(c), P_{F_f} is a probability measure on $\mathscr{E}_g(F,f)$ so $(g^{-1}(\mathbb{C}),\mathscr{E}_g(F,f),P_{F,f})$ becomes a probability space. We then define the Lebesgue integral $\int g(x)P_{F_f}(dx)$ in the usual way.

Theorem 2.5: (a) if $g\in L^2(F,f)$ and f is F orthogonal at every $A\in\mathscr{E}_g(F,f)$, then

$$E_{F_f}(g) = \int g(x)P_{F_f}(dx).$$

(b) If $g\colon Y_F\to\mathbb{C}$ is Σ_F measureable and $gF(f)\in H_F$, then $g\circ F$ satisfies the conditions of (a) and

$$E_{F_f}(g)\equiv E_{F_f}(g\circ F) = \int g(y)P_{F_f}(dy) = \langle gF(f), F(f)\rangle.$$

Proof: (a) Suppose g is a simple function $g=\Sigma c_i\chi_{A_i}$, $c_i\in\mathbb{C}$, $A_i\in\mathscr{E}_g(F,f)$, $i=1,\ldots,n$. Then by Theorem 2.2(c) we have

$$E_{F_f}(g) = \Sigma c_i E_{F_f}(\chi_{A_i}) = \Sigma c_i P'_{F_f}(A_i)$$
$$= \Sigma c_i P_{F_f}(A_i) = \int g(x)P_{F_f}(dx).$$

A standard limit argument now gives the result.

(b) It follows from Theorem 2.2(a) that $g\circ F$ satisfies the conditions of (a). Applying (a) and a change of variables we have

$$E_{F_f}(g\circ F) = \int g(y)P_{F_f}(dy).$$

Since

$$P_{F_f}(B) = \int_B |F(f)|^2\,d\nu_F$$

for all $B\in\Sigma_F$, we have $dP_{F_f}/d\nu_F = |F(f)|^2$. The Radon–Nikodym theorem gives

$$\int g(y)P_{F_f}(dy) = \int g(y)\frac{dP_{F_f}}{d\nu_F}\,d\nu_F = \int g|F(f)|^2\,d\nu_F$$
$$= \langle gF(f), F(f)\rangle.$$ □

Under the conditions of Theorem 2.5, E_{F_f} has the desirable properties of an expectation.

Corollary 2.6: Suppose f is F orthogonal at every $A\in\mathscr{E}_g(F,f)$ where $g\in L^2(F,f)$.

(a) If $g\colon D(F)\to\mathbb{R}$, then $E_{F_f}(g)\in\mathbb{R}$.

(b) If $g\colon D(F)\to\mathbb{R}^+$, then $E_{F_f}(g)\in\mathbb{R}^+$.

In certain quantum mechanical models we do not have the F-orthogonality condition of Theorem 2.5(a), so we must use the general form for E_{F_f} even though it does not always have satisfactory properties. [Alternatively, we might say that $E_{F_f}(g)$ does not exist or that g is not an observable function.] This same difficulty occurs in traditional quantum mechanics when one does not properly symmetrize a set of operators.

We can also define pseudoconditional expectations in a natural way. Let $A\in\mathscr{E}(F,f)$, $g\in L^2_w(F,f)$, and define

$$f_F(g|A)(y) = \int_{F^{-1}(y)\cap A} gf\,d\mu_y\,.$$

The (F,f) *pseudoconditional expectation of g given A* is defined as

$$E_{F_f}(g|A) = \frac{\langle f_F(g|A), f_F(1|A)\rangle}{P_{F_f}(A)}$$
$$= \frac{\langle f_F(g|A), f_F(A)\rangle}{\|f_F(A)\|^2}.$$

Notice if $g=\chi_{F^{-1}(B)}$ for $B\in\Sigma_F$, then

$$f_F(\chi_{F^{-1}(B)}|A) = \chi_{F^{-1}(B)}f_F(A)$$

and

$$E_{F_f}(\chi_{F^{-1}(B)}|A) = P_{F_f}(B\,|A).$$

III. CATALOGS AND AMPLITUDE SPACES

In Sec. II we considered single measurements on a sample space X. The present section discusses collections of measurements on X. For example, an important class of mea-

surements are those whose outcomes are real numbers. More precisely, an *observable* is a measurement F where $Y_F \in B(\mathbb{R})$ and $\Sigma_F = Y_F \cap B(\mathbb{R})$. In quantum theory it is usually essential to consider several observables in order to describe a physical system.

A *catalog* $\mathscr{A}(X)$ on a sample space X is a nonempty collection of measurements on X. There are two ways that one can proceed in applications of this theory. Given a sample space X and a collection of complex-valued functions $\mathscr{F} = \{f_i : i \in I\}$ on X, we can form the catalog $\mathscr{A}(X,\mathscr{F})$ (if it exists) of all measurements on X such that each f_i is an amplitude density for every $F \in \mathscr{A}(X,\mathscr{F})$. Or one can begin with a catalog $\mathscr{A}(X)$. A function $f: X \to \mathbb{C}$ is an *amplitude density* for $\mathscr{A}(X)$, if f is an amplitude density for every $F \in \mathscr{A}(X)$. Denote the set of all amplitude densities on $\mathscr{A}(X)$ by $\mathscr{F}(\mathscr{A})$. We say that a catalog $\mathscr{A}(X)$ is *complete* if there are sufficiently many measurements to distinguish different sample points. That is, for every $x \neq y$ in X there is an $F \in \mathscr{A}(X)$ such that $F(x) \neq F(y)$.

This framework generalizes the operational statistics framework of Foulis and Randall.[6] A catalog corresponds to a quasimanual in operational statistics and a measurement corresponds to an operation. A trivial example of a complete catalog is given by a measure space (Y,Σ,ν). Let $X = Y$ and let F: $X \to Y$ be the identity map. On $F^{-1}(y) = \{y\}$ let $\Sigma_y = \{\varnothing,\{y\}\}$ and define $\mu_y(\{y\}) = 1$. Moreover, let $(Y_F,\Sigma_F,\nu_F) = (Y,\Sigma,\nu)$. Then $\mathscr{A} = \{F\}$ is a complete catalog on X. Notice that $\mathscr{F}(\mathscr{A})$ is the unit sphere of the Hilbert space $L^2(Y,\Sigma,\nu)$.

Let $\mathscr{A}(X)$ be a catalog. The set of functions f: $X \to \mathbb{C}$ satisfying (A1) and (A2) of Sec. II for all $F \in \mathscr{A}(X)$ and

(A3') $\|f\| \equiv \|F(f)\|_{H_F} = \|G(f)\|_{H_G}$ for all $F,G \in \mathscr{A}(X)$

is called the *amplitude space* of $\mathscr{A}$ and is denoted $H(\mathscr{A})$. The elements of $H(\mathscr{A})$ are called *amplitude functions*. Notice that

$$H(\mathscr{A}) = \{\alpha f : \alpha \in \mathbb{C},\ f \in \mathscr{F}(\mathscr{A})\}$$

and if $f \in H(\mathscr{A})$ with $\|f\| \neq 0$, then $f/\|f\| \in \mathscr{F}(\mathscr{A})$. Of course, if $f \in H(\mathscr{A})$ then $\alpha f \in H(\mathscr{A})$ for every $\alpha \in \mathbb{C}$. However, if $f,g \in H(\mathscr{A})$ we need not have $f+g \in H(\mathscr{A})$. If $\alpha f + \beta g \in H(\mathscr{A})$, $f,g \in H(\mathscr{A})$, $\alpha,\beta \in \mathbb{C}$, we call $\alpha f + \beta g$ a *superposition* of f and g. We now characterize pairs $f,g \in H(\mathscr{A})$ for which superpositions are possible. For $f,g \in H(\mathscr{A})$ we write fsg if for every $F,G \in \mathscr{A}(X)$ we have

$$\langle F(f),F(g)\rangle_{H_F} = \langle G(f),G(g)\rangle_{H_G} .$$

Notice that s is a reflexive, symmetric relation and fsg implies $(\alpha f)s(\beta g)$ for all $\alpha,\beta \in \mathbb{C}$.

Lemma 3.1: Let $f,g \in H(\mathscr{A})$. Then fsg if and only if $f+g, f+ig \in H(\mathscr{A})$.

Proof: For $F \in \mathscr{A}(X)$ we have

$$\|F(f+g)\|^2 = \|F(f)\|^2 + \|F(g)\|^2 + 2\,\mathrm{Re}\langle F(f),F(g)\rangle . \tag{3.1}$$

Hence, if fsg, then $f+g \in H(\mathscr{A})$. Moreover, $fs(ig)$ so $f+ig \in H(\mathscr{A})$. Conversely, if $f+g \in H(\mathscr{A})$, then from Eq. (3.1) we have

$$\mathrm{Re}\langle F(f),F(g)\rangle = \mathrm{Re}\langle G(f),G(g)\rangle$$

for every F, $G \in \mathscr{A}(X)$. If in addition, $f+ig \in H(\mathscr{A})$, then since

$$\|F(f+ig)\|^2 = \|F(f)\|^2 + \|F(g)\|^2 + 2\,\mathrm{Im}\langle F(f),F(g)\rangle$$

we have

$$\mathrm{Im}\langle F(f),F(g)\rangle = \mathrm{Im}\langle G(f),G(g)\rangle$$

for every $F,G \in \mathscr{A}(X)$. Hence fsg. □

Corollary 3.2: For $f,g \in H(\mathscr{A})$ we have fsg if and only if $\alpha f + \beta g \in H(\mathscr{A})$, for every $\alpha,\beta \in \mathbb{C}$.

For $f,g \in \mathscr{F}(\mathscr{A})$, $F \in \mathscr{A}(X)$, the *F-transition amplitude of f to g* is

$$T_F(f,g) \equiv \langle F(f),F(g)\rangle_{H_F} .$$

Notice that fsg if and only if $T_F(f,g) = T_G(f,g)$ for every $F,G \in \mathscr{A}$.

For $A \subseteq H(\mathscr{A})$ we write

$$A^s = \{g \in H(\mathscr{A}) :\ gsf \text{ for all } f \in A\} .$$

A set $A \subseteq H(\mathscr{A})$ is a *superposition set* (or *s set*) if $A \subseteq A^s$. Every s set is contained in a maximal s set. Moreover, A is a maximal s set if and only if $A = A^s$ (Ref. 7). Denote the set of maximal s sets by $M(H)$. Let $A \in M(H)$. If $f \in A$, $\alpha \in \mathbb{C}$, then $(\alpha f)sg$ for every $g \in A$. Since A is maximal $\alpha f \in A$. If $f,g \in A$, then $f+g \in H(\mathscr{A})$ and $(f+g)sh$ for every $h \in A$. Again, by maximality, $f+g \in A$. Hence A is a linear space. For $f,g \in A$, define $\langle f,g\rangle = \langle F(f),F(g)\rangle$ for every $F \in \mathscr{A}$. Then it is clear that $\langle f,g\rangle$ is an inner product on A. Thus each $A \in M(H)$ forms an inner product space. Moreover, it is clear that for every $A \in M(H)$ and $F \in \mathscr{A}$, $F: A \to H_F$ is a linear isometry. The pair $(H(\mathscr{A}),s)$ gives a *partial inner product space*.[7] One can complete $(H(\mathscr{A}),s)$ in a natural way to obtain a partial Hilbert space. Such structures generalize direct sums of Hilbert spaces.[7] We call the sets $A \in M(H)$ *superselection sectors*.

We can apply our work in Sec. II and compute probabilities and expectations for a catalog $\mathscr{A} = \mathscr{A}(X)$. Let $F,G \in \mathscr{A}$ and let $f \in \mathscr{F}(\mathscr{A})$. If $A \in \Sigma_G$ and $G^{-1}(A) \in \mathscr{E}(F,f)$, then $P_{F,f}[G^{-1}(A)]$ is interpreted as the "probability" of the G event A upon execution of an F measurement. Operationally this means that if a G measurement is executed followed by an F measurement then $P_{F,f}[G^{-1}(A)]$ is the probability that G results in an outcome in A. If $A \in \Sigma_G$, $B \in \Sigma_F$, and $G^{-1}(A) \in \mathscr{E}(F,f)$ then $P_{F,f}(B\,|\,G^{-1}(A))$ is the probability that an F measurement results in an outcome in B given that a previous G measurement resulted in an outcome in A. If G is an observable with $G \in L^2_w(F,f)$, then $E_{F,f}(G)$ is interpreted as the "expectation" of G determined by an F measurement. In general, this differs from the ordinary expectation $E_{G,f}(G)$ of G.

If $f \in H(\mathscr{A})$, then we have seen that f can be represented by a vector $F(f)$ in the Hilbert space H_F. In particular, if $f \in \mathscr{F}(\mathscr{A})$, then $\|F(f)\| = 1$, so $F(f)$ is a traditional quantum state. Thus after an F measurement, the amplitude densities "reduce" to vectors in H_F. We now show that observables reduce to operators on H_F after an F measurement is executed. Let $F \in \mathscr{A}$ and suppose that g: $Y_F \to \mathbb{R}$ is injective, Σ_F measurable, and $gF(f) \in H_F$. Then $g \circ F$ is an observable and

$$f_F(g\circ F) = gF(f)\in H_F .$$

Moreover, $E_{F_f}(g\circ F) = \langle gF(f),F(f)\rangle$. Define the self-adjoint operator $\tilde{g}$ on H_F by $(\tilde{g}\psi)(y) = g(y)\psi(y)$. We then have $f_F(g\circ F) = \tilde{g}F(f)$ and $E_{F_f}(g\circ F) = \langle \tilde{g}F(f),F(f)\rangle$ for all $f\in\mathcal{F}(\mathcal{A})$ with $F(f)\in D(\tilde{g})$. We say that the observable $g\circ F$ is F *represented* by the operator $\tilde{g}$ and write $(g\circ F)^{\sim} = \tilde{g}$. In particular, if F itself is an observable, then letting g be the identity function $g(y) = y$ we have $\tilde{F} = \tilde{g}$. More generally, let $F\in\mathcal{A}$ and let G: $D(F)\to\mathbf{R}$. Define the transformation $\tilde{G}$ by setting

$$D_0(\tilde{G}) = \{F(f)\colon G\in L^2(F,f),\ f\in\mathcal{F}(\mathcal{A})\}\subseteq H_F$$

and for $F(f)\in D_0(\tilde{G})$, $\tilde{G}F(f) = f_F(G)$. Then for all $F(f)\in D_0(\tilde{G})$ we have $E_{F_f}(G) = \langle \tilde{G}F(f),F(f)\rangle$. Define

$$D(\tilde{G})\{F(f)\colon 0\neq f\in H(\mathcal{A}),\ F(f/\|f\|)\in D_0(\tilde{G})\}\cup\{0\} .$$

Now $\tilde{G}$ is easily extended to $D(\tilde{G})$ as follows: If $F(f')\in D(\tilde{G})$ then $f' = \alpha f$ for $\alpha\in\mathbf{C}$, $F(f)\in D_0(\tilde{G})$. Setting $\tilde{G}F(f') = \alpha\tilde{G}F(f)$ it is easily seen that $\tilde{G}$ is well defined on $D(\tilde{G})$. If $S\subseteq H(\mathcal{A})$ is a superselection sector we have seen that $FS\to H_F$ is a linear isometry. Then $D(\tilde{G})\cap F(S)$ is a subspace of H_F and $\tilde{G}$ is a linear operator on this subspace. Thus G is represented by the linear operator $\tilde{G}$ (depending on S). In general, $\tilde{G}$ is not essential self-adjoint even if its domain is dense. In particular, this work applies to observables.

IV. TWO-SPINS

In this section and the ones to follow, we shall illustrate our theory in terms of various physically relevant models. We begin with a model that corresponds to measuring the spin of a spin-$\frac{1}{2}$ particle in two different directions. Let $x_1 = \frac{1}{2}$, $x_2 = -\frac{1}{2}$ and form the sample space

$$X = \{(x_i,x_j)\colon\ i,j = 1,2\} .$$

Then X is the set of all possible configurations of the system where $(\frac{1}{2},\frac{1}{2})$ designates a spin of $\frac{1}{2}$ in both directions, $(\frac{1}{2},-\frac{1}{2})$ a spin of $\frac{1}{2}$ in the first direction, and $-\frac{1}{2}$ in the second direction, etc.

If F, G are the observables that measure the spin in each of the two directions, we have $F(x_i,x_j) = x_i$, $G(x_i,x_j) = x_j$. Then $Y_F = Y_G = \{\frac{1}{2},-\frac{1}{2}\}$ and assuming we have no prior knowledge about the spins we take $\Sigma_F = \Sigma_G = \{\varnothing,\{\frac{1}{2}\},\{-\frac{1}{2}\},Y_F\}$ and $\nu_F = \nu_G$ are the counting measures. Moreover,

$$F^{-1}(\tfrac{1}{2}) = \{(\tfrac{1}{2},\tfrac{1}{2}),(\tfrac{1}{2},-\tfrac{1}{2})\}$$

and

$$\mu_{F^{-1}(1/2)}$$

is the counting measure on $F^{-1}(\frac{1}{2})$. A similar definition applies to the other fibers. We now define the complete catalog $\mathcal{A} = \{F,G\}$. If $f\in\mathcal{F}(\mathcal{A})$ we can represent f by the 2×2 matrix M_f, where $(M_f)_{ij} = f(x_i,x_j)$, $i,j = 1,2$. We then have

$$M_f = \begin{bmatrix} a & b \\ c & d \end{bmatrix},\ a,b,c,d\in\mathbf{C} , \tag{4.1}$$

where

$$|a+b|^2 + |c+d|^2 = |a+c|^2 + |b+d|^2 = 1 . \tag{4.2}$$

Notice that

$$F(f) = (a+b,c+d)\in H_F,\quad G(f) = (a+c,b+d)\in H_G .$$

We now compute probabilities and expectations for $f\in\mathcal{F}(\mathcal{A})$ of the form (4.1). Since

$$P_{F_f}[F^{-1}(\tfrac{1}{2})] = |a+b|^2 ,$$
$$P_{F_f}[F^{-1}(-\tfrac{1}{2})] = |c+d|^2 ,$$
$$P_{G_f}[G^{-1}(\tfrac{1}{2})] = |a+c|^2 ,$$
$$P_{G_f}[G^{-1}(-\tfrac{1}{2})] = |b+d|^2 ,$$

P_{F_f} and P_{G_f} are indeed probability measures on $\mathcal{E}(F)$ and $\mathcal{E}(G)$, respectively. Since $f_F[G^{-1}(\frac{1}{2})] = (a,c)$, $f_F[G^{-1}(\frac{1}{2})] = (b,d)$ we have

$$P_{F_f}[G^{-1}(\tfrac{1}{2})] = |a|^2 + |c|^2 ,$$
$$P_{F_f}[G^{-1}(-\tfrac{1}{2})] = |b|^2 + |d|^2 .$$

In general, P_{F_f} is not a probability measure on $\mathcal{E}(G)$ since it need not be additive. Similarly,

$$P'_{F_f}[G^{-1}(\tfrac{1}{2})] = |a|^2 + |c|^2 + a\bar{b} + c\bar{d} ,$$
$$P'_{F_f}[G^{-1}(-\tfrac{1}{2})] = |b|^2 + |d|^2 + b\bar{a} + d\bar{c} .$$

In general, P'_{F_f} is not a probability measure on $\mathcal{E}(G)$ since it may have complex values. However, as we know, the conditional probability is a probability measure. These have the following values:

$$P_{F_f}(F^{-1}(\tfrac{1}{2})|G^{-1}(\tfrac{1}{2})) = |a|^2/(|a|^2 + |c|^2) ,$$
$$P_{F_f}(F^{-1}(-\tfrac{1}{2})|G^{-1}(\tfrac{1}{2})) = |c|^2/(|a|^2 + |c|^2) ,$$
$$P_{F_f}(F^{-1}(\tfrac{1}{2})|G^{-1}(-\tfrac{1}{2})) = |b|^2/(|b|^2 + |d|^2) ,$$
$$P_{F_f}(F^{-1}(-\tfrac{1}{2})|G^{-1}(-\tfrac{1}{2})) = |d|^2/(|b|^2 + |d|^2) .$$

To compute expectations, first note that

$$f_F(F) = (\tfrac{1}{2}(a+b),-\tfrac{1}{2}(c+d)) ,$$
$$f_F(G) = (\tfrac{1}{2}(a-b),\tfrac{1}{2}(c-d)) .$$

We then have from (4.2)

$$E_{F_f}(F) = \tfrac{1}{2}|a+b|^2 - \tfrac{1}{2}|c+d|^2 = |a+b|^2 - \tfrac{1}{2} .$$

This is reasonable since $-\frac{1}{2}\leqslant E_{F_f}(F)\leqslant\frac{1}{2}$; and in traditional quantum mechanics, one gets the same result. Moreover,

$$E_{F_f}(G) = \tfrac{1}{2}(a-b)(\bar{a}+\bar{b}) + \tfrac{1}{2}(c-d)(\bar{c}+\bar{d})$$
$$= \tfrac{1}{2} - (|b|^2 + |d|^2 + \bar{a}b + \bar{c}d) .$$

This is not very satisfactory since it may have complex values. Finally, we have

$$E_{F_f}(F|G^{-1}(\tfrac{1}{2})) = (|a|^2 - |c|^2)/2(|a|^2 + |c|^2)$$

and

$$E_{F_f}(F|G^{-1}(-\tfrac{1}{2})) = (|b|^2 - |d|^2)/2(|b|^2 + |d|^2) .$$

As with conditional probabilities, the conditional expectations have reasonable values.

To give a concrete example, suppose $f\in\mathcal{F}(\mathcal{A})$ has the form

$$M_f = \begin{bmatrix} \cos\theta/2 & -e^{i\theta/2} \\ -i\sin\theta/2 & e^{i\theta/2} \end{bmatrix} .$$

Then $P_{G_f}(\frac{1}{2}) = 1$, $P_{G_f}(-\frac{1}{2}) = 0$, so the spin is $\frac{1}{2}$ in the G

direction with certainty. We also have

$$P_{F_f}(\tfrac{1}{2}|G^{-1}(\tfrac{1}{2})) = \cos^2\theta/2\,,$$
$$E_{F_f}(F|G^{-1}(\tfrac{1}{2})) = \tfrac{1}{2}\cos\theta, \quad E_{G_f}(G) = \tfrac{1}{2}\,.$$

These are the usual quantum formulas when the angle between the two directions is θ. The first equation says that if a G measurement is made so that we are sure that the spin is $\frac{1}{2}$ in the G direction, then the probability that the spin is $\frac{1}{2}$ in the F direction is $\cos^2\theta/2$. However, if we do not perform a G measurement first, then we obtain

$$P_{F_f}(\tfrac{1}{2}) = \sin^2\theta/2, \quad E_{F_f}(F) = -\tfrac{1}{2}\cos\theta\,.$$

Moreover, we have the following unusual results:

$$P_{F_f}(\tfrac{1}{2}|G^{-1}(-\tfrac{1}{2})) = \tfrac{1}{2}, \quad P_{F_f}[G^{-1}(-\tfrac{1}{2})] = 2\,,$$
$$E_{F_f}(F|G^{-1}(-\tfrac{1}{2})) = 0, \quad E_{F_f}(G) = -\tfrac{1}{2}\,,$$
$$P'_{F_f}[G^{-1}(\tfrac{1}{2})] = 0, \quad P'_{F_f}[G^{-1}(-\tfrac{1}{2})] = 1\,.$$

We say that $f\in\mathcal{F}(\mathcal{A})$ is $\mathcal{A}$ *scattered* if f is F-orthogonally scattered over $\mathcal{E}(G)$ and is G-orthogonally scattered over $\mathcal{E}(F)$. It follows from Theorem 2.2(c) that f is $\mathcal{A}$ scattered if and only if $P_{F_f}, P'_{F_f}, P_{G_f}, P'_{G_f}$ are probability measures on $\mathcal{E}(F)$ and $\mathcal{E}(G)$. It is therefore of interest to characterize the $\mathcal{A}$-scattered amplitude densities.

Theorem 4.1: An $f\in\mathcal{F}(\mathcal{A})$ is $\mathcal{A}$ scattered if and only if, except for a multiplicative factor of modulus 1, M_f has one of the following forms:

$$\begin{bmatrix}1&0\\0&0\end{bmatrix}, \quad \begin{bmatrix}0&1\\0&0\end{bmatrix}, \quad \begin{bmatrix}0&0\\1&0\end{bmatrix}, \quad \begin{bmatrix}0&0\\0&1\end{bmatrix},$$
$$\begin{bmatrix}ae^{i\theta}&0\\0&(1-a^2)^{1/2}\end{bmatrix}, \quad \begin{bmatrix}0&ae^{i\theta}\\(1-a^2)^{1/2}&0\end{bmatrix}, \quad 0\leqslant\theta<2\pi, \quad 0<a<1\,,$$
$$\begin{bmatrix}ae^{i\theta}&(\frac{1}{2}-a^2)^{1/2}e^{i\phi}\\(\frac{1}{2}-a^2)^{1/2}e^{-i\phi}&-ae^{-i\theta}\end{bmatrix}, \quad \begin{bmatrix}ae^{i\theta}&(\frac{1}{2}-a^2)^{1/2}e^{i\phi}\\-(\frac{1}{2}-a^2)^{1/2}e^{-i\phi}&ae^{-i\theta}\end{bmatrix}, \quad 0\leqslant\theta,\ \phi<2\pi, \quad 0<a<1/\sqrt{2}\,.$$

Proof: It follows from Theorem 2.2(c) that f is $\mathcal{A}$ scattered if and only if $P_{F_f}=P'_{F_f}$ on $\mathcal{E}(G)$ and $P_{G_f}=P'_{G_f}$ on $\mathcal{E}(F)$. Applying our above computations we conclude that f is $\mathcal{A}$ scattered if and only if the entries in M_f satisfy

$$a\bar{b}+c\bar{d}=0, \quad a\bar{c}+b\bar{d}=0\,. \tag{4.3}$$

It is clear that all considerations are independent of a multiplicative factor of modulus 1, so we shall ignore such factors. It is easy to check that if M_f has one of the above forms, then its entries satisfy (4.3). Conversely, suppose $f\in\mathcal{F}(\mathcal{A})$ and the entries of M_f satisfy (4.3). If one of the entries in M_f is 0, then either all the entries, except one, are 0 or two diagonal entries are nonzero. In the first case we have the first four forms and in the second case we have the next two forms. Now suppose all the entries are nonzero. We can assume without loss of generality that $a>0$. From the first equation of (4.3) we have $c=-a\bar{b}/\bar{d}$. The second equation then gives $|d|=a$. Hence, $d=ae^{i\alpha}$ for some $0\leqslant\alpha<2\pi$. Then $c=-\bar{b}e^{i\alpha}$. We can then assume that M_f has the form

$$M_f=\begin{bmatrix}a&b\\-\bar{b}e^{i\alpha}&ae^{i\alpha}\end{bmatrix}.$$

Applying (4.2) gives $|b|=(\frac{1}{2}-a^2)^{1/2}$, so $b=(\frac{1}{2}-a^2)^{1/2}e^{i\beta}$ for some $0\leqslant\beta<2\pi$. Multiplying by a factor of modulus 1 gives the last form. Proceeding as before with $d=-ae^{i\alpha}$ gives the next to last form. □

The probabilities and expectations given by the first six forms in Theorem 4.1 are trivial. Using our previous formulas, we obtain the following probabilities and expectations for the last form in Theorem 4.1:

$$P_{F_f}[F^{-1}(\tfrac{1}{2})] = \tfrac{1}{2}+2a(\tfrac{1}{2}-a^2)^{1/2}\cos(\theta-\phi)\,,$$
$$P_{F_f}[F^{-1}(-\tfrac{1}{2})] = \tfrac{1}{2}-2a(\tfrac{1}{2}-a^2)^{1/2}\cos(\theta-\phi)\,,$$
$$P_{G_f}[G^{-1}(\tfrac{1}{2})] = \tfrac{1}{2}-2a(\tfrac{1}{2}-a^2)^{1/2}\cos(\theta+\phi)\,,$$
$$P_{G_f}[G^{-1}(-\tfrac{1}{2})] = \tfrac{1}{2}+2a(\tfrac{1}{2}-a^2)^{1/2}\cos(\theta+\phi)\,,$$
$$P_{F_f}[G^{-1}(\tfrac{1}{2})] = P_{F_f}[G^{-1}(-\tfrac{1}{2})] = \tfrac{1}{2}\,,$$
$$P_{F_f}(F^{-1}(\tfrac{1}{2})|G^{-1}(\tfrac{1}{2})) = P_{F_f}(F^{-1}(-\tfrac{1}{2})|G^{-1}(-\tfrac{1}{2})) = 2a^2\,,$$
$$P_{F_f}(F^{-1}(\tfrac{1}{2})|G^{-1}(-\tfrac{1}{2})) = P_{F_f}(F^{-1}(-\tfrac{1}{2})|G^{-1}(\tfrac{1}{2})) = 1-2a^2\,,$$
$$E_{F_f}(F) = 2a(\tfrac{1}{2}-a^2)^{1/2}\cos(\theta-\phi)\,,$$
$$E_{F_f}(G) = 0\,,$$
$$E_{F_f}(F|G^{-1}(\tfrac{1}{2})) = 2a^2-\tfrac{1}{2}\,,$$
$$E_{F_f}(F|G^{-1}(-\tfrac{1}{2})) = \tfrac{1}{2}-2a^2\,.$$

Another interesting class of amplitude densities is those that have superpositions with all others. We say that $f\in\mathcal{F}(\mathcal{A})$ is *central* if $f\in\mathcal{F}(\mathcal{A})^s$. The corollary to the next result characterizes central amplitude densities.

Theorem 4.2: Let $f,g\in\mathcal{F}(\mathcal{A})$ and suppose

$$M_f=\begin{bmatrix}a&b\\c&d\end{bmatrix}, \quad M_g=\begin{bmatrix}a'&b'\\c'&d'\end{bmatrix}.$$

Then fsg if and only if

$$(a-d)(\bar{b}'-\bar{c}')+(b-c)(\bar{a}'-\bar{d}')=0\,. \tag{4.4}$$

Proof: The following statements are equivalent

(1) fsg,
(2) $\langle F(f),F(g)\rangle=\langle G(f),G(g)\rangle$,
(3) $\langle(a+b,c+d),(a'+b',c'+d')\rangle$
$\quad=\langle(a+c,b+d),(a'+c',b'+d')\rangle$,
(4) Eq. (4.4). □

Corollary 4.3: An $f\in\mathcal{F}(\mathcal{A})$ is central if and only if M_f

has the form

$$M_f = \begin{bmatrix} a & b \\ b & a \end{bmatrix}, \quad |a+b|^2 = \frac{1}{2}.$$

Proof: If M_f has the above form, then clearly (4.4) holds for every $g \in \mathcal{F}(\mathcal{A})$, so f is central. Conversely, if f is central, then (4.4) holds for every $g \in \mathcal{F}(\mathcal{A})$. If g has the first form in Theorem 4.1, we conclude that $b = c$. If g has the second form in Theorem 4.1, we conclude that $a = d$. □

Notice that Corollary 4.3 states that f is central if and only if $f(\frac{1}{2},\frac{1}{2}) = f(-\frac{1}{2},-\frac{1}{2})$ and $f(\frac{1}{2},-\frac{1}{2}) = f(-\frac{1}{2},\frac{1}{2})$. Thus centrality is equivalent to a symmetry between spin $\frac{1}{2}$ and spin $-\frac{1}{2}$. This is also evident in the next characterization.

Corollary 4.4: An $f \in \mathcal{F}(\mathcal{A})$ is central if and only if $F(f) = G(f)$ and $F(f)(\frac{1}{2}) = F(f)(-\frac{1}{2})$.

Proof: Observe that $F(f) = G(f)$ is equivalent to $b = c$ and then $F(f)(\frac{1}{2}) = F(f)(-\frac{1}{2})$ is equivalent to $a = d$. □

Notice that $F(f) = G(f)$ means that f does not distinguish between the two spin directions. Moreover, $F(f)(\frac{1}{2}) = F(f)(-\frac{1}{2})$ means that f shows no preference between spin $\frac{1}{2}$ and spin $-\frac{1}{2}$.

Note that an $f \in \mathcal{F}(\mathcal{A})^s$ is not $\mathcal{A}$ scattered except for trivial cases. Also, since $F(f)(\frac{1}{2}) = F(f)(-\frac{1}{2})$ for $f \in \mathcal{F}(\mathcal{A})^s$, it follows that span $\mathcal{F}(a)^s$ is a one-dimensional subspace of every superselection sector of $H(\mathcal{A})$. A typical $f \in \mathcal{F}(\mathcal{A})^s$ has the form

$$M_f = [2(1+\sin\theta)]^{-1/2} \begin{bmatrix} \cos\theta/2 & \sin\theta/2 \\ \sin\theta/2 & \cos\theta/2 \end{bmatrix}, \qquad (4.5)$$

where $0 \leqslant \theta \leqslant \pi$. This f corresponds to an experiment in which the angle between the two directions is θ. Using an f with the form (4.5) we obtain the following probabilities and expectations:

$$P_{Ff}[F^{-1}(\tfrac{1}{2})] = P_{Ff}[F^{-1}(-\tfrac{1}{2})] = \tfrac{1}{2},$$
$$P_{Gf}[G^{-1}(\tfrac{1}{2})] = P_{Gf}[G^{-1}(-\tfrac{1}{2})] = \tfrac{1}{2},$$
$$P_{Ff}(F^{-1}(\tfrac{1}{2})|G^{-1}(\tfrac{1}{2})) = \cos^2\theta/2,$$
$$P_{Ff}(F^{-1}(-\tfrac{1}{2})|G^{-1}(\tfrac{1}{2})) = \sin^2\theta/2,$$
$$E_{Ff}(F) = E_{Gf}(G) = 0,$$
$$E_{Ff}(F|G^{-1}(\tfrac{1}{2})) = -E_{Ff}(F|G^{-1}(-\tfrac{1}{2})) = \tfrac{1}{2}\cos\theta.$$

These are the usual formulas of traditional quantum mechanics for this situation. The following formulas do not have traditional counterparts. Notice that the first of these does not make sense as a probability,

$$P_{Ff}[G^{-1}(\tfrac{1}{2})] = P_{Ff}[G^{-1}(-\tfrac{1}{2})] = 1/2(1+\sin\theta),$$
$$P'_{Ff}[G^{-1}(\tfrac{1}{2})] = P'_{Ff}[G^{-1}(-\tfrac{1}{2})] = \tfrac{1}{2},$$
$$E_{Ff}(G) = E_{Gf}(F) = 0.$$

V. DIRECTION-SPIN

Let S^2 be the unit sphere in $\mathbb{R}^3$ and form the sample space $X = S^2 \times \{\frac{1}{2}, -\frac{1}{2}\}$. If we think of a point in S^2 as designating a direction in space, then X gives the set of all possible spin configurations for a spin-$\frac{1}{2}$ particle. Define F: $X \to \{\frac{1}{2}, -\frac{1}{2}\}$ by $F(u,s) = s$. Let Σ_F be the power set on Y_F and let ν_F be the counting measure. Endow the fibers $F^{-1}(\pm\frac{1}{2}) = S^2 \times \{\pm\frac{1}{2}\}$ with their natural Borel sets $\Sigma_{\pm 1/2}$ and Lebesgue measures $\mu_{\pm 1/2}$. Then F becomes a measurement on X that we interpret as the spin measurement. Define G: $X \to S^2$ by $G(u,s) = u$. Let $\Sigma_G = B(S^2)$ and let $\nu_G = b\,dx$, where dx is Lebesgue measure and $b = 128\pi/9(4+\pi)$ is a normalization constant. Let Σ_u be the power set on the fiber

$$G^{-1}(u) = \{(u,\tfrac{1}{2}),(u,-\tfrac{1}{2})\}$$

and let μ_u be the counting measure. We interpret the measurement G as the direction measurement. We now form the complete catalog $\mathcal{A} = \{F,G\}$ on X.

If $f \in \mathcal{F}(\mathcal{A})$, then $f(u,s)$ is the amplitude density that the spin is s in the direction u. In the last section we gave a fairly complete analysis of the two-spin experiment. For brevity, we shall now only consider amplitude densities corresponding to traditional quantum states and shall only compute the usual probabilities and expectations. A direction $u \in S^2$ is specified by a pair (θ,ϕ), where $0 \leqslant \theta \leqslant \pi$ is the polar angle and $0 \leqslant \phi < 2\pi$ is the azimuthal angle. Define f: $X \to \mathbb{C}$ by $f(u,\frac{1}{2}) = a\cos\theta/2$, $f(u,-\frac{1}{2}) = a\sin\theta/2$, where $u = (\theta,\phi)$, and $a = 3\sqrt{2}/16\pi$ is a normalization constant. It is easy to check that $f \in \mathcal{F}(\mathcal{A})$. We can interpret f as the amplitude density for a particle that initially has spin $\frac{1}{2}$ in the (0,0,1) direction.

For $0 < \theta < \pi$, let $0 < \epsilon < \min(\theta, \pi, -\theta)$ and define

$$B_{\theta,\epsilon} = \{(\theta',\phi) \in S^2 : \theta - \epsilon \leqslant \theta' \leqslant \theta + \epsilon\}.$$

Then $B_{\theta,\epsilon} \in \Sigma_G$ and $A_{\theta,\epsilon} = G^{-1}(B_{\theta,\epsilon}) \in \mathcal{E}(F,f)$. By defining $A'_{\theta,\epsilon} = A_{\theta,\epsilon} \cap F^{-1}(\frac{1}{2})$ we have

$$f_F(A_{\theta,\epsilon})\left(\frac{1}{2}\right) = \int_{A'_{\theta,\epsilon}} f\,d\mu_{1/2} = a\int_{A'_{\theta,\epsilon}} \cos\frac{\theta}{2}\,d\mu_{1/2}.$$

By applying the integral mean value theorem, there exists a θ' with $\theta - \epsilon \leqslant \theta' \leqslant \theta + \epsilon$ such that

$$f_F(A_{\theta,\epsilon})(\tfrac{1}{2}) = a(\cos\theta'/2)\mu_{1/2}(A'_{\theta,\epsilon}).$$

Using a similar method we have

$$f_F(A_{\theta,\epsilon})(-\tfrac{1}{2}) = a(\sin\theta''/2)\mu_{1/2}(A'_{\theta,\epsilon}),$$

where $\theta - \epsilon \leqslant \theta'' \leqslant \theta + \epsilon$. Hence

$$P_{Ff}(A_{\theta,\epsilon}) = \|f_F(A_{\theta,\epsilon})\|^2$$
$$= a^2[\mu_{1/2}(A'_{\theta,\epsilon})]^2(\cos^2\theta'/2 + \sin^2\theta''/2).$$

We then obtain

$$P_{Ff}(\tfrac{1}{2}|A_{\theta,\epsilon}) = [f_F(A_{\theta,\epsilon})(\tfrac{1}{2})]^2 \big/ P_{Ff}(A_{\theta,\epsilon})$$
$$= (\cos^2\theta'/2)/(\cos^2\theta'/2) + (\sin^2\theta''/2).$$

By defining

$$P_{Ff}(\tfrac{1}{2}|\theta) = \lim_{\epsilon \to 0} P_{Ff}(\tfrac{1}{2}|A_{\theta,\epsilon}),$$

we obtain

$$P_{Ff}(\tfrac{1}{2}|\theta) = \cos^2\theta/2.$$

Similarly,

$$P_{Ff}(-\tfrac{1}{2}|\theta) = \sin^2\theta/2.$$

These are the usual results in traditional quantum mechanics.

Moreover,

$$f_F(F|A_{\theta,\epsilon})(\tfrac{1}{2}) = \tfrac{1}{2} f_F(A_{\theta,\epsilon})(\tfrac{1}{2}),$$
$$f_F(F|A_{\theta,\epsilon})(-\tfrac{1}{2}) = -\tfrac{1}{2} f_F(A_{\theta,\epsilon})(-\tfrac{1}{2}).$$

Hence

$$\begin{aligned} E_{Ff}(F|A_{\theta,\epsilon}) &= \langle f_F(F|A_{\theta,\epsilon}), f_F(A_{\theta,\epsilon})\rangle / P_{Ff}(A_{\theta,\epsilon}) \\ &= (\cos^2\theta'/2 - \sin^2\theta''/2)/2 \\ &\quad \times (\cos^2\theta'/2 + \sin^2\theta''/2). \end{aligned}$$

Defining

$$E_{Ff}(F|\theta) = \lim_{\epsilon\to 0} E_{Ff}(F|A_{\theta,\epsilon}),$$

we obtain

$$E_{Ff}(F|\theta) = \tfrac{1}{2}(\cos^2\theta/2 - \sin^2\theta/2) = \tfrac{1}{2}\cos\theta.$$

Again, this is the traditional result.

VI. TWO-SLIT EXPERIMENT

This section presents a simple model for the two-slit experiment. There is a source s of identical noninteracting particles. To the right of s there is a screen with two slits followed by a detection screen. If a particle moving to the right from s is not absorbed by the first screen, it goes through one of the two slits and is eventually absorbed at a point x on the detection screen. For simplicity, we assume that the screens are one dimensional and infinite in extent. Let the origin O of the detection screen be halfway between the two slits and let x_1, x_2 be points on this screen directly to the right of slit 1 and slit 2, respectively.

We assume that if a particle is absorbed at x then it has only two possible paths from s to x; one path goes through slit 1 and the other through slit 2. The set of all relevant paths is described by the sample space

$$X = \{(x,j): x\in\mathbb{R}, j = 1,2\}.$$

The sample point (x,j) designates the path from s to x going through slit j, $j = 1,2$. Define $F: X\to\mathbb{R}$ by $F(x,j) = x$. Let $\Sigma_F = B(\mathbb{R})$ and let ν_F be the Lebesgue measure dx. On the fiber $F^{-1}(x)$, let μ_x be the counting measure. The measurement F gives the absorption location. Define $G: X\to\{1,2\}$ by $G(x,j) = j$. Let ν_G be the counting measure on the power set of $\{1,2\}$. On the fiber $G^{-1}(j)$, let $\Sigma_j = B(\mathbb{R})\times\{j\}$ and let $\mu_j = b\,dx, j = 1,2$, where $b>0$ is a normalization given by

$$b^2 = (2\pi a_1)^{-1/2}|a|[1 + \exp[-(|a|d)^2/2a_1]].$$

The letters a, a_1, d denote geometric and physical constants that are determined by the experiment and are assumed to be known. In fact, $d = |x_1 - x_2|$ and $a = a_1 + ia_2\in\mathbb{C}$, $a_1, a_2 > 0$. The constant a_1 depends on the various dimensions in the experiment and a_2 depends on the frequency (momentum) of the particles. Of course, the measurement G gives the slit through which a path travels. We now form the complete catalog $\mathscr{A} = \{F,G\}$ on X.

Let $f: X\to\mathbb{C}$ be defined by

$$f(x,j) = c\exp[-a(x - x_j)^2],$$

where $c>0$ is the normalization constant given by

$$c^2 = |a|/2\pi b^2.$$

It is straightforward to show that $f\in\mathscr{F}(\mathscr{A})$. We shall see that f gives distributions that one expects for this experiment. Since

$$\begin{aligned} F(f)(x) = c[&\exp[-a(x - x_1)^2] \\ &+ \exp[-a(x - x_2)^2]], \end{aligned}$$

the probability density for F becomes

$$\begin{aligned} &|F(f)(x)|^2 \\ &= c^2\{\exp[-2a_1(x - x_1)^2] + \exp[-2a_1(x - x_2)^2] \\ &\quad + 2\exp[-a_1(x - x_1)^2]\exp[-a_1(x - x_2)^2] \\ &\quad \times\cos a_2[(x - x_1)^2 - (x - x_2)^2]\}. \end{aligned}$$

This gives the typical interference pattern for the two-slit experiment. For the measurement G we have $G(f)(j) = (|a|/2a)^{1/2}, j = 1,2$. Hence

$$P_{Gf}(1) = P_{Gf}(2) = |G(f)(1)|^2 = \tfrac{1}{2},$$

so the particle travels through each slit with equal likelihood.

Since

$$f_F[G^{-1}(1)](x) = f(x,1) = c\exp[-a(x - x_1)^2],$$

we have

$$\begin{aligned} P_{Ff}[G^{-1}(1)] &= c^2\int \exp[-2a_1(x - x_1)^2]dx \\ &= \tfrac{1}{2}[1 + \exp[-(|a|d)^2/2a_1]]^{-1}. \end{aligned}$$

Similarly, $P_{Ff}[G^{-1}(2)]$ has this same value so we do not obtain a probability measure, except in the classical limit $d\to\infty$ or $|a|^2/2a_1\to\infty$. However, for $B\in\Sigma_F$, $P_{Ff}(B\,|G^{-1}(1))$ does give a probability distribution. This is given by

$$\begin{aligned} P_{Ff}(B\,|G^{-1}(1)) &= \int_B \frac{|f_F[G^{-1}(1)](x)|^2\,dx}{P_{Ff}[G^{-1}(1)]} \\ &= \left(\frac{2a_1}{\pi}\right)^{1/2}\int_B \exp[-2a_1(x - x_1)^2]dx. \end{aligned}$$

This is the classical Gaussian distribution obtained by closing the second slit.

Again, if $B\in\Sigma_F$ we have

$$f_G[F^{-1}(B)](1) = bc\int_B \exp[-a(x - x_1)^2]dx$$

and

$$f_G[F^{-1}(B)](2) = bc\int_B \exp[-a(x - x_2)^2]dx.$$

Hence

$$\begin{aligned} &P_{Gf}[F^{-1}(B)] \\ &\quad = |f_G[F^{-1}(B)](1)|^2 + |f_G[F^{-1}(B)](2)|^2 \\ &\quad = \frac{|a|}{2\pi}\left[\left|\int_B \exp[-a(x - x_1)^2]dx\right|^2\right. \\ &\qquad \left. + \left|\int_B \exp[-a(x - x_2)^2]\right|^2\right]. \end{aligned}$$

Notice that this does not give a probability measure. However,

$$P_{G_f}(1|F^{-1}(B)) = |f_G[F^{-1}(B)](1)|^2/P_{G_f}[F^{-1}(B)]$$

$$= \frac{|\int_B \exp[-a(x-x_1)^2]dx|^2}{|\int_B \exp[-a(x-x_1)^2]dx|^2 + |\int_B \exp[-a(x-x_2)^2]dx|^2}.$$

A similar expression holds for $P_{G_f}(2|F^{-1}(B))$ and $P_{G_f}(\cdot|F^{-1}(B))$ gives a probability distribution.

If the reader finds the normalization constant b in this and the previous section unnatural, it can be eliminated. Then f becomes an amplitude density for F, but not for G. However, all the (F,f) probabilities, conditional probabilities and expectations can still be computed.

VII. PHASE SPACE MODEL

This section presents an amplitude phase space model for a simple quantum system. The system consists of a single nonrelativistic, spinless particle constrained to one dimension. (The model can easily be generalized to three dimensions.) We take for our sample space the two-dimensional phase space

$$X = \mathbb{R}^2 = \{(q,p): q,p \in \mathbb{R}\}.$$

Define the measurements $Q: X \to \mathbb{R}$, $P: X \to \mathbb{R}$ by $Q(q,p) = q$, $P(q,p) = p$, and let $\Sigma_Q = \Sigma_P = B(\mathbb{R})$, $dv_Q = dq$, $dv_p = dp$. On the fiber $Q^{-1}(q) = q \times \mathbb{R}$ we let $\Sigma_q = q \times B(\mathbb{R})$, $du_q = dp$, and on $P^{-1}(p) = \mathbb{R} \times p$ we let $\Sigma_p = B(\mathbb{R}) \times p$, $du_p = dq$. Of course, Q,P correspond to position and momentum measurements, respectively. Then $\mathscr{A} = \{Q,P\}$ is a complete catalog on X. If $f \in \mathscr{F}(\mathscr{A})$ we have

$$Q(f)(q) = \int f(q,p)dp \in L^2(\mathbb{R},dq),$$

$$P(f)(p) = \int f(q,p)dq \in L^2(\mathbb{R},dp)$$

and $\|Q(f)\| = \|P(f)\| = 1$.

Simple examples of amplitude densities can be constructed as follows. Let

$$\psi,\phi \in L^2(\mathbb{R},dx) \cap L^1(\mathbb{R},dx)$$

with

$$\int |\psi|^2 \, dx = \int |\phi|^2 \, dx = \int \psi \, dx = \int \phi \, dx = 1.$$

Then $f(q,p) = \psi(q)\phi(p)$ is an amplitude density for $\mathscr{A}$. We then have $Q(f)(q) = \psi(q)$ and $P(f)(p) = \phi(p)$. Therefore, for $A \in \Sigma_Q$, $B \in \Sigma_p$ we have

$$P_{Q_f}(A) = P'_{Q_f}(A) = \int |\psi(q)|^2 \, dq,$$

$$P_{P_f}(B) = P'_{P_f}(B) = \int_B |\phi(p)|^2 \, dp.$$

Moreover, since

$$f_Q[P^{-1}(B)](q) = \psi(q)\int_B \phi(p)dp,$$

$$f_p[Q^{-1}(A)](p) = \phi(p)\int_A \psi(q)dq,$$

we have

$$P_{Q_f}[P^{-1}(B)] = \left|\int_B \phi(p)dp\right|^2,$$

$$P_{P_f}[Q^{-1}(A)] = \left|\int_A \psi(q)dq\right|^2,$$

$$P'_{Q_f}[P^{-1}(B)] = \int_B \phi(p)dp,$$

$$P'_{P_f}[Q^{-1}(A)] = \int_A \psi(q)dq.$$

In general, we see that P_{Q_f}, P_{P_f} are not measures and P'_{Q_f}, P'_{P_f} are not real valued. Hence f is not F orthogonal on $\mathscr{E}(G)$ or G orthogonal on $\mathscr{E}(F)$.

The conditional probabilities become

$$P_{Q_f}(A|P^{-1}(B)) = \int_A \frac{|f_Q[P^{-1}(B)]|^2 dq}{P_{Q_f}[P^{-1}(B)]} = \int_A |\psi(q)|^2 \, dq,$$

$$P_{P_f}(B|Q^{-1}(A)) = \int_B |\phi(p)|^2 \, dp.$$

Hence Q and P are stochastically independent relative to f. The pseudoexpectations become

$$E_{Q_f}(Q) = \int q|\psi(q)|^2 \, dq, \quad E_{P_f}(P) = \int p|\phi(p)|^2 \, dp.$$

Moreover, since

$$f_Q(P) = \int pf(q,p)dp = \psi(q)\int p\phi(p)dp,$$

we have

$$E_{Q_f}(P) = \langle f_Q(P), Q(f)\rangle = \int p\phi(p)dp,$$

$$E_{P_f}(Q) = \int q\psi(q)dq.$$

Notice that the latter two pseudoexpectations need not be real valued.

Although the amplitude densities just considered served as illustrative examples, they are unphysical. We now construct a class of physical amplitude densities that correspond to the traditional quantum states. For

$$\psi \in L^2(\mathbb{R},dq) \cap L^1(\mathbb{R},dq), \tag{7.1}$$

we denote the Fourier transform by

$$\hat{\psi}(p) = (2\pi\hbar)^{-1/2}\int \psi(q)e^{-iqp/\hbar}\,dq$$

and the inverse Fourier transform by

$$\check{\psi}(q) = (2\pi\hbar)^{-1/2}\int \psi(p)e^{iqp/\hbar}\,dp.$$

We say that $f \in \mathscr{F}(\mathscr{A})$ is *regular* if $f_Q(A \times \mathbb{R})\hat{} = f_P(A \times \mathbb{R})$

and $f_Q(\mathbb{R}\times A)\hat{} = f_P(\mathbb{R}\times A)$ for every $A\in B(\mathbb{R})$. It is shown in Ref. 8 that f is regular if and only if

(1) for every $p\in\mathbb{R}$,

$$f(q,p) = (2\pi\hbar)^{-1/2}Q(f)(q)e^{-iqp/\hbar} \text{ a.e. } [q],$$

(2) for every $q\in\mathbb{R}$

$$f(q,p) = (2\pi\hbar)^{-1/2}Q(f)\hat{}(p)e^{iqp/\hbar} \text{ a.e. } [p].$$

It is not clear that regular amplitude densities exist and from (1) and (2) we see that if they exist, they must be nonmeasurable. Nevertheless, it is shown in Ref. 5 that for every ψ satisfying (7.1) there exists a regular f such that $Q(f) = \psi$. Intuitively, the regular f are those for which the Fourier transform of position is momentum. Moreover, $\psi = Q(f)$ is a traditional quantum state.

Let f be regular with $Q(f) = \psi$. We then have

$$P(f)(p) = \int f(q,p)dq$$

$$= (2\pi\hbar)^{-1/2}\int \psi(q)e^{-iqp/\hbar}dq = \hat{\psi}(p).$$

Hence for every $A\in B(\mathbb{R})$ we have

$$P_{Qf}(A) = \int_A |\psi(q)|^2\,dq, \quad P_{Pf}(A) = \int_A |\hat{\psi}(p)|^2\,dp,$$

which are the usual quantum mechanical formulas. Moreover, we have

$$f_Q[P^{-1}(A)](q) = \int_A f(q,p)dp = (2\pi\hbar)^{-1/2}$$

$$\times\int_A \hat{\psi}(p)e^{iqp/\hbar}\,dp = (\chi_A\hat{\psi})\check{}(q).$$

Hence

$$P_{Qf}[P^{-1}(A)] = \|(\chi_A\hat{\psi})\check{}\|^2 = \|\chi_A\hat{\psi}\|^2 = \int_A |\hat{\psi}(p)|^2\,dp.$$

We can thus get information about P by measuring Q. Notice that $P_{Qf}[P^{-1}(\cdot)]$ is a probability measure. It follows that f is Q orthogonal on $\mathscr{E}(P)$. In fact, f is Q-orthogonally scattered on $\mathscr{E}(P)$. Indeed, suppose $A,B\in B(\mathbb{R})$ and $A\cap B = \varnothing$. Then $(\mathbb{R}\times A)\cap(\mathbb{R}\times B) = \varnothing$ and we have

$$\langle f_Q[P^{-1}(A)], f_Q[P^{-1}(B)]\rangle$$
$$= \langle(\chi_A\hat{\psi})\check{}(\chi_B\hat{\psi})\check{}\rangle = \langle\chi_A\hat{\psi},\chi_B\hat{\psi}\rangle = 0.$$

Similarly,

$$P_{Pf}[Q^{-1}(A)] = \int_A |\psi(q)|^2\,dq$$

and f is P-orthogonally scattered on $\mathscr{E}(Q)$.

If $A,B\in B(\mathbb{R})$, then

$$P_{Qf}(A\,|P^{-1}(B)) = \|\chi_A f_Q[P^{-1}(B)]\|^2/\|f_Q[P^{-1}(B)]\|^2$$
$$= \|\chi_A(\chi_B\hat{\psi})\check{}\|^2/\|\chi_B\hat{\psi}\|^2.$$

Similarly,

$$P_{Pf}(A\,|P^{-1}(B)) = \|\chi_A(\chi_B\psi)\hat{}\|^2/\|\chi_B\psi\|^2.$$

It is shown in Ref. 5 that these reduce to the traditional von Neumann–Lüders formulas

$$P_{Qf}(A\,|P^{-1}(B))$$
$$= \mathrm{tr}[E^Q(A)E^P(B)P_\psi E^P(B)]/[\mathrm{tr}\,E^P(B)P_\psi],$$
$$P_{Pf}(A\,|Q^{-1}(B))$$
$$= \mathrm{tr}[E^P(A)E^Q(B)P_\psi E^Q(B)]/[\mathrm{tr}\,E^Q(B)P_\psi],$$

where P_ψ is the one-dimensional projection onto ψ and E^Q, E^P are the spectral measures for Q and P, respectively.

In the sequel, we shall assume that ψ is a Schwartz test function and that f is regular with $Q(f) = \psi$. We then obtain

$$E_{Pf}(P\,|Q^{-1}(A))$$
$$= \langle f_P(P\,|Q^{-1}(A)), f_P(Q^{-1}(A))\rangle/P_{Pf}[Q^{-1}(A)]$$
$$= \langle pf_P(Q^{-1}(A)), f_P(Q^{-1}(A))\rangle/\|\chi_A\psi\|^2$$
$$= \int p|(\chi_A\psi)\hat{}|^2\,dp\left(\int_A |\psi|^2\,dq\right)^{-1}.$$

In particular,

$$E_{Pf}(P) = \int p|\hat{\psi}(p)|^2\,dp.$$

Similarly,

$$E_{Qf}(Q) = \int q|\psi(q)|^2\,dq.$$

We also have

$$f_Q(P)(q) = \int pf(q,p)dp$$

$$= (2\pi\hbar)^{-1/2}\int p\hat{\psi}(p)e^{iqp/\hbar}\,dp$$

$$= (2\pi\hbar)^{-1/2}\left(-i\hbar\frac{d}{dq}\right)\int \hat{\psi}(p)e^{iqp/\hbar}\,dp$$

$$= -i\hbar\frac{d\psi}{dq}(q).$$

Let $R(\mathscr{A})\subseteq H(\mathscr{A})$ be the set of all scalar multiples of regular amplitude functions. Then it is clear that $R(\mathscr{A})\subseteq R(\mathscr{A})^s$, so $R(\mathscr{A})$ is a linear subspace of a superselection sector of $H(\mathscr{A})$. We can then represent Q, P as operators $\tilde{Q}$, $\tilde{P}$ on $H_Q = L^2(\mathbb{R},dq)$ as follows:

$$\tilde{Q}\psi(q) = q\psi(q),$$

$$\tilde{P}\psi(q) = \tilde{P}(Q)(f)(q) = f_Q(P)(q) = \left(-i\hbar\frac{d}{dq}\right)\psi(q).$$

If we define $\tilde{\tilde{P}}\hat{\psi}(p) = p\hat{\psi}(p)$, the following theorem holds.[8,9]

Theorem 7.1: If $G(q,p) = \Sigma a_{mn}q^mp^n$ is a polynomial, then

$$f_Q(G)(q) = \Sigma a_{mn}\tilde{Q}^m\tilde{P}^n\psi(q),$$
$$f_P(G)(p) = \Sigma a_{mn}\tilde{\tilde{P}}^n(\tilde{Q}^m\psi)\hat{}(p). \qquad \square$$

It follows that any polynomial $G(q,p) = \Sigma a_{mn}q^mp^n$ can be represented on $L^2(\mathbb{R},dq)$ by the operator $\tilde{G} = \Sigma a_{mn}\tilde{Q}^m\tilde{P}^n$. Moreover,

$$E_{Qf}(G) = \langle\Sigma a_{mn}\tilde{Q}^m\tilde{P}^n\psi,\psi\rangle,$$
$$E_{Pf}(G) = \langle\Sigma a_{mn}\tilde{P}^n\tilde{Q}^m\psi,\psi\rangle.$$

Thus, the pseudoexpectation of G depends on which measurement is executed. For example, let $G(q,p) = qp$. We

then obtain the following version of the Heisenberg commutation relation:

$$E_{Q,f}(G) - E_{P,f}(G) = \langle [\tilde{Q},\tilde{P}]\psi,\psi\rangle = i\hbar.$$

The reason that $E_{Q,f}(G)$ and $E_{P,f}(G)$ are complex valued is that f is not Q orthogonal or P orthogonal on the σ algebra $\{G^{-1}(A): A\in B(\mathbb{R})\}$. The usual uncertainty relation can be obtained by taking variances.

We now consider Schrödinger's equation. Suppose the classical Hamiltonian is

$$H(q,p) = p^2/2m + V(q).$$

If the system is closed, then we have conservation of energy

$$H(q,p) = E. \tag{7.2}$$

Now suppose $f\in\mathcal{F}(\mathcal{A})$ is regular with $\psi = Q(f)$. Taking the Q-amplitude average of (7.2) gives $f_Q[H(q,p)] = f_Q(E)$. Using the linearity of f_Q and Theorem 7.1 we obtain

$$H(\tilde{Q},\tilde{P})\psi = (\tilde{p}^2/2m)\psi + V(\tilde{Q})\psi = E\psi.$$

Of course, this is the time independent Schrödinger equation and E,ψ form an eigenpair for the operator $H(\tilde{Q},\tilde{P})$.

Now suppose the classical dynamics is generated by the Hamilton equation

$$\frac{dp}{dt} = -\frac{\partial H}{\partial q}. \tag{7.3}$$

We assume that for any time $t\in\mathbb{R}$ the system is described by a regular amplitude density $f(q,p,t)$ with corresponding vector $\psi(q,t) = Q(f)(q,t)$ and moreover, f and ψ are differentiable with respect to t. Suppose (7.3) holds in the Q-amplitude average in the sense that

$$\frac{d}{dt}\int pf(q,p,t)dp = -\frac{\partial}{\partial q}\int H(q,p)f(q,p,t)dp. \tag{7.4}$$

Then (7.4) has the form

$$\frac{d}{dt}f_Q(p) = -\frac{\partial}{\partial q}f_q(H).$$

Applying Theorem 7.1 gives

$$\frac{d}{dt}\left(-i\hbar\frac{d\psi}{dq}\right) = -\frac{\partial}{\partial q}H(\tilde{Q},\tilde{P})\psi. \tag{7.5}$$

Interchanging the order of differentiation in (7.5) gives

$$\frac{\partial}{\partial q}\left(-i\hbar\frac{\partial\psi}{\partial t}\right) = -\frac{\partial}{\partial q}H(\tilde{Q},\tilde{P})\psi. \tag{7.6}$$

Integrating both sides of (7.6) we obtain (except for a constant that we can set equal to 0)

$$i\hbar\frac{\partial\psi}{\partial t} = H(\tilde{Q},\tilde{P})\psi.$$

This, of course, is the time-dependent Schrödinger equation. We concluded that Schrödinger's equation is an amplitude averaged version of the Hamilton equation of classical mechanics.

If we use the other Hamilton equation $dq/dt = \partial H/\partial p$ and take the P-amplitude average we obtain

$$\frac{d}{dt}f_P(q) = \frac{\partial}{\partial p}f_P(H).$$

Proceeding in a similar way, we obtain the Fourier transformed time-dependent Schrödinger equation

$$i\hbar\frac{\partial\hat{\psi}}{\partial t} = \frac{p^2}{2m}\hat{\psi} + (V\psi)^{\wedge}.$$

VIII. DISCRETE FEYNMAN AMPLITUDES

We now illustrate the theory developed in Secs. II and III by presenting a model for a quantum random walk. As we shall see, this gives a discrete counterpart to the Feynman path-integral formalism.

Let S be a nonempty set. We interpret S as a set of "states" that a quantum particle can occupy. The elements of S may represent discrete positions, momenta, spins, or other physical quantities. A function $f_1: S\times S\to\mathbb{C}$ is a *one-step transition amplitude* if for every $s_1,s_2\in S$ we have

$$\Sigma_s f_1(s_1,s)\bar{f}_1(s_2,s) = \Sigma_s f_1(s,s_1)\bar{f}_1(s,s_2) = \delta_{s_1,s_2}. \tag{8.1}$$

Equation (8.1) is a type of unitarity condition. We interpret $f_1(s,s')$ as the probability amplitude that a particle moves from s to s' in one time step and we interpret $\bar{f}_1(s,s')$ as the probability amplitude that a particle moves from s' to s in minus one time step. Alternatively, if a particle is at s' at time n, then $\bar{f}_1(s,s')$ is the probability amplitude that it was at s at time $n-1$. We then can interpret $f_1(s_1,s)\bar{f}_1(s_2,s)$ as the probability amplitude that a particle moves from s_1 to s_2 via s in zero time steps. It follows that $\Sigma_s f_1(s_1,s)\bar{f}_1(s_2,s)$ is the total probability amplitude that a particle moves from s_1 to s_2 in zero time steps. It is clear that this should equal δ_{s_1,s_2}. Similar reasoning applies to the other equality in (8.1).

We denote the set of one-step transition amplitudes on S by $T(S)$. We say that an $f_1\in T(S)$ is *stochastic* if for every $s_0\in S$ we have

$$\Sigma_s|f_1(s_0,s)| < \infty \text{ and } \Sigma_s f_1(s_0,s) = 1. \tag{8.2}$$

Equation (8.2) means that a particle that is initially at s_0 moves somewhere with amplitude 1. We denote the stochastic $f_1\in T(S)$ by $T_s(S)$.

An *n path from s_0 to s* is an $(n+1)$-tuple

$$(s_0,s_1,\ldots,s_{n-1},s)\in S^{n+1}.$$

Let $\mathcal{P}_n(s_0,s)$ be the set of n paths from s_0 to s and let $\mathcal{P}_n(s_0) = \cup_s \mathcal{P}_n(s_0,s)$. For $f_1\in T(S)$ and

$$x = (s_0,s_1,\ldots,s_{n-1},s)\in\mathcal{P}_n(s_0,s),$$

we define the *amplitude* $f(x)$ of x as

$$f(x) = f_1(s_0,s_1)f_1(s_1,s_2)\cdots f_1(s_{n-1},s). \tag{8.3}$$

The *n-step transition amplitude* from s_0 to s is defined as $f_0(s_0,s) = \delta_{s_0,s}$ and

$$f_n(s_0,s) = \Sigma\{f(x): x\in\mathcal{P}_n(s_0,s)\}, \quad n\geqslant 1. \tag{8.4}$$

It is not hard to show that the summation in (8.4) converges so $f_n(s_0,s)$ exists. The next result shows that probability is conserved and that an amplitude Chapman–Kolmogorov equation holds.

Theorem 8.1: If $f_1\in T(S)$, then

$$\Sigma_s|f_n(s_0,s)|^2 = \Sigma_s|f_n(s,s_0)|^2 = 1, \quad n\in\mathbb{N}, \tag{8.5}$$

$$f_n(s_0,s) = \Sigma_{s'} f_m(s_0,s')f_{n-m}(s',s), \quad m,n\in\mathbb{N}, \quad m<n. \tag{8.6}$$

Proof: See Ref. 10. □

Let N be a large positive integer and form the sample space $X=\{x:\ x\in\mathscr{P}_N(s_0)\}$. For $j=1,...,N$, define F_j: $X\to Y_j=F_j(X)=S$ by

$$F_j(s_0,s_1,...,s_{N-1},s_N)=s_j.$$

Let Σ_j be the power set on Y_j and let ν_j be the counting measure on Σ_j. For $s\in Y_j$, let Σ_s be the power set on $F_j^{-1}(s)$ and let μ_s be the counting measure on Σ_s. Then F_j is a measurement on $X, j=1,...,N$, and $\mathscr{A}=\{F_j: 1\leqslant j\leqslant N\}$ is a complete catalog on X. The sample points represent discrete trajectories over a time period N for a quantum particle that starts at s_0. The measurement F_j gives the state of the particle at time $j, j=1,...,N$. We now show that the amplitude f defined earlier provides an amplitude density for $\mathscr{A}$.

Theorem 8.2: If $f_1\in T_s(S)$, then $f\in\mathscr{F}(\mathscr{A})$ and $F_j(f)(s)=f_j(s_0,s)$.

Proof: Applying (8.2) and (8.6) we have for $n\leqslant N$,

$$\begin{aligned}\Sigma_s f_n(s_1,s)&=\Sigma_s\Sigma_{s'} f_{n-1}(s_1,s')f_1(s',s)\\&=\Sigma_{s'} f_{n-1}(s_1,s')\Sigma_s f_1(s',s)=\Sigma_{s'} f_{n-1}(s_1,s').\end{aligned}$$

It follows by induction that

$$\Sigma_s f_n(s_1,s)=1. \tag{8.7}$$

Let $F_j^{-1}(s)$ be a fiber for F_j. It follows from (8.2) that $f\in L^1(F_j^{-1}(s),\Sigma_s,\mu_s)$. Moreover,

$$\begin{aligned}F_j(f)(s)&=\Sigma\{f(x): F_j(x)=s\}\\&=f_j(s_0,s)\Sigma_{s'} f_{N-j}(s',s).\end{aligned}$$

By applying (8.7) one obtains

$$F_j(f)(s)=f_j(s_0,s).$$

From (8.5) we conclude that $F_j\in L^2(Y_j,\Sigma_j,\nu_j)$ and $\|F_j(f)\|=1$. Hence $f\in\mathscr{F}(\mathscr{A})$. □

Let Σ_0 be the collection of subsets $B\subseteq X$ such that $\Sigma_{x\in B}|f(x)|<\infty$. For $B\in\Sigma_0$, define $f(B)$ by

$$f(B)=\Sigma_{x\in B} f(x).$$

If $A,B\in\Sigma_0$ with $f(B)\neq 0$ we define

$$f(A|B)=f(A\cap B)/f(B).$$

When we write $f(A|B)$ we always assume that $f(B)\neq 0$. Using these definitions, F_j, $j=1,...,N$, becomes an amplitude Markov chain in the following sense. For any $A_j\subseteq S$ we have

$$\begin{aligned}&f(F_j^{-1}(A_j)|F_{j-1}^{-1}(A_{j-1})\cap\cdots\cap F_1^{-1}(A_1))\\&\quad=f(F_j^{-1}(A_j)|F_{j-1}^{-1}(A_{j-1})).\end{aligned}\tag{8.8}$$

To prove (8.8), let $s_1,...,s_j\in S$. We then have

$$\begin{aligned}&f(F_j^{-1}(s_j)|F_{j-1}^{-1}(s_{j-1})\cap\cdots\cap F_1^{-1}(s_1))\\&\quad=\frac{f(F_j^{-1}(s_j)\cap\cdots\cap F_1^{-1}(s_1))}{f(F_{j-1}^{-1}(s_{j-1})\cap\cdots\cap F_1^{-1}(s_1))}\\&\quad=\frac{f(s_0,...,s_{j-1},s_j)\Sigma_{s'} f_{N-j}(s_j,s')}{f(s_0,...,s_{j-1})\Sigma_{s'} f_{N-j+1}(s_{j-1},s')}\\&\quad=f_1(s_{j-1},s_j)=\frac{f(F_j^{-1}(s_j)\cap F_{j-1}^{-1}(s_{j-1}))}{f(F_{j-1}^{-1}(s_{j-1}))}\\&\quad=f(F_j^{-1}(s_j)|F_{j-1}^{(1)}(s_{j-1})).\end{aligned}\tag{8.9}$$

If we sum (8.9) over the $s_i\in A_i$, $i=1,...,j$, one by one we obtain (8.8). In general, F_j is not a Markov chain in the classical sense.

As in previous sections, we can compute various probabilities and expectations for the measurements $F_j, j=1,...,N$. The simplest of these is

$$P_{F_j}(A)=\Sigma_{s\in A}|f_j(s_0,s)|^2. \tag{8.10}$$

Of course, (8.10) gives a probability measure. If $n\leqslant j$, we have

$$P_{F_j f}[F_n^{-1}(B)]=\Sigma_s|\Sigma_{s'\in B} f_n(s_0,s')f_{j-n}(s',s)|^2. \tag{8.11}$$

In general, (8.11) does not give a probability measure and f is not F_j orthogonal on $\mathscr{E}(F_n)$. Physically, this is because the generalized event $F_n^{-1}(B)$ interferes with the later measurement F_j. However, if $j\leqslant n$, we have

$$\begin{aligned}P_{F_j f}[F_n^{-1}(B)]&=\Sigma_s|\Sigma_{s'\in B} f_j(s_0,s)f_{n-j}(s,s')|^2\\&=\Sigma_s|f_j(s_0,s)|^2\Sigma_{s'\in B}|f_{n-j}(s,s')|^2.\end{aligned}\tag{8.12}$$

Notice that (8.12) does give a probability measure and hence f is F_j orthogonal on $\mathscr{E}(F_n)$. This is because the generalized event $F_n^{-1}(B)$ does not interfere with the earlier measurement F_j. The corresponding conditional probabilities become

$$\begin{aligned}&P_{F_j f}(A|F_n^{-1}(B))\\&\quad=\Sigma_{s\in A}|\Sigma_{s'\in B} f_n(s_0,s')f_{j-n}(s',s)|^2/P_{F_j f}\\&\qquad\times[F_n^{-1}(B)],\quad n\leqslant j,\\&P_{F_j f}(A|F_n^{-1}(B))\\&\quad=\Sigma_{s\in A}|f_j(s_0,s)|^2\Sigma_{s'\in B}|f_{n-j}(s,s')|^2/P_{F_j f}\\&\qquad\times[F_n^{-1}(B)],\quad j\leqslant n.\end{aligned}$$

The main problem in this theory (as in traditional quantum mechanics) is to find an explicit expression for $f_n(s_0,s)$. Fixing $s_0\in S$, we define the *discrete wave function* $\psi_n(s)=f_n(s_0,s)$. It follows from (8.6) that

$$\psi_{n+1}(s)=\Sigma_{s'} f_1(s',s)\psi_n(s'). \tag{8.13}$$

We call the difference equation (8.13) the *discrete wave equation*. It is shown in Ref. 10 that for certain f_1, (8.13) is a discrete analog of Dirac's equation.

Lest the reader question the existence of stochastic one-step transition amplitudes, we now give some simple examples (others are given in Ref. 10 and later in this section.) If $S=\{s_1,...,s_n\}$ is finite, then $f_1\in T_s(S)$ is equivalent to a stochastic, unitary matrix with entries $f_1(s_i,s_j)$. Examples of these are

$$\begin{bmatrix}\frac{1}{2}+\frac{1}{2}i & \frac{1}{2}-\frac{1}{2}i\\ \frac{1}{2}-\frac{1}{2}i & \frac{1}{2}+\frac{1}{2}i\end{bmatrix}\begin{bmatrix}\frac{1}{3}+(1\sqrt{3})i & \frac{1}{3} & \frac{1}{3}-(1\sqrt{3})i\\ \frac{1}{3}-(1\sqrt{3})i & \frac{1}{3}+(1\sqrt{3})i & \frac{1}{3}\\ \frac{1}{3} & \frac{1}{3}-(1\sqrt{3})i & \frac{1}{3}+(1\sqrt{3})i\end{bmatrix}.$$

We now present a dynamical model whose one-step transition amplitude is a discrete analog of a free Feynman amplitude. We shall work in a two-dimensional configuration space and the extension to three dimensions will be straightforward.

Let a and n be positive integers such that for any integer

$0 < t < n$, we have

$$at \neq 0 \pmod{n}. \tag{8.14}$$

(For example, let $a = 3$, $n = 16$.) Let α be an angle with radian measure $2\pi/n$ and let k_0 be a unit vector in $\mathbb{R}^2$. Let $k_1,...,k_{n-1}$ be unit vectors in $\mathbb{R}^2$ such that $k_j \cdot k_0 = \cos j\alpha$, $j = 1,...,n-1$. Thus each k_j forms an angle α with its predecessor $j = 1,...,n-1$. Let $V \subseteq \mathbb{R}^2$ denote the set of points of the form $q = \Sigma e_j$, $e_j \in \{k_0,...,k_{n-1}\}$. We think of V as a discrete configuration space and form a discrete phase space

$$S = \{(q,k_j): q \in V, j = 0,...,n-1\}. \tag{8.15}$$

The q's correspond to discrete positions and the k_j's correspond to discrete momenta. (Actually, the k_j's give the momenta directions; the magnitude will depend on a.) We define the *discrete Feynman amplitude* $g_1: S \times S \to \mathbb{C}$ by

$$g_1((q,k_r),(q+k_r,k_t)) = n^{-1/2}\exp[ia\pi(t-r)^2/n] \tag{8.16}$$

and g_1 is zero, otherwise. Equation (8.16) says that if a particle is at $q \in V$ and is moving in the k_r direction, then it moves to $q + k_r$ and can then change direction with the given amplitude.

Lemma 8.3: If g_1 is a discrete Feynman amplitude, then $g_1 \in T(S)$.

Proof: It is clear that (8.1) holes for $s_1 = s_2$. Now suppose that $s_1, s_2 \in S$ with $s_1 \neq s_2$. We can assume that $s_1 = (q_1,k_r)$, $s_2 = (q_2,k_t)$, where $r \neq t$ and $q_1 + k_r = q_2 + k_t$. Since $am \neq 0 \pmod{n}$ for a $0 < m < n$ we obtain

$$\begin{aligned}
&\Sigma_s g_1(s_1,s)\bar{g}_1(s_2,s) \\
&\quad = \sum_{j=0}^{n-1} g_1((q_1,k_r),(q_1+k_r,k_j))\bar{g}_1((q_2,k_t),(q_1+k_r,k_j)) \\
&\quad = n^{-1}\sum_{j=0}^{n-1} \exp\left[\frac{ia\pi(j-r)^2}{n}\right]\exp\left[\frac{-ia\pi(j-t)^2}{n}\right] \\
&\quad = n^{-1}\exp\left[\frac{ia\pi(r^2-t^2)}{n}\right]\sum_{j=0}^{n-1}\left[\exp\left[\frac{i2a\pi(t-r)}{n}\right]\right]^j \\
&\quad = [n^{-1}\exp[ia\pi(r^2-t^2)/n]] \\
&\qquad \times [1-\exp[i2a\pi(t-r)/n]]^{-1} \\
&\qquad \times [1-\exp[-i2a\pi(t-r)]] = 0.
\end{aligned}$$

In a similar way, it follows that for $s_1 \neq s_2$,

$$\Sigma_s g_1(s,s_1)\bar{g}_1(s,s_2) = 0. \qquad \square$$

We now show that as $n \to \infty$, the discrete Feynman amplitude approaches the traditional free particle continuum Feynman amplitude in a certain sense. Let

$$x = \{(q_0,k_{j_0}),...,(q_r,k_{j_r})\}$$

be an r path, where $q_t + k_{j_t} = q_{t+1}$, $t = 0,...,r-1$. Let $\beta_t = 2\pi(j_t - j_{t-1})/n$, $t = 1,...,r$, and suppose that the β_t are small (or close to 2π). This will be the case if n is large and the path does not turn very far. Then the distance between q_{t+1} and q_{t-1} becomes

$$\begin{aligned}
&\|q_{t+1} - q_{t-1}\|^2 \\
&\quad = \|q_{t-1} + k_{j_{t-1}} + k_{j_t} - q_{t-1}\|^2 \\
&\quad = 2 + 2k_{j_t}\cdot k_{j_{t-1}} = 2 + 2\cos\beta_t \\
&\quad \approx 2 + 2(1 - \beta_t^2/2) = 4 - \beta_t^2.
\end{aligned}$$

Hence

$$\beta_t^2 \approx 4 - \|q_{t+1} - q_{t-1}\|^2.$$

If v_t denotes the particle's "velocity" at time t, we have

$$v_t^2 \approx \|q_{t+1} - q_{t-1}\|^2/4.$$

Thus $\beta_t^2 \approx 4(1 - v_t^2)$. Then the amplitude of the path x becomes

$$\begin{aligned}
g(x) &= n^{-r/2}\exp\left[ia\pi n^{-1}\sum_{t=1}^{r}(j_t - j_{t-1})^2\right] \\
&= n^{-r/2}\exp\left[ian(4\pi)^{-1}\sum_{t=1}^{r}\beta_t^2\right] \\
&\approx n^{-r/2}\exp\left[ian\pi^{-1}\sum_{t=1}^{r}(1-v_t^2)\right] \\
&= n^{-r/2}e^{ianr/\pi}\exp\left[-i2n\pi^{-1}\sum_{t=1}^{r}\frac{av_t^2}{2}\right].
\end{aligned}$$

If we let a correspond to the mass of the particle (which would be an integer in certain mass units), then the summation corresponds to the integral of the kinetic energy over the path. In this sense $g(x)$ approaches the free particle continuum Feynman amplitude.

In general, g_1 need not be stochastic. However, for certain n, we shall show that g_1 multiplied by a constant of modulus 1 is stochastic.

Lemma 8.4: If n is even, then for any $t = 0,...,n-1$, we have

$$\sum_{s=0}^{n-1}\exp\left[\frac{ia\pi(s-t)^2}{n}\right] = \sum_{s=0}^{n-1}\exp\left(\frac{ia\pi s^2}{n}\right).$$

Proof: If $t \in \{0,...,n-1\}$, then

$$\begin{aligned}
&\sum_{s=0}^{n-1}\exp\left[\frac{ia\pi(s-t)^2}{n}\right] \\
&\quad = \sum_{s=0}^{t-1}\exp\left[\frac{ia\pi(t-s)^2}{n}\right] \\
&\qquad + \sum_{s=t}^{n-1}\exp\left[\frac{ia\pi(s-t)^2}{n}\right].
\end{aligned}$$

On the right-hand side of the above equation, replace $t - s$ in the first sum by $n - j$ and replace $s - t$ in the second sum by j to obtain

$$\begin{aligned}
&\sum_{s=0}^{n-1}\exp\left[\frac{ia\pi(s-t)^2}{n}\right] \\
&\quad = \sum_{j=n-t}^{n-1}\exp\left[\frac{ia\pi(n-j)^2}{n}\right] + \sum_{j=0}^{n-t-1}\exp\left(\frac{ia\pi j^2}{n}\right) \\
&\quad = \sum_{j=n-t}^{n-1}\exp\left[\frac{ia\pi(n^2-2nj+j)^2}{n}\right] \\
&\qquad + \sum_{j=0}^{n-t-1}\exp\left(\frac{ia\pi j^2}{n}\right) \\
&\quad = \sum_{j=0}^{n-1}\exp\left(\frac{ia\pi j^2}{n}\right). \qquad \square
\end{aligned}$$

In the sequel, we shall assume that n has the form $n = 2^k$ for some positive integer k and that a is odd and

satisfies (8.14). Let

$$I_k = \sum_{j=0}^{2^k-1} \exp\left(\frac{ia\pi j^2}{2^k}\right).$$

Lemma 8.5: Under the above assumptions we have

$$I_k = \begin{cases} 2^{k/2}e^{ia\pi/4}, & k \text{ even}, \\ 2^{k/2}e^{ia\pi/4}\,\mathrm{sgn}(\cos a\pi/4), & k \text{ odd}. \end{cases}$$

Proof: It is not hard to show that I_k satisfies the recurrence relation $I_{k+2} = 2I_k$. Since the initial conditions are $I_1 = 1 + e^{ia\pi/2}$ and $I_2 = 2e^{ia\pi/4}$, we obtain

$$I_k = \begin{cases} 2^{k/2}e^{ia\pi/4}, & k \text{ even}, \\ 2^{(k-1)/2}(1 + e^{ia\pi/2}), & k \text{ odd}. \end{cases}$$

However,

$$\begin{aligned} 2^{(k-1)/2}(1 + e^{ia\pi/2}) &= 2^{(k-1)/2}e^{ia\pi/4}(e^{ia\pi/4} + e^{-ia\pi/4}) \\ &= 2^{(k-1)/2}e^{ia\pi/4}\cos a\pi/4 \\ &= 2^{k/2}e^{ia\pi/4}\,\mathrm{sgn}(\cos a\pi/4) \end{aligned}$$

and the result follows. □

We now define $f_1: S \times S \to \mathbb{C}$ by

$$f_1(s_1,s_2) = \begin{cases} e^{-ia\pi/4}g_1(s_1,s_2), & k \text{ even}, \\ e^{-ia\pi/4}\,\mathrm{sgn}(\cos a\pi/4)g_1(s_1,s_2), & k \text{ odd}. \end{cases}$$

It follows from Lemmas 8.3–8.5 that $f_1 \in T_s(S)$. We denote the constant multiplying $g_1(s_1,s_2)$ by b.

Let N be a large positive integer and form the sample space $X = \{x: x \in \mathscr{P}_N(s_0)\}$ on the discrete phase space S given by (8.15). For $j = 1,...,N$, define $Q_j: X \to Y_j = V$ and $P_j: X \to Y_j' = \{k_0,...,k_{n-1}\}$ by $Q_j(x) = q_j$, $p_j(x) = k_{i_j}$, where

$$x = ((q_0,k_{i_0}),...,(q_N,k_{i_N})).$$

Then Q_j, P_j, $j = 1,...,N$, can be made into measurements in terms of counting measures in the usual way and $\mathscr{A} = \{Q_j, P_j: 1 \leqslant j \leqslant N\}$ is a complete catalog on X. We see that Q_j gives a position measurement and P_j gives a momentum measurement at time j. Let f_1 be the discrete Feynman amplitude defined above. Define the function $f: X \to \mathbb{C}$ as in Eq. (8.3). We now show that $f \in \mathscr{F}(\mathscr{A})$.

Letting $f_j(s_0,s)$ be the j-step transition amplitude defined as in Eq. (8.4), we have

$$P_j(f)(k_r) = \Sigma\{f(x): P_j(x) = k_r\} = \Sigma_q f_j(s_0,(q,k_r)), \tag{8.17}$$

$$Q_j(f)(q) = \Sigma\{f(x): Q_j(x) = q\} = \Sigma_r f_j(s_0,(q,k_r)). \tag{8.18}$$

Let $\psi_j(s) = f_j(s_0,s)$ be the discrete wave function and let $\phi_j(r) = \Sigma_q \psi_j(q,k_r)$ be the *momentum wave function* and $\psi_j(q) = \Sigma_r \psi_j(q,k_r)$ be the *position wave function*. Since the summations have only a finite number of nonzero terms there is no problem with convergence. It follows from (8.17) and (8.18) that

$$P_j(f)(k_r) = \phi_j(r), \quad Q_j(f)(q) = \psi_j(q). \tag{8.19}$$

It follows from (8.19) that $f \in \mathscr{F}(\mathscr{A})$ if and only if for $j = 1,...,N$ we have

$$\Sigma_r|\phi_j(r)|^2 = \Sigma_q|\psi_j(q)|^2 = 1. \tag{8.20}$$

To prove (8.20) we consider the discrete wave equation (8.13). In the present case, (8.13) becomes

$$\begin{aligned} \psi_j(q,k_r) &= \Sigma_t f_1((q - k_t,k_t),(q,k_r))\psi_{j-1}(q - k_t,k_t) \\ &= bn^{-1/2}\Sigma_t \exp[ia\pi(r - t)^2/n]\psi_{j-1}(q - k_t,k_t). \end{aligned} \tag{8.21}$$

By summing (8.21) over $q \in V$ we obtain

$$\phi_j(r) = bn^{-1/2}\Sigma_t \exp[ia\pi(r - t)^2/n]\phi_{j-1}(t). \tag{8.22}$$

By applying Lemmas 8.4 and 8.5 and summing (8.21) over k_r, $r = 0,...,n-1$, one obtains

$$\psi_j(q) = \Sigma_t \psi_{j-1}(q - k_t,k_t). \tag{8.23}$$

Define the $n \times n$ matrix

$$M_{rt} = bn^{-1/2}[\exp[ia\pi(r - t)^2/n]]. \tag{8.24}$$

It follows from Lemma 8.3 that M is unitary. Moreover, (8.22) becomes $\phi_j = M\phi_{j-1}$ and iteration now gives $\phi_j = M^j\phi_0$. Since ϕ_0 is the unit vector $\phi_0(r) = \delta_{ri_0}$ and M^j is unitary, we conclude that ϕ_j is a unit vector. Hence the first part of (8.20) holds.

If $s_0 = (q_0,k_u)$, then initially we have

$$\psi_0(q,k_r) = \delta_{(q_0,k_u)(q,k_r)}.$$

Applying (8.21) and iterating (8.23) with $c = ia\pi/n$ gives the following expression for $\psi_j(q)$:

$$(bn^{-1/2})^{j-1} \sum_{t_1,...,t_{j-1}} \left\{\exp\left[c\sum_{r=1}^{j-2}(t_r - t_{r+1})^2\right]\exp[c(t_{j-1} - u)^2]: q = q_0 + k_{t_1} + \cdots + k_{t_{j-1}} + k_u\right\}. \tag{8.25}$$

Theorem 8.6: For $j = 0,...,N$, we have $\Sigma_q|\psi_j(q)|^2 = 1$.

Proof: We obtain this result by varying the initial condition. Let $\psi_j^{(u)}(q)$ denote the $\psi_j(q)$ with initial condition (q_0,k_u). Then

$$\Sigma_q|\psi_j^{(u)}(q)|^2 = \Sigma_q\psi_j^{(u)}(q)\bar{\psi}_j^{(u)}(q),$$

where $\psi_j^{(u)}$ is given by (8.25). By rotational symmetry, we conclude that

$$\Sigma_q|\psi_j^{(u)}(q)|^2 = \Sigma_q|\psi_j^{(v)}(q)|^2.$$

More precisely, there is a bijection between the paths from $q_0 + k_u$ and the paths from $q_0 + k_v$ that preserves the amplitudes. We then have

$$n\Sigma_q|\psi_j(q)|^2=\Sigma_{q,u}|\psi_j^{(u)}(q)|^2$$

$$=n^{1-j}\Sigma_u\sum_{t_1,\ldots,t_{j-1}}\exp\left[c\sum_{r=1}^{j-2}(t_r-t_{r+1})^2\right]\exp[c(t_{j-1}-u)^2]$$

$$\times\sum_{t'_1,\ldots,t'_{j-1}}\exp\left[\bar{c}\sum_{r=1}^{j-2}(t'_r-t'_{r+1})^2\right]\exp[\bar{c}(t'_{j-1}-u)^2]$$

$$=n^{2-j}\sum_{t_1,\ldots,t_{j-1}}\exp\left[c\sum_{r=1}^{j-2}(t_r-t_{r+1})^2\right]\sum_{t'_1,\ldots,t'_{j-2}}\exp\left[\bar{c}\sum_{r=1}^{j-3}(t'_r-t'_{r+1})^2\right]\exp[\bar{c}(t'_{j-2}-t_{j-1})^2]$$

$$=n^{3-j}\sum_{t_1,\ldots,t_{j-2}}\exp\left[c\sum_{r=1}^{j-3}(t_r-t_{r+1})^2\right]\sum_{t'_1,\ldots,t'_{j-3}}\exp\left[\bar{c}\sum_{r=1}^{j-4}(t'_r-t'_{r+1})^2\right]\exp[\bar{c}(t'_{j-3}-t_{j-2})^2]$$

$$=n^{-1}\Sigma_{t_1,t_2}\exp[c(t_1-t_2)^2]\Sigma_{t'_1}\exp[\bar{c}(t'_1-t_2)^2]=\Sigma_{t_1}1=n. \qquad \square$$

The following theorem enables us to find an explicit expression for ϕ_j.

Theorem 8.7: For $r=0,\ldots,n-1$, the eigenvalues of M are $\lambda_r=\exp(-ia\pi r^2/n)$. The corresponding unit eigenvectors are

$$e_r=n^{-1/2}(1,e^{-i2a\pi r/n},e^{-i(2)2\pi ar/n},\ldots,e^{-i(n-1)2\pi ar/n}).$$

Proof: The jth entry of Me_r is

$$\Sigma_t M_{jt}(e_r)_t=bn^{-1/2}\Sigma_t\exp[ia\pi(j-t)^2/n]$$

$$\times\exp(-i2a\pi tr/n)$$

$$=bn^{-1/2}\sum_{t=-j}^{n-1-j}\exp(ia\pi t^2/n)$$

$$\times\exp[-i2a\pi(t+j)r/n]$$

$$=bn^{-1/2}\exp[-ia\pi(r^2+2jr)/n]$$

$$\times\sum_{t=-j}^{n-1-j}\exp[ia\pi(t-r)^2/n].$$

By Lemmas 8.4 and 8.5, the last summation becomes

$$\sum_{t=0}^{n}\exp[ia\pi(t-j-r)^2/n]=n^{1/2}\bar{b}.$$

It follows that $(Me_r)_j=\lambda_r(e_r)_j$. $\square$

Corollary 8.8: For $j=0,\ldots,N$, $t=0,\ldots,n-1$, we have

$$\phi_j(t)=n^{-1}\sum_{t=0}^{n-1}\exp[i2\pi ar(u-t)/n]$$

$$\times\exp(-ia\pi jr^2/n).$$

Proof: Expanding the initial momentum wave function $\phi_j(r)=\delta_{ru}$ in terms of the orthonormal basis e_r, $r=0,\ldots,n-1$, we have

$$\phi_0=\Sigma_r\langle\phi_0,e_r\rangle e_r.$$

Hence applying Theorem 8.7 gives

$$\phi_j=M^j\phi_0=\Sigma_r\langle\phi_0,e_r\rangle M^j e_r$$

$$=\Sigma_r\langle\phi_0,e_r\rangle\lambda_r^j e_r=n^{-1/2}\Sigma_r e^{i2\pi aur/n}\lambda_r^j e_r.$$

Again, by Theorem 8.7 we have

$$\phi_j(t)=n^{-1}\Sigma_r e^{i2\pi aur/n}\exp(-ia\pi jr^2/n)e^{-i2\pi atr/n}$$

$$=n^{-1}\Sigma_r e^{i2\pi ar(u-t)/n}\exp(-ia\pi jr^2/n). \qquad \square$$

Using Corollary 8.8 we can compute various momentum probabilities. For example,

$$P_{p_j,f}(k_u)=|P_j(f)(k_u)|^2=|\phi_j(u)|^2$$

$$=n^{-2}|\Sigma_r\exp(-ia\pi jr^2/n|^2).$$

It appears to be quite difficult to simplify (8.25) to obtain an explicit expression for $\psi_j(q)$. One can further investigate discrete Feynman amplitudes for a particle moving under the influence of a potential.[10]

[1] P. Lahti and P. Mittlestaedt, *Symposium on the Foundations of Modern Physics* (World Scientific, Singapore, 1985).

[2] I. Pitowski, "The range of quantum probability," J. Math. Phys. **27**, 1156 (1986).

[3] R. Feynman and A. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).

[4] M. Munroe, *Introduction to Measure and Integration* (Addison–Wesley, Reading, MA, 1953).

[5] S. Gudder, "Probability manifolds," J. Math. Phys. **25**, 2397 (1984).

[6] D. Foulis and C. Randall, "Empirical logic and tensor products," in *Interpretations and Foundations of Quantum Theory*, edited by H. Neumann (Wissenschaftsverlag, Mannheim, 1981), pp. 9–20.

[7] S. Gudder, "Partial Hilbert spaces and amplitude functions," Ann. Inst. H. Poincaré **45**, 311 (1986).

[8] S. Gudder, "Amplitude phase-space model for quantum mechanics," Int. J. Theor. Phys. **24**, 343 (1985).

[9] I. Pitowski, "A phase-space model for quantum mechanics in which all operators commute," in *Fundamental Problems in Quantum Mechanics Conference Proceedings* (CUNY, Albany, NY, 1984).

[10] S. Gudder, "Discrete quantum mechanics," J. Math. Phys. **27**, 1782 (1986).

Conditions for the convolution of two Wigner distributions to be itself a Wigner distribution

Francis J. Narcowich
Department of Mathematics, Texas A&M University, College Station, Texas 77843-3368

(Received 26 October 1987; accepted for publication 18 May 1988)

The convolution of two Wigner distribution functions (WDF's) is always non-negative, but not always a WDF, as Jagannathan *et al.* [Phys. Lett. A **120**, 161 (1987)] have shown. In this paper conditions are given that are sufficient, and probably necessary, for such a convolution to be a WDF, and a new characterization of Gaussian WDF's is obtained as a by-product.

I. INTRODUCTION

Recently, Jagannathan *et al.*[1] looked at an interesting problem. Consider a spinless quantum mechanical system with n degrees of freedom. In the Wigner–Weyl phase-space formulation of quantum mechanics,[2] the states of such a system, both pure and mixed, are represented by certain continuous functions defined on the corresponding classical phase space, the Wigner distribution functions (WDF's).[2,3] The problem then is the following: If $P_0(q_1,\ldots,q_n,p_1,\ldots,p_n)$ is a fixed WDF, then what are necessary and sufficient conditions for the convolution

$$P_0{*}P(\mathbf{q},\mathbf{p}) = \int P_0(\mathbf{q}-\mathbf{q}',\mathbf{p}-\mathbf{p}')P(\mathbf{q}',\mathbf{p}')d^n p'\, d^n q' \quad (1.1)$$

to be a WDF whenever P is one?

The convolution $P_0{*}P$ has an interesting physical interpretation. If P_0 is taken to be a state associated with an apparatus, then $P_0{*}P$ can be interpreted as being the state that results when one uses that apparatus to make a measurement on a system in the state P. (See Ref. 1 and the references listed there.)

When P_0 is a Gaussian, $P_0{*}P$ can be used to produce the "smoothed" or Husimi[1,2,4] distribution function, which turns out to be entire in $(\mathbf{q},\mathbf{p})$ and pointwise non-negative.[5] Actually, the fact that $P_0{*}P$ is pointwise non-negative is not peculiar to Gaussian P_0's; it occurs whenever both P_0 and P are WDF's.[1,5]

There are two main results that Jagannathan *et al.* obtained in Ref. 1. The first is that whenever P_0 is a *pointwise* non-negative WDF, the convolution $P_0{*}P$ is also a pointwise non-negative WDF, provided P is itself a WDF. Put another way, they found that a sufficient condition for $P_0{*}P$ to be a WDF for every WDF P is that P_0 be a *non-negative* WDF.

The second result is that *not* every WDF will work when used as a P_0. Indeed, they construct WDF's P_0 and P such that $P_0{*}P$ is *not* a WDF, although it is, of course, a pointwise non-negative function.

In this paper, I will obtain another set of conditions sufficient for P_0 to result in $P_0{*}P$ being a WDF whenever P is. I will use these conditions to construct an example of a WDF P_0 for which $P_0{*}P$ is a WDF if P is, but which is *not* pointwise non-negative. These conditions allow for P_0 being a non-negative WDF, and so they include the sufficient condition found in Ref. 1. Indeed, it is quite possible that the set I give is necessary as well as sufficient for P_0 to produce the desired result.

The conditions I obtain will follow in part from the KLM conditions, which are necessary and sufficient conditions for a function defined on phase space to be a WDF,[3] and which are themselves based on Kastler's concept of a function of *$\hbar$-positive type*[6] and on the notion of a symplectic Fourier transform.[3] In Sec. II, I will establish notation, discuss symplectic Fourier transforms, and then introduce a class of functions that I call of *η-positive type*. When the parameter $\eta = \hbar$, one again obtains Kastler's class of functions of $\hbar$-positive type. When $\eta = 0$, one recovers the familiar class of functions of ordinary (Bochner) positive type. The latter class arises in connection with classifying states in classical mechanics. Other values of η would have the physical significance of a scaled, or even negative, $\hbar$. Here, these other η's play only a mathematical role, albeit an important one. I close Sec. II by stating the KLM conditions, and then briefly comparing them with the conditions necessary and sufficient for a function to be a classical state.

Constructing the example mentioned earlier requires knowing the values of η for which a Gaussian will be of η-positive type. In Sec. III, I will find these values, and as a by-product I will obtain a very simple, easily applied characterization of those phase-space Gaussians that are WDF's. Simon *et al.*[7] have recently characterized these using Williamson's normal form of a non-negative matrix under symplectic transformations.[8] What they obtain is a characterization that relies heavily on symplectic invariants of a matrix, and that is somewhat more complicated and harder to apply than what I obtain here. With some work, one can show that the characterizations are equivalent, although I will not do that in this paper.

In Sec. IV, I state and establish the main results of this paper. To do this, I need to take a careful look at the set of all η for which a given function F is of η-positive type. This set, which I term the Wigner spectrum of F, and its properties are essential for obtaining and even stating the main results; Wigner spectra may eventually prove to be useful in semiclassical theories.

Section V is a discussion section. It contains a brief summary of the results from Sec. IV, and also an argument—but not a proof—making it plausible that the conditions derived in Sec. IV are necessary as well as sufficient for the map $P \to P_0{*}P$ to take WDF's to WDF's. Indeed, what I do is reduce the question of necessity to two open technical questions involving Wigner spectra.

 0022-2488/88/092036-06$02.50

II. THE KLM CONDITIONS

The KLM conditions, which comprise a set of conditions that are necessary and sufficient for a phase-space function to be a WDF, are named after their discoverers, Kastler, Loupias, and Miracle-Sole.[3,6,9] In order to state them, I have to introduce the idea of a *symplectic* Fourier transform. To do that, I need the notation that was introduced in Ref. 3.

A point in phase space will be denoted by z; that is, let

$$z \equiv (q_1,\ldots,q_n,p_1,\ldots,p_n) = (\mathbf{q},\mathbf{p}) . \tag{2.1}$$

[This differs from Ref. 1, where $z = (q + ip)/\sqrt{2}$.] Next, set

$$\sigma(z,z') = \mathbf{q}' \cdot \mathbf{p} - \mathbf{p}' \cdot \mathbf{q} = \sum_{j=1}^{n} (q_j' p_j - p_j' q_j) , \tag{2.2}$$

which is an antisymmetric, bilinear form defined on phase space. Here σ is a familiar object from classical mechanics, and is called a symplectic form.[10] In terms of matrices, σ looks like

$$\sigma(z,z') = z^T \beta z' , \tag{2.3}$$

where z,z' are to be regarded as column vectors, z^T is the transpose of z, and

$$\beta = \begin{pmatrix} 0_{n\times n} & -I_{n\times n} \\ I_{n\times n} & 0_{n\times n} \end{pmatrix} . \tag{2.4}$$

In (2.4), $I_{n\times n}$ and $0_{n\times n}$ are the usual $n\times n$ identity and zero matrices, respectively.

For later reference, I want to point out that a *symplectic matrix* is a $2n\times 2n$ matrix that preserves σ. That is, S is a symplectic matrix if and only if

$$\sigma(Sz,Sz') = \sigma(z,z') , \tag{2.5}$$

for every pair of points z,z' in phase space. Equivalently, S is symplectic if and only if[7,11]

$$S^T \beta S = \beta . \tag{2.6}$$

The phase-space volume element will come up in a number of different places in this work. To avoid endlessly repeating the usual expression for this volume element, let

$$dv(z) \equiv dq_1 \cdots dq_n \, dp_1 \cdots dp_n . \tag{2.7}$$

The symplectic Fourier transform of a phase-space function $g(z)$ is defined by

$$\tilde{g}(a) = \int g(z) e^{i\sigma(a,z)} \, dv(z) , \tag{2.8}$$

where $a = (\mathbf{u},\mathbf{v}) = (u_1,\ldots,u_n,v_1,\ldots,v_n)$ is a point in the dual of phase space; it is related to the usual Fourier transform via

$$\tilde{g}(a) = \tilde{g}(\mathbf{u},\mathbf{v}) = (\mathcal{F}g)(\mathbf{v},-\mathbf{u}) . \tag{2.9}$$

Because of this simple relationship, many of the usual properties of the Fourier transform carry over to symplectic Fourier transforms, with only minor modifications being necessary. For example, the inversion formula becomes

$$g(z) = \frac{1}{(2\pi)^{2n}} \int \tilde{g}(a) e^{i\sigma(z,a)} \, dv(a) . \tag{2.10}$$

Another equally important formula is the one that relates pointwise products and convolution products. If f and g are functions defined on phase space, then the convolution of f and g is

$$f*g(z) \equiv \int f(z') g(z-z') \, dv(z') , \tag{2.11}$$

and the symplectic Fourier transform of $f*g$ is

$$\widetilde{f*g}(a) = \tilde{f}(a)\tilde{g}(a) . \tag{2.12}$$

The final ingredient needed for stating the KLM conditions is the concept of an $\hbar$-positive-type function.[3,6,9] For my purposes, it is better to regard $\hbar$ as a parameter rather than a fixed physical constant. Consequently, I will define η-positive-type functions; these reduce to $\hbar$-positive-type when one sets $\eta = \hbar$, and to (Bochner) positive-type functions[12] when $\eta = 0$.

Let F be a function defined and continuous on the dual of phase space. The function F will be termed of *η-positive type* if, for every finite set of points $\{a_1,\ldots,a_m\}$ in the dual of phase space, the $m\times m$ matrix M with entries

$$M_{jk} = F(a_j - a_k) e^{(i\eta/2)\sigma(a_k,a_j)} \tag{2.13}$$

is self-adjoint and non-negative.

A function $\tilde{P}(a)$ is the symplectic Fourier transform of a WDF P if and only if it satisfies the KLM conditions

(i) $\tilde{P}(a)$ is continuous and of $\hbar$-positive type,

(ii) $\tilde{P}(0) = 1$.

Similar conditions also characterize classical states, which are simply probability measures defined on phase space: $\tilde{P}(a)$ is the symplectic Fourier transform of a probability measure if and only if these hold:

(i') $\tilde{P}(a)$ is continuous and of 0-positive type,

(ii') $\tilde{P}(0) = 1$.

Since functions of 0-positive type are identical with Bochner's class of functions of positive type, and the conditions listed are nothing more than a restatement of Bochner's famous theorem,[12] I will call (i') and (ii') the Bochner conditions.

III. GAUSSIANS

Gaussian distributions play a prominent role in classical statistical mechanics. In the phase-space formulation of quantum mechanics, they arise in connection with the ground state of an harmonic oscillator.[2] As I pointed out in Sec. I, Simon *et al.*[7] have recently characterized those Gaussians that are also WDF's. In this section, I will find all η for which a given Gaussian has a symplectic Fourier transform of η-positive type. Using the results I obtain, I will then employ the KLM conditions to characterize the Gaussians that are also WDF's. What I obtain will be conditions that are equivalent to the ones in Ref. 7, but that are simpler to state and easier to apply.

A typical Gaussian centered at a point $z_0 = (\mathbf{q}_0,\mathbf{p}_0)$ has the form

$$P_A(z) = \pi^{-n} (\det A)^{1/2} \exp(-(z-z_0)^T A (z-z_0)) , \tag{3.1}$$

where A is a real, symmetric, positive definite $2n\times 2n$ matrix. This A must be positive definite—i.e., have only positive eigenvalues—to ensure that $P_A(z)$ is integrable. The factors in front of the exponential normalize P_A so that

$$\int P_A(z)dv(z) = 1\,. \tag{3.2}$$

The symplectic Fourier transform of P_A is easy to calculate; the result is

$$\widetilde{P}_A(a) = \exp(-\tfrac{1}{4}a^TBa + i\sigma(a,z_0))\,, \tag{3.3}$$

where, with β as in (2.4),

$$B = -\beta A^{-1}\beta\,. \tag{3.4}$$

Note that $\widetilde{P}_A(0) = 1$, so that (ii) of the KLM conditions is always satisfied.

I now want to look for all values of η that make $\widetilde{P}_A(a)$ of η-positive type. Clearly, $\widetilde{P}_A$ is continuous, and it will be of η-positive type if and only if the matrix in (2.13) is self-adjoint and non-negative. By inserting (3.3) into (2.13) and then doing a little algebra, one finds that

$$M_{jk} = \widetilde{P}_A(a_j)\widetilde{P}_A(a_k)^* \exp(\tfrac{1}{2}a_k^T(B + i\eta\beta)a_j)\,. \tag{3.5}$$

In deriving (3.5), I used the assumption that A is a real, symmetric matrix to find that B is real and symmetric, which in turn allows me to replace $a_j^TBa_k$ by $a_k^TBa_j$.

Let N be the matrix with entries

$$N_{jk} = \exp(\tfrac{1}{2}a_k^T(B + i\eta\beta)a_j)\,. \tag{3.6}$$

Observe that M and N are related this way:

$$M = DND^\dagger\,, \quad N = D^{-1}M(D^{-1})^\dagger\,, \tag{3.7}$$

where $D = \mathrm{diag}(\widetilde{P}_A(a_1),...,\widetilde{P}_A(a_m))$ is invertible because $\widetilde{P}_A(a)$ never vanishes. From (3.7), one sees that M is self-adjoint and non-negative if and only if N is.

The self-adjointness of N is an obvious consequence of B being a real, symmetric matrix and β being a real, antisymmetric matrix. Only the non-negativity of N requires discussion.

Recall that N is non-negative if and only if, for every set of m complex numbers $\{\lambda_1,...,\lambda_m\}$,

$$\sum_{j,k=1}^{m} \lambda_j^*\lambda_k N_{jk} \geqslant 0\,. \tag{3.8}$$

To simplify the notation slightly, let

$$J = B + i\eta\beta\,. \tag{3.9}$$

Using this in (3.6), and then inserting the result in (3.8), one gets

$$\sum_{j,k=1}^{m} \lambda_j^*\lambda_k \exp\left(\frac{1}{2}a_k^TJa_j\right) \geqslant 0\,. \tag{3.10}$$

Next, replace the exponential in (3.10) by its power series, and then split off the zeroth-order term; (3.10) becomes

$$\left|\sum_{j=1}^{m}\lambda_j\right|^2 + \sum_{\nu=1}^{\infty}\left\{\sum_{j,k=1}^{m}\frac{\lambda_j^*\lambda_k}{\nu!}\left(\frac{1}{2}a_k^TJ_a\right)^\nu\right\} \geqslant 0\,. \tag{3.11}$$

I claim that a *necessary and sufficient condition for* $\widetilde{P}_A$ *to be of* η*-positive type is that the matrix* J *be non-negative.* To see that this is sufficient, first observe that the $m\times m$ matrix L with entries

$$L_{jk} = a_k^TJa_j \tag{3.12}$$

satisfies

$$\sum_{j,k=1}^{m}\lambda_j^*L_{jk}\lambda_k = a^\dagger Ja\,, \quad a = \sum_{k=1}^{m}\lambda_k^*a_k\,. \tag{3.13}$$

Hence, if J is non-negative, L will be, too. But if L is non-negative, the matrices with entries that are powers of $L_{jk} - L_{jk}^2, L_{jk}^3$, etc., will also be non-negative: Each of them is the Schur product (see Ref. 3, p. 3, and Ref. 12) of non-negative matrices, and so each is itself non-negative. This means that

$$\sum_{j,k=1}^{m}\lambda_j^*\lambda_k(a_k^TJa_j)^\nu \geqslant 0\,, \tag{3.14}$$

for every $\nu \geqslant 1$. Dividing (3.14) by $2^\nu\nu!$, summing from $\nu = 1$ to ∞, and adding the term $|\Sigma\lambda_j|^2$, one recovers (3.11). Since the a's and λ's are arbitrary, one has that $\widetilde{P}_A$ is of η-positive type.

Conversely, suppose that (3.11) holds for arbitrary a's and λ's, and in particular, when $m = 3$, for

$$\lambda_1 = 1\,, \quad \lambda_2 = i\,, \quad \lambda_3 = i - 1\,,$$
$$a_1 = \epsilon b_1\,, \quad a_2 = \epsilon b_2\,, \quad a_3 = 0\,,$$

where $\epsilon > 0$ and both b_1 and b_2 are just arbitrary points in the dual of phase space. Inserting these in (3.11), dividing out ϵ^2, and then letting $\epsilon \to 0$, one gets

$$\sum_{j,k=1}^{3}\lambda_j^*\lambda_k(b_k^TJb_j) \geqslant 0\,.$$

This is equivalent to

$$(b_1 + ib_2)^\dagger J(b_1 + ib_2) \geqslant 0\,. \tag{3.15}$$

Since $b_1 + ib_2$ is an arbitrary $2n$-dimensional *complex* vector, J is non-negative, which is what I needed to complete the proof of the assertion.

The set of all η for which $J = B + i\eta B$ is non-negative has simple structure. First of all, if η is in it, then so is $-\eta$. The reason for this is that J and $J^T = B - i\eta\beta$ have the same eigenvalues, and so the non-negativity of one is equivalent to that of the other. Second, if $0 \leqslant r \leqslant 1$ and if η is a value that makes J non-negative, then $r\eta$ is also such a value. This follows from observing that

$$B + i\eta r\beta = (1 - r)B + r(B + i\eta\beta)$$

is non-negative because $0 \leqslant r \leqslant 1$ and both of the matrices B and $B + i\eta\beta$ are non-negative. Taken together, these facts imply that the set has the structure $[-\eta_0,\eta_0]$, where η_0 is the largest value of η for which J is non-negative.

As an example, take $A = (1/c)I$, and note that $B = -c\beta^2 = cI$. Thus

$$J = cI + i\eta\beta\,. \tag{3.16}$$

The eigenvalues of $i\beta$ are ± 1, and so those of J are $c \pm \eta$. Thus $J \geqslant 0$ for η if and only if

$$\eta \in [-c,c]\,. \tag{3.17}$$

The question of whether P_A is a WDF hinges on whether $\hbar$ belongs to the set described above, that is, on whether $\widetilde{P}_A$ is of $\hbar$-positive type. My analysis shows that this will be the case if and only if

$$J = -\beta A^{-1}\beta + i\hbar\beta \tag{3.18}$$

is non-negative. Since $-\beta^3 = \beta$, and $\beta^T = -\beta$, one may rewrite (3.18) as

$$\beta^TJ\beta = A^{-1} + i\hbar\beta\,,$$

which will be non-negative if and only if J is. Thus one has that P_A is a WDF if and only if $A^{-1} + i\hbar\beta$ is a non-negative matrix. In the special case treated earlier, namely $A = c^{-1}I$, P_A will be a WDF if and only if $\hbar \leqslant c$.

Two comments are in order. First, one can show that the condition that $B + i\hbar\beta$ be non-negative is invariant under a linear, symplectic change of coordinates in $P_A(z)$. Second, one can put the conditions for P_A to be a WDF in a form identical to that of Ref. 7. I omit the details.

IV. SUFFICIENT CONDITIONS ON P_0

One of the important results that Jagannathan *et al.*[1] obtained was that for the map $P \to P_0 * P$ to result in a pointwise non-negative WDF it is sufficient for P_0 to be a pointwise non-negative WDF itself. They established the result by exploiting the connection between WDF's and density matrices.

There is another way to get the same result. From (2.12), one has that

$$\widetilde{P_0 * P}(a) = \widetilde{P}_0(a)\widetilde{P}(a) . \tag{4.1}$$

The fact that P_0 is a WDF implies that $\widetilde{P}_0(a)$ satisfies the KLM conditions, so $\widetilde{P}_0(a)$ is continuous and of $\hbar$-positive type, and $\widetilde{P}_0(0) = 1$. On the other hand, P_0 is non-negative, so $\widetilde{P}_0$ must be of 0-positive type (i.e., of ordinary "Bochner" positive type). What needs to be done is to show that $\widetilde{P}_0(a)\widetilde{P}(a)$ satisfies the same conditions that $\widetilde{P}_0$ does, since these guarantee that $P_0 * P$ is a non-negative WDF.

Since P_0 and P are WDF's, they are normalized so that $\widetilde{P}_0(0) = \widetilde{P}(0) = 1$. Obviously, $\widetilde{P}_0(a)\widetilde{P}(a)$ is, too. What remains is the more difficult task of showing that $\widetilde{P}_0(a)\widetilde{P}(a)$ is both of 0-positive type and $\hbar$-positive type.

As in the case for $\widetilde{P}_0$ above, a function $F(a)$ may be of η-positive type for several values of η. It is useful to collect all these values into a set, and to examine the properties of that set. Hence let $\mathscr{W}(F)$ be the collection of all real η for which F is of η-positive type. I will call $\mathscr{W}(F)$ the *Wigner spectrum of* F, and I will now examine a few of its properties.

First of all, *if S is a symplectic matrix, then the functions $F(a)$ and $F_S(a) = F(Sa)$ have the same Wigner spectrum; that is,*

$$\mathscr{W}(F) = \mathscr{W}(F_s) . \tag{4.2}$$

Recall that if F is of η-positive type, then the $m \times m$ matrix M with entries given by (2.13) must be self-adjoint and non-negative for every finite set $\{a_1,...,a_m\}$ of points in the dual of phase space. Let $b_1 = S^{-1}a_1$, $b_2 = S^{-1}a_2,...,b_m = S^{-1}a_m$, and note that, because S is symplectic [see (2.5)],

$$\sigma(a_k,a_j) = \sigma(b_k,b_j) . \tag{4.3}$$

From (2.13) and (4.3), one sees that

$$M_{jk} = F_S(b_j - b_k)e^{(i\eta/2)\sigma(b_k,b_j)} . \tag{4.4}$$

However, M is already non-negative and self-adjoint, and the set of b's is clearly arbitrary, so F_s is of η-positive type and $\eta \in \mathscr{W}(F_s)$. Since $F = (F_S)_{S^{-1}}$, if $\eta \in \mathscr{W}(F_S)$, the argument above shows that $\eta \in \mathscr{W}(F)$, which proves that (4.2) is true.

The invariance of the Wigner spectrum under linear symplectic transformations has a simple but important consequence. *If $\eta \in \mathscr{W}(F)$, then $-\eta \in \mathscr{W}(F)$.* Let $S = -I$. This is clearly a symplectic matrix, so $\eta \in \mathscr{W}(F_S)$. Thus, for every set $\{a_1,...,a_m\}$ in the dual of phase space, the matrix N with entries

$$N_{jk} = F_S(a_j - a_k)e^{(i\eta/2)\sigma(a_k,a_j)} \tag{4.5}$$

is non-negative and self-adjoint. Since $F_S(a) = F(-a)$ and $\sigma(a_k,a_j) = -\sigma(a_j,a_k)$, one may rewrite (4.5) as

$$N_{jk} = F(a_k - a_j)e^{-(i\eta\sigma/2)(a_j,a_k)} . \tag{4.6}$$

From (4.6), one sees that

$$(N^T)_{jk} = F(a_j - a_k)e^{-(i\eta\sigma/2)(a_k,a_j)} . \tag{4.7}$$

Taking the transpose preserves self-adjointness and non-negativity, so N^T has both of these properties, and it has them for every finite subset $\{a_1,...,a_m\}$ of the dual of phase space. Clearly, F is of $(-\eta)$-positive type, and thus $-\eta \in \mathscr{W}(F)$.

There is an important connection between the Wigner spectrum of a product of two functions and the "sum" of the spectra of those functions. *If F,G are continuous functions defined on the dual of phase space, then*

$$\mathscr{W}(FG) \supseteq \mathscr{W}(F) + \mathscr{W}(G) . \tag{4.8}$$

The set on the right in (4.8) comprises all real numbers of the form $\eta + \eta'$, with $\eta \in \mathscr{W}(F)$ and $\eta' \in \mathscr{W}(G)$. To see that (4.8) holds, observe that for every set $\{a_1,...,a_m\}$, both $m \times m$ matrices M—given by (2.13)—and M', with entries

$$M'_{jk} = G(a_j - a_k)e^{(i\eta'/2)\sigma(a_k,a_j)} , \tag{4.9}$$

are non-negative and self-adjoint. The Schur product of these matrices is the matrix M'' with entries

$$M''_{jk} = M_{jk}M'_{jk}$$

$$= F(a_j - a_k)G(a_j - a_k)e^{[i(\eta+\eta')/2]\sigma(a_k,a_j)} , \tag{4.10}$$

and it is also a non-negative and self-adjoint $m \times m$ matrix (see Ref. 3, p. 3, and Ref. 12). Once again one sees that the product FG is of $(\eta + \eta')$-positive type; hence $\eta + \eta' \in \mathscr{W}(FG)$, and (4.8) holds.

The Wigner spectrum has one other interesting property: *$\mathscr{W}(F)$ is a compact (closed, bounded) subset of the real numbers.* Since this plays no role in the present discussion, I will not include a proof of this statement.

I now can complete the task of showing that $P_0 * P$ is a non-negative Wigner function for every WDF P, provided that P_0 is a non-negative WDF. Recall that I had reduced the problem to showing that $\widetilde{P}_0(a)\widetilde{P}(a)$ was both of 0-positive type and of $\hbar$-positive type. In the language of Wigner spectra, I want to show that both 0 and $\hbar$ are in $\mathscr{W}(\widetilde{P}_0\widetilde{P})$.

This is easy to do. Here $\widetilde{P}_0$ is of 0-positive type and of $\hbar$-positive type. Consequently, by the discussion above, it is also of $(-\hbar)$-positive type. Hence it is true that

$$\{0,\hbar,-\hbar\} \subseteq \mathscr{W}(\widetilde{P}_0) . \tag{4.11}$$

On the other hand, all that is known about $\widetilde{P}$ is that P is a WDF. Again, from the KLM conditions one has that $\widetilde{P}$ is of $\hbar$-positive type, and, as above, that $\widetilde{P}$ is also of $(-\hbar)$-positive type. Again, the following holds:

$$\{\hbar,-\hbar\} \subseteq \mathscr{W}(\widetilde{P}) . \tag{4.12}$$

From (4.8), (4.11), and (4.12), one gets that

$$\mathscr{W}(\tilde{P}_0\tilde{P}) \supseteq \{-2\hbar, -\hbar, 0, \hbar, 2\hbar\}, \tag{4.13}$$

and so $P_0 * P$ is a non-negative WDF.

One can also use a similar demonstration to show that if P_0 and P are WDF's, then $P_0 * P$ will be pointwise non-negative, although it may not be a WDF.[1] By Bochner's theorem, $P_0 * P$ being non-negative is equivalent to $\tilde{P}_0(a)\tilde{P}(a)$ being of 0-positive type. If P_0 and P are WDF's, all that one knows is that $+\hbar$ and $-\hbar$ are in both $\mathscr{W}(\tilde{P}_0)$ and $\mathscr{W}(\tilde{P})$. By (4.8), one then has that

$$\mathscr{W}(\tilde{P}_0\tilde{P}_2) \supseteq \{-2\hbar, 0, 2\hbar\}. \tag{4.14}$$

Since $0 \in \mathscr{W}(\tilde{P}_0\tilde{P})$, $\tilde{P}_0(a)\tilde{P}(a)$ is, as required, of 0-positive type, and $P_0 * P$ is a pointwise non-negative function.

Notice that $\hbar$ is missing from the right side of (4.14). This absence is reflected in there being WDF's for which $P_0 * P$ is *not* a WDF, which was found by Jagannathan *et al.*[1]

An obvious thing to do now is to let P_0 be an arbitrary WDF, and then try to determine what numbers $\mathscr{W}(\tilde{P}_0)$ must contain in order that $P_0 * P$ be a non-negative WDF for every WDF P. As before, letting both P_0 and P be WDF's guarantees that $\tilde{P}_0(a)\tilde{P}(a)$ is properly normalized; i.e., $\tilde{P}_0(0) \cdot \tilde{P}(0) = 1$. Thus all that needs to be done is to obtain conditions on $\mathscr{W}(\tilde{P}_0)$ that ensure that 0 and $\hbar$ both belong to $\mathscr{W}(\tilde{P}_0\tilde{P})$.

Suppose that

$$\mathscr{W}(\tilde{P}_0) \supseteq \{\eta, -\eta, \hbar, -\hbar\}. \tag{4.15}$$

Since P is an arbitrary WDF, the only information known about it is that $\mathscr{W}(\tilde{P})$ satisfies (4.12). Using (4.8), (4.12), and (4.15), one gets

$$\mathscr{W}(\tilde{P}_0\tilde{P}) \supseteq \{0, \pm 2\hbar, \pm(\eta + \hbar), \pm(\eta - \hbar)\}. \tag{4.16}$$

If $\eta = 0$, (4.14) results. If $\eta = 2\hbar$, then

$$\mathscr{W}(\tilde{P}_0\tilde{P}) \supseteq \{0, \pm 2\hbar, \pm 3\hbar, \pm\hbar\}, \tag{4.17}$$

and $\tilde{P}_0\tilde{P}$ satisfies the KLM conditions, making $P_0 * P$ a WDF.

What all this means is that, *for $P_0 * P$ to be a non-negative WDF for every WDF P, it is sufficient that $\mathscr{W}(P_0)$ contain $\hbar$ and either 0 or $2\hbar$.* This, of course, raises the question of whether there is a WDF P_0 for which $0 \notin \mathscr{W}(\tilde{P}_0)$, and $2\hbar \in \mathscr{W}(P_0)$. If one exists, it would provide us with a WDF P_0 that takes on both positive and negative values, but that forces $P_0 * P$ to be a non-negative WDF whenever P is a WDF.

Such a P_0 does exist, as I will now demonstrate by an explicit construction. To simplify matters, I will work with a quantum mechanical system having one degree of freedom.

The construction begins with the WDF that represents the first excited state of an harmonic oscillator with $m = 1$ and $\omega = 1$. From Ref. 2 (see p. 143), one has

$$P_H(z) = -(1/\pi\hbar)(1 - 2|z|^2/\hbar)e^{-|z|^2/\hbar}, \tag{4.18}$$

where $|z|^2 = p^2 + q^2$. A straightforward, rather standard computation gives

$$\tilde{P}_H(a) = (1 - (\hbar/2)|a|^2)e^{-\hbar|a|^2/4}. \tag{4.19}$$

Two things are clear about $\tilde{P}_H$. First, 0 does *not* belong to the Wigner spectrum of $\tilde{P}_H$ because $P_H(z)$ takes on negative values, and, second, $\hbar$ and $-\hbar$ *do* belong to it because P_H is a WDF.

Carrying out the next step requires adding yet another property of the Wigner spectrum to the list given earlier. This one relates to how it behaves when a function is "scaled." *If λ is a positive number and if F is a continuous function defined in the dual of phase space, then*

$$\mathscr{W}(F(\lambda a)) = \lambda^2 \mathscr{W}(F(a)). \tag{4.20}$$

Establishing (4.20) can be done in a way that is quite similar to the ways used for the other properties, and so I will not bother to do it here. My main reason for introducing this property is that it enables me to obtain that

$$\pm\tfrac{3}{2}\hbar \in \mathscr{W}(\tilde{P}_H(\sqrt{3/2}\, a)), \tag{4.21}$$

but that $0 \notin \mathscr{W}(\tilde{P}_H(\sqrt{3/2}\, a))$.

The third step involves the Gaussians discussed in Sec. III. For P_A with $A = c^{-1}I$, $z_0 = 0$, the Wigner spectrum of $\tilde{P}_A$ was actually found in Sec. III; it is the interval in (3.17). That is,

$$\mathscr{W}(\exp(-(c/4)|a|^2)) = [-c, c]. \tag{4.22}$$

Choosing $c = \hbar/2$ in (4.22) results in

$$\mathscr{W}(\exp(-(\hbar/8)|a|^2)) = [-\hbar/2, \hbar/2]. \tag{4.23}$$

The last step is to form the product

$$\tilde{P}_0(a) \equiv \tilde{P}_H(\sqrt{3/2}\, a)e^{-\hbar|a|^2/8} = (1 - (3\hbar/4)|a|^2)e^{-\hbar|a|^2/2}. \tag{4.24}$$

Observe that from (4.8), (4.21), and (4.23), $\tilde{P}_0(a)$ has both $\hbar$ and $2\hbar$ in its Wigner spectrum. This is enough to guarantee that $P_0(z)$, the inverse symplectic Fourier transform of $\tilde{P}_0$, is a WDF, and that the map $P \to P_0 * P$ takes WDF's to non-negative WDF's.

But $P_0(z)$ takes on negative values. A straightforward computation yields

$$P_0(z) = (1/3\pi\hbar)(3|z|^2/2\hbar - 1)e^{-|z|^2/2\hbar}, \tag{4.25}$$

which is negative for all z sufficiently small.

The significance of the existence of a P_0 of the form (4.25) is that it shows the conditions given in Ref. 1 for the map $P \to P_0 * P$ to take WDF's to non-negative WDF's are only sufficient, and *not* necessary ones. Although I cannot give a proof, I believe that the sufficient conditions derived here *are* in fact necessary, too.

V. DISCUSSION

The main result of this paper is this: If P_0 is a WDF, the map $P \to P_0 * P$ takes WDF's to WDF's if either 0 or $2\hbar$ belongs to the Wigner spectrum of $\tilde{P}_0$. By way of comparison, the chief result of Ref. 1, in the language introduced here, is that $P \to P_0 * P$ has the desired mapping properties if 0 is in $\tilde{P}_0$'s Wigner spectrum.

In other words, $0 \in \mathscr{W}(\tilde{P}_0)$ is a sufficient condition for $P \to P_0 * P$ to behave as required. The existence of the WDF P_0 given in (4.25) shows that the condition derived in Ref. 1 is not a necessary one, since $\mathscr{W}(\tilde{P}_0)$ contains $\pm 2\hbar$ but not 0.

Of course, it is natural to ask whether the sufficient conditions derived in Sec. IV are necessary as well. I think they

are, and I will now present an argument that supports my belief.

There are two propositions that, if true, would imply necessity. The first is that (4.8) holds with the inclusion replaced by an equality; that is,

$$\mathscr{W}(FG) = \mathscr{W}(F) + \mathscr{W}(G) . \tag{5.1}$$

The second is that there is a family of WDF's $P_\epsilon(z)$ defined for all sufficiently small $\epsilon > 0$ and satisfying

$$\mathscr{W}(\tilde{P}_\epsilon) \subseteq [-\hbar - \epsilon, -\hbar + \epsilon] \cup [\hbar - \epsilon, \hbar + \epsilon] . \tag{5.2}$$

That necessity is a consequence of these two is not hard to see. If P_0 is a WDF such that $P_0 * P$ is also a WDF as long as P is, then, by the KLM conditions, $\tilde{P}_0(a)\tilde{P}(a)$ is always of $\hbar$-positive type. Thus

$$\hbar \in \mathscr{W}(\tilde{P}_0 \tilde{P}) . \tag{5.3}$$

In particular, from (5.1) one has that, for each of the P_ϵ's,

$$h \in \mathscr{W}(\tilde{P}_0 \tilde{P}_\epsilon) = \mathscr{W}(\tilde{P}_0) + \mathscr{W}(\tilde{P}_\epsilon) . \tag{5.4}$$

Combining (5.4) and (5.2) and letting $\epsilon \to 0$, one arrives at

$$\hbar \in \mathscr{W}(\tilde{P}_0) + \{-\hbar, \hbar\} , \tag{5.5}$$

which is possible if and only if either 0 or $\pm 2\hbar$ belong to $\mathscr{W}(\tilde{P}_0)$.

This argument reduces the question of necessity to showing that (5.1) and (5.2) are true. Since both propositions are plausible, necessity is plausible, too. Of course, whether or not these propositions are true is an open question.

[1] R. Jagannathan, R. Simon, E. C. G. Sudarshan, and R. Vasudevan, Phys. Lett. A **120**, 161 (1987).
[2] M. Hillery, R. F. O'Connell, M. O. Scully, and E. P. Wigner, Phys. Rep. **106**(3), 121 (1984).
[3] F. J. Narcowich and R. F. O'Connell, Phys. Rev. A **34**, 1 (1986).
[4] K. Husimi, Proc. Phys. Math. Soc. Jpn. **22**, 264 (1940).
[5] R. F. O'Connell and E. P. Wigner, Phys. Lett. A **25**, 121 (1981); R. F. O'Connell and A. K. Rajagopal, Phys. Rev. Lett. **48**, 525 (1982); R. F. O'Connell and D. F. Walls, Nature **312**, 257 (1985).
[6] D. Kastler, Commun. Math. Phys. **1**, 14 (1965).
[7] R. Simon, E. C. G. Sudarshan, and N. Mukunda, Phys. Lett. A **124**, 223 (1987).
[8] J. Williamson, Am. J. Math. **58**, 141 (1936).
[9] G. Loupias and S. Maracle-Sole, Commun. Math. Phys. **2**, 31 (1966); Ann. Inst. H. Poincaré **6**, 39 (1967).
[10] V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1978).
[11] H. Pollard, *Mathematical Introduction to Celestial Mechanics* (Prentice–Hall, Englewood Cliffs, NJ, 1966).
[12] I. M. Gel'fand and N. Ya. Vilenkin, *Generalized Functions*, Vol. 4 (Academic, New York, 1964); Y. Katznelson, *An Introduction to Harmonic Analysis* (Wiley, New York, 1968).

Berry phases and unitary transformations

Thomas F. Jordan
Physics Department, University of Minnesota, Duluth, Minnesota 55812

(Received 4 August 1987; accepted for publication 4 May 1988)

Berry phases for spin are defined for any closed loop made by a vector changing direction in three-dimensional space. A sequence of rotations moves the vector along the loop. Each rotation is around the axis perpendicular to the moving vector. The Berry phases come from the eigenvalues of the unitary operator for the product of these rotations. The angle of the product rotation is shown to be the solid angle enclosed by the loop. The proof uses the ordinary language of quantum mechanics. The product is calculated from the commutation relations for spin. A general framework is set up to define Berry phases for other transformations and states like those for rotations and spin. The integral formula is derived. Alternatives for dynamics are shown to provide different applications and interpretations of the same mathematics. An example is used to show how one Hamiltonian may be simpler than others. Adiabatic evolution is obtained in the limit as a coupling constant goes to zero, so the adiabatic changes are made by a weak perturbation that acts over a long time. The Berry phase is the same whether the coupling constant is large or small. A rationale for the definition of Berry phases is obtained from any of the alternatives for dynamics. It is particularly clear in both the limit of adiabatic evolution, where the Hamiltonian does not contain the operator that generates Berry phases, and the opposite extreme where the Hamiltonian is the operator that generates Berry phases. The latter is illustrated with an example for spin. The general definition of Berry phases and the method of calculating them are illustrated by obtaining Berry phases from Lorentz transformations. They are similar to those obtained from rotations. A vector traces a loop on a unit hyperboloid instead of the unit sphere. In place of the solid angle, the Berry phases contain the analogous measure of the area enclosed by the loop on the hyperboloid. The sign is opposite what it is for rotations. The general definition is shown to fit any unitary representation of a semisimple Lie group. The complete set of commuting operators is chosen to contain a basis for a Cartan subalgebra of the Lie algebra of generators. If a sequence of transformations by unitary operators in the group representation takes each operator in the Cartan subalgebra around a loop in the Lie algebra back to the same operator, it can be made by unitary operators in the group representation that have all the properties required for the definition of Berry phases. The loops are made by vectors moving in a real space, the Lie algebra. The motion of the vectors must maintain the lengths and angles defined by the Cartan metric of the Lie algebra. The Berry phases are determined by one loop made by one suitably chosen operator from the Cartan subalgebra.

I. INTRODUCTION

A Berry phase[1-7] for spin is the result of a sequence of rotations. Consider a familiar example.[1,3,4,6,7] We watch the spin component $\hat{k}(t) \cdot \mathbf{S}$ in the direction of a unit vector $\hat{k}(t)$ that changes as a function of time. This component has a constant value. At each time t the spin state is represented by an eigenvector of $\hat{k}(t) \cdot \mathbf{S}$. Between times 0 and T the vector $\hat{k}(t)$ goes around a closed loop, so $\hat{k}(T)$ is the same as $\hat{k}(0)$.

We can guess what happens. From time t to $t + \Delta t$ the state is changed by a rotation that takes $\hat{k}(t)$ to $\hat{k}(t + \Delta t)$. The sequence of these rotations from time 0 to time T takes $\hat{k}(0)$ around the loop back to $\hat{k}(0)$. The product of all these rotations is a rotation that does not change $\hat{k}(0)$, so it is a rotation around the axis in the direction of $\hat{k}(0)$. The state at time 0 is not changed by this rotation because it is represented by an eigenvector of $\hat{k}(0) \cdot \mathbf{S}$. The product rotation just multiplies this state vector by a phase factor. The phase is minus the angle of the product rotation multiplied by the eigenvalue of $\hat{k}(0) \cdot \mathbf{S}$. This is the Berry phase.

There are some questions to be answered. Which rotations? At each time t there are many different rotations that take $\hat{k}(t)$ to $\hat{k}(t + \Delta t)$. Which one do we use? We choose the most direct one, the rotation around the axis perpendicular to $\hat{k}(t)$ and $\hat{k}(t + \Delta t)$. This is part of our definition of Berry phases. A different choice would put another rotation into the sequence. It would be a rotation around the axis in the direction of $\hat{k}(t)$. Its angle would add to the angle of the product rotation, so it would change the Berry phases. The definition of Berry phases excludes it. This definition of Berry phases is described in Sec. II A. With the choice of rotations thus made, the Berry phases are determined by the loop made by $\hat{k}(t)$ between times 0 and T. We can start with a given loop and construct the sequence of rotations.

How do we multiply these rotations? The angle of the product rotation is the solid angle circumscribed by $\hat{k}(t)$ as it goes around the loop.[1,6] This is purely a property of rotations. It is proved in Sec. II B by a method that uses the ordinary language of quantum mechanics. The rotations are described by their representation in terms of spin. The loop is

 0022-2488/88/092042-11$02.50

divided into infinitesimal loops and the product of the rotations for each infinitesimal loop is calculated from the commutation relations for spin.

A general framework is set up in Sec. II C to define Berry phases for other transformations and states like those for rotations and spin. The generalization of our choice of the rotation from $\hat{k}(t)$ to $\hat{k}(t+\Delta t)$ is the condition introduced by Anandan and Stodolsky to make the transformations unique.[7]

Does our definition of the Berry phase agree with Berry's integral formula? Certainly. The integral formula is derived in Sec. II D.

What is the role of dynamics? All this can be formulated without any reference to dynamics, without considering time or a Hamiltonian. That is what we do in Secs. II A–II D. Then in Sec. II E we describe alternatives for dynamics that provide different applications and interpretations of the same mathematics. One Hamiltonian may be simpler than others; an example of Aharonov and Anandan[8] is used to illustrate that.

Adiabatic evolution is obtained in the limit as a coupling constant goes to zero, so the adiabatic changes are made by a weak perturbation that acts over a long time. The Berry phase is the same whether the coupling constant is large or small. The limit is needed only to get adiabatic evolution.

A rationale for our definition of Berry phases is obtained from any of our alternatives for dynamics. The definition fits the dynamics, whichever the dynamics may be. The rationale is particularly clear for two kinds of dynamics. One is obtained in the limit of adiabatic evolution where the Hamiltonian does not contain the operator that generates Berry phases. The other is at the opposite extreme where the Hamiltonian is the operator that generates Berry phases. The latter is illustrated with an example for spin. It is a modification of an example of Aharonov and Anandan that forgoes the simplicity of their Hamiltonian but makes the Berry phases stand out more clearly. This is a particularly clean example of Berry phases produced in evolution that is not adiabatic with states that are not represented by eigenvectors of the Hamiltonian.

In Sec. III, our general definition of Berry phases and our method of calculating them are illustrated by obtaining Berry phases from Lorentz transformations. In Sec. IV, the definition is shown to fit any unitary representation of a semisimple Lie group. The results are described in the first paragraphs of Secs. III and IV.

II. BERRY PHASES AND ROTATIONS

A. Berry phases

Here we consider Berry phases for spins. They are the same for helicities, so they are the same whether the particle with spin is nonrelativistic or relativistic or whether it has nonzero or zero mass.[6] Berry phases are described in a more general framework in Secs. II C–II E.

Consider a spin represented by matrices $\mathbf{S}$. Let $\mathbf{k}(q)$ be a real three-vector that depends on a real variable q so that $\mathbf{k}$ is a differentiable function of q. Let $\hat{k}(q)$ be the unit vector in the direction of $\mathbf{k}(q)$. Then $\hat{k}(q)\cdot\mathbf{S}$ represents the projection of the spin in that direction.

We can change $\hat{k}(q)$ to $\hat{k}(q+\Delta q)$ with a rotation around the axis perpendicular to $\hat{k}(q)$ and $\hat{k}(q+\Delta q)$, so we can change $\hat{k}(q)\cdot\mathbf{S}$ to $\hat{k}(q+\Delta q)\cdot\mathbf{S}$ with the unitary operator for this rotation. To describe this rotation we write $\mathbf{k}$ for $\mathbf{k}(q)$ and $\Delta\mathbf{k}$ for $(d\mathbf{k}/dq)\Delta q$. Then $\mathbf{k}+\Delta\mathbf{k}$ is $\mathbf{k}(q+\Delta q)$. The axis of the rotation is the direction of $\mathbf{k}\times\Delta\mathbf{k}$. The angle is the magnitude of $k^{-2}\mathbf{k}\times\Delta\mathbf{k}$. The unitary operator for this rotation is $e^{-ik^{-2}\mathbf{k}\times\Delta\mathbf{k}\cdot\mathbf{S}}$. It changes $\hat{k}(q)\cdot\mathbf{S}$ to

$$\hat{k}(q+\Delta q)\cdot\mathbf{S}=e^{-ik^{-2}\mathbf{k}\times\Delta\mathbf{k}\cdot\mathbf{S}}\hat{k}(q)\cdot\mathbf{S}e^{ik^{-2}\mathbf{k}\times\Delta\mathbf{k}\cdot\mathbf{S}}. \quad (2.1)$$

Suppose that as q runs from 0 to Q the vector $\mathbf{k}(q)$ goes around a closed loop C, so $\mathbf{k}(Q)$ is the same as $\mathbf{k}(0)$. For each interval Δq between 0 and Q we have a rotation around the $\mathbf{k}\times\Delta\mathbf{k}$ axis from $\hat{k}(q)$ to $\hat{k}(q+\Delta q)$. The sequence of these rotations from 0 to Q takes $\hat{k}(0)$ around the loop back to $\hat{k}(0)$. The product of all these rotations is a rotation that does not change $\hat{k}(0)$, so it is a rotation around the axis in the direction of $\hat{k}(0)$. The unitary operator for this rotation is $e^{-i\phi\hat{k}(0)\cdot\mathbf{S}}$, where ϕ is the angle of the product rotation. It is the product of the operators $e^{-ik^{-2}\mathbf{k}\times\Delta\mathbf{k}\cdot\mathbf{S}}$ from 0 to Q.

Let $|m\rangle$ be a normalized eigenvector of $\hat{k}(0)\cdot\mathbf{S}$ for the eigenvalue m. Then

$$e^{-i\phi\hat{k}(0)\cdot\mathbf{S}}|m\rangle=e^{-im\phi}|m\rangle. \quad (2.2)$$

The Berry phases $\gamma_m(Q)$ are defined by

$$e^{i\gamma_m(Q)}=e^{-im\phi}. \quad (2.3)$$

They are determined by the loop C made by $\hat{k}(q)$ as q runs from 0 to Q, because it determines the sequence of rotations.

The angle ϕ of the product rotation is the solid angle Ω circumscribed by $\hat{k}(q)$ as $\hat{k}(q)$ moves around the loop C, the solid angle that will be on your left if you go around the loop with $\hat{k}(q)$. This is proved in the next section. It gives

$$e^{i\gamma_m(Q)}=e^{-im\Omega}. \quad (2.4)$$

For any sequence of rotations, the product is a rotation. Let $\mathbf{k}(0)$ be a vector along the axis of the product rotation. Then the product rotation does not change $\hat{k}(0)$. The sequence of rotations takes $\hat{k}(0)$ around some loop back to $\hat{k}(0)$. The loop may be trivial. It may come back to $\hat{k}(0)$ more than once. In any case, the product rotation is represented by the unitary operator $e^{-i\phi\hat{k}(0)\cdot\mathbf{S}}$ with ϕ the angle of the product rotation. For an eigenvector $|m\rangle$ of $\hat{k}(0)\cdot\mathbf{S}$ we have a phase $\gamma_m(Q)$ defined by (2.2) and (2.3). A sequence of rotations used to define Berry phases is distinguished by one special feature: each rotation between a $\hat{k}(q)$ and $\hat{k}(q+\Delta q)$ is around the $\mathbf{k}\times\Delta\mathbf{k}$ axis. Then ϕ is the solid angle Ω.

This feature is the key assumption in our definition of Berry phases for spin. There are many different rotations that take $\hat{k}(q)$ to $\hat{k}(q+\Delta q)$. We choose the rotation around the $\mathbf{k}\times\Delta\mathbf{k}$ axis. A different choice would put a rotation around $\hat{k}(q)$ in front of the rotation around $\mathbf{k}\times\Delta\mathbf{k}$. It would have no other effect. To see this let $R(\boldsymbol{\theta})$ be the rotation through the angle $|\boldsymbol{\theta}|$ around the axis in the direction of $\boldsymbol{\theta}$. In particular, let $R(\boldsymbol{\theta})$ be the rotation around $\mathbf{k}\times\Delta\mathbf{k}$ that takes $\hat{k}(q)$ to $\hat{k}(q+\Delta q)$ and let $R(\boldsymbol{\theta}')$ be another rotation that takes $\hat{k}(q)$ to $\hat{k}(q+\Delta q)$, so

$$R(\theta')\hat{k}(q) = R(\theta)\hat{k}(q)\,. \tag{2.5}$$

Then

$$R(\theta)^{-1}R(\theta')\hat{k}(q) = \hat{k}(q)\,. \tag{2.6}$$

That means $R(\theta)^{-1}R(\theta')$ is a rotation around $\hat{k}(q)$. Since

$$R(\theta') = R(\theta)R(\theta)^{-1}R(\theta') \tag{2.7}$$

we see that choosing $R(\theta')$ instead of $R(\theta)$ would just put a rotation around $\hat{k}(q)$ in front of $R(\theta)$.

The rotation around $\hat{k}(q)$ would not move $\hat{k}(q)$ along the loop. It would not change $\hat{k}(q)\cdot\mathbf{S}$. It would be represented by a unitary operator that commutes with $\hat{k}(q)\cdot\mathbf{S}$. Nevertheless, it would change the Berry phase. We explain how in language we can also use to generalize from rotations and spin to other transformations and states. Let

$$G(q) = k^{-2}\mathbf{k}\times\frac{d\mathbf{k}}{dq}\cdot\mathbf{S}\,, \tag{2.8}$$

so

$$e^{-ik^{-2}\mathbf{k}\times\Delta\mathbf{k}\cdot\mathbf{S}} = e^{-iG(q)\Delta q}\,, \tag{2.9}$$

and let $U(q)$ be the product of these operators $e^{-iG\Delta q}$ from 0 to q, so $U(0)$ is 1 and

$$U(q+\Delta q) = e^{-iG(q)\Delta q}U(q)\,. \tag{2.10}$$

From (2.1) we can see that

$$\hat{k}(q)\cdot\mathbf{S} = U(q)\hat{k}(0)\cdot\mathbf{S}U(q)^{-1}\,. \tag{2.11}$$

The unitary operator $e^{-i\phi\hat{k}(0)\cdot\mathbf{S}}$ for the product rotation is $U(Q)$. It is the product of all the $e^{-iG(q)\Delta q}$ from 0 to Q. From (2.2) and (2.3) we see that the Berry phases $\gamma_m(Q)$ are defined by

$$U(Q)|m\rangle = e^{i\gamma_m(Q)}|m\rangle\,. \tag{2.12}$$

Consider how $U(Q)$ would change if $R(\theta)$ were changed to $R(\theta')$ for an interval Δq at a particular value of q. A rotation around $\hat{k}(q)$ would be put in front of the rotation from $\hat{k}(q)$ to $\hat{k}(q+\Delta q)$. Let the angle of the rotation around $\hat{k}(q)$ be $\delta(q)\Delta q$. Then the unitary operator for this rotation is $e^{-i\delta(q)\Delta q\hat{k}(q)\cdot\mathbf{S}}$. It would be put into the sequence of operators that are multiplied together to make $U(Q)$, so $U(Q)$ would be changed to

$$\begin{aligned} U'(Q) &= U(Q)U(q)^{-1}e^{-i\delta(q)\Delta q\hat{k}(q)\cdot\mathbf{S}}U(q) \\ &= U(Q)e^{-i\delta(q)\Delta q\hat{k}(0)\cdot\mathbf{S}}\,. \end{aligned} \tag{2.13}$$

If there were similar changes for the other intervals Δq, altogether $U(Q)$ would be changed to

$$U'(Q) = U(Q)\exp\left(-i\int_0^Q \delta(q)dq\,\hat{k}(0)\cdot\mathbf{S}\right). \tag{2.14}$$

From (2.12) we see that the Berry phases $\gamma_m(Q)$ would be changed to

$$\gamma'_m(Q) = \gamma_m(Q) - m\int_0^Q \delta(q)\,dq\,. \tag{2.15}$$

Thus we see that an important part of our definition of Berry phases is the assumption that each rotation from a $\hat{k}(q)$ to $\hat{k}(q+\Delta q)$ is around the $\mathbf{k}\times\Delta\mathbf{k}$ axis. We can describe this assumption in a form that is easy to generalize. A change from $R(\theta)$ to $R(\theta')$ would mean a change from

$$\theta = k^{-2}\mathbf{k}\times\Delta\mathbf{k} = k^{-2}\mathbf{k}\times\frac{d\mathbf{k}}{dq}\Delta q \tag{2.16}$$

to

$$\theta' = k^{-2}\mathbf{k}\times\frac{d\mathbf{k}}{dq}\Delta q + \hat{k}(q)\delta(q)\Delta q\,. \tag{2.17}$$

Then $G(q)$ defined by (2.8) would be changed to

$$G'(q) = k^{-2}\mathbf{k}\times\frac{d\mathbf{k}}{dq}\cdot\mathbf{S} + \delta(q)\hat{k}(q)\cdot\mathbf{S}\,. \tag{2.18}$$

This is what would be allowed if we did not assume that each rotation from a $\hat{k}(q)$ to $\hat{k}(q+\Delta q)$ is around the $\mathbf{k}\times\Delta\mathbf{k}$ axis. The only requirement would be that δ is a differentiable real function of q. What we are assuming is that $\delta(q)$ is zero for all q from 0 to Q. From (2.11) we can see that $U(q)|m\rangle$ is an eigenvector of $\hat{k}(q)\cdot\mathbf{S}$ for the eigenvalue m, so

$$\langle m|U(q)^{-1}G'(q)U(q)|m\rangle = m\delta(q)\,. \tag{2.19}$$

What we are assuming is that this is zero. In terms of $G(q)$, what we are assuming is

$$\langle m|U(q)^{-1}G(q)U(q)|m\rangle = 0\,, \tag{2.20}$$

for all m and all q from 0 to Q.

This is the form of the assumption that we will generalize. The rationale for it will be described further in the general framework in Sec. II C and in terms of dynamics in Sec. II E.

B. Rotations

Here we establish the property of rotations that we used to obtain the Berry phase. Consider the sequence of rotations that moves the direction of the vector $\mathbf{k}$ around the closed loop C from $\hat{k}(0)$ back to $\hat{k}(0)$. The product of these rotations is a rotation that does not change $\hat{k}(0)$, so it is a rotation around the axis in the direction of $\hat{k}(0)$. Suppose each rotation between a $\mathbf{k}$ and $\mathbf{k}+\Delta\mathbf{k}$ is around the axis in the direction of $\mathbf{k}\times\Delta\mathbf{k}$. Then the angle of the product rotation is the solid angle Ω circumscribed by $\mathbf{k}$ as $\mathbf{k}$ moves around the loop C, the solid angle that will be on your left if you go around the loop with $\mathbf{k}$. That is what we have to prove.

We prove it first for an infinitesimal loop at $\mathbf{k}(0)$, then for an infinitesimal loop that is removed from $\mathbf{k}(0)$ except for a connecting path. Finally we show how the infinitesimal loops can be joined to form an arbitrary loop.

Let $\mathbf{k}(0)$ be $\hat{z}$ and let C be the infinitesimal loop made by

$$\begin{aligned} &\Delta\mathbf{k} = \epsilon\hat{x}\,, \quad \Delta\mathbf{k} = \eta\hat{y}\,, \\ &\Delta\mathbf{k} = -\epsilon\hat{x}\,, \quad \Delta\mathbf{k} = -\eta\hat{y}\,, \end{aligned} \tag{2.21}$$

where $\hat{x}$, $\hat{y}$, $\hat{z}$ are unit vectors in the x, y, z directions and ϵ and η are small real numbers. There are four rotations in the sequence. To second order in ϵ and η the vectors $k^{-2}\mathbf{k}\times\Delta\mathbf{k}$ that specify the rotations are

$$\begin{aligned} &\hat{z}\times\epsilon\hat{x} = \epsilon\hat{y}\,, \\ &(\hat{z}+\epsilon\hat{x})\times\eta\hat{y} = -\eta\hat{x}+\epsilon\eta\hat{z}\,, \\ &(\hat{z}+\epsilon\hat{x}+\eta\hat{y})\times(-\epsilon\hat{x}) = -\epsilon\hat{y}+\epsilon\eta\hat{z}\,, \\ &(\hat{z}+\eta\hat{y})\times(-\eta\hat{y}) = \eta\hat{x}\,, \end{aligned} \tag{2.22}$$

so the operators $e^{-ik^{-2}\mathbf{k}\times\Delta\mathbf{k}\cdot\mathbf{S}}$ that represent the rotations are

$$e^{-i\epsilon S_2},\quad e^{i\eta S_1 - i\epsilon\eta S_3},$$
$$e^{i\epsilon S_2 - i\epsilon\eta S_3},\quad e^{-i\eta S_1}. \tag{2.23}$$

The product of these four rotations is the rotation represented by

$$\begin{aligned} &e^{-i\eta S_1}e^{i\epsilon S_2 - i\epsilon\eta S_3}e^{i\eta S_1 - i\epsilon\eta S_3}e^{-i\epsilon S_2} \\ &\quad = 1 - i2\epsilon\eta S_3 + \epsilon\eta(S_1S_2 - S_2S_1) \\ &\quad = 1 - i2\epsilon\eta S_3 + i\epsilon\eta S_3 \\ &\quad = 1 - i\epsilon\eta S_3 = e^{-i\epsilon\eta S_3} = e^{-i\Omega S_3}. \end{aligned} \tag{2.24}$$

This is the rotation by $\epsilon\eta$ around the z axis. Since the solid angle Ω is $\epsilon\eta$ and $\mathbf{k}(0)$ is $\hat{z}$, the product rotation is the rotation through the angle Ω around the axis in the direction of $\mathbf{k}(0)$.

This will not be changed by any change in the length of $\mathbf{k}$ or the component of $\Delta\mathbf{k}$ in the direction of $\mathbf{k}$ as $\mathbf{k}$ moves around the loop. The cone of directions traced by $\mathbf{k}$ is all that matters.

Now consider a loop C made by a path that takes $\hat{k}(0)$ to $\hat{k}(P)$, an infinitesimal loop at $\hat{k}(P)$, and a path right next to the first segment that takes $\hat{k}(P)$ back to $\hat{k}(0)$, as shown in Fig. 1. The product of the rotations for the infinitesimal loop at $\hat{k}(P)$ is the rotation through the angle Ω around the $\hat{k}(P)$ axis, with Ω the solid angle of the infinitesimal loop at $\hat{k}(P)$. The product of the rotations along the path that takes $\hat{k}(0)$ to $\hat{k}(P)$ is a rotation that takes $\hat{k}(0)$ to $\hat{k}(P)$. Let it be represented by $e^{-i\theta\cdot\mathbf{S}}$. The product of the rotations along the final segment that takes $\hat{k}(P)$ to $\hat{k}(0)$ is the inverse of this rotation. The product of the rotations along the entire loop C is the rotation represented by

$$e^{i\theta\cdot\mathbf{S}}e^{-i\Omega\hat{k}(P)\cdot\mathbf{S}}e^{-i\theta\cdot\mathbf{S}} = e^{-i\Omega\hat{k}(0)\cdot\mathbf{S}} \tag{2.25}$$

because

$$e^{-i\theta\cdot\mathbf{S}}\hat{k}(0)\cdot\mathbf{S}e^{i\theta\cdot\mathbf{S}} = \hat{k}(P)\cdot\mathbf{S}. \tag{2.26}$$

The product rotation is the rotation through the angle Ω around the axis in the direction of $\hat{k}(0)$, and Ω is the solid angle enclosed by the entire loop C.

All that remains is to show how infinitesimal loops can be joined to form an arbitrary loop. First, consider a chain of loops joined as shown in Fig. 2. The product of the rotations along any one of the little loops is the rotation around the $\hat{k}(0)$ axis through the angle equal to the solid angle enclosed by the little loop. Then the product of the rotations along all the little loops in succession is the rotation around the $\hat{k}(0)$ axis through the angle equal to the sum of the solid angles enclosed by the little loops. This is the same as the product of the rotations along the larger loop. The rotations along adjoining segments are inverses of each other and cancel out. Therefore the product of the rotation around the larger loop is the rotation around the $\hat{k}(0)$ axis through the angle equal to the solid angle enclosed by the larger loop.

The result is thus established for long thin loops. The proof is completed by putting long thin loops together as shown in Fig. 3.

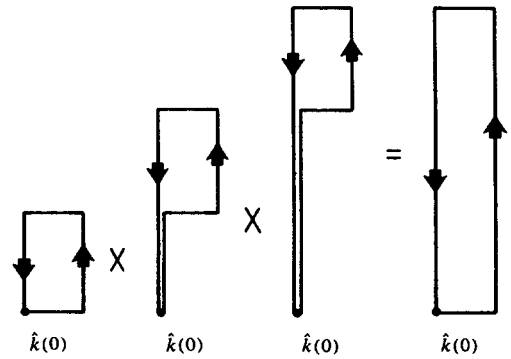


FIG. 2. Joining infinitesimal loops to form a long thin loop.

C. Generalization

The basic outline of what we did for rotations and spin states can be used to define phase changes for other transformations and states. It might also provide a method to calculate the phases. We set up a general framework here.

Consider a complete set of commuting Hermitian operators $A_1, A_2, \ldots$ with an orthonormal basis of eigenvectors. Let m label the different sets of eigenvalues $a_m^{(1)}, a_m^{(2)}, \ldots$ of $A_1, A_2, \ldots$ and for each m let $|m\rangle$ be the eigenvector of $A_1, A_2, \ldots$ for the eigenvalues $a_m^{(1)}, a_m^{(2)}, \ldots$, so

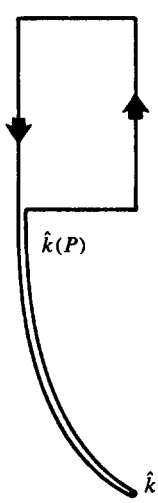


FIG. 1. The loop C around an infinitesimal solid angle Ω that is removed from $\hat{k}(0)$.

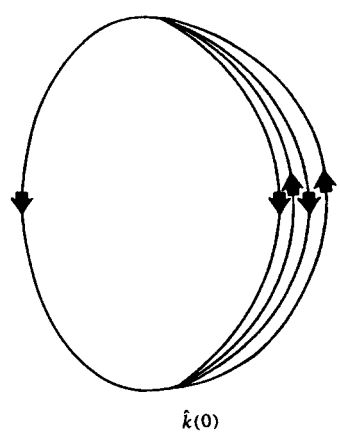


FIG. 3. Joining long thin loops to form an arbitrary loop.

$$A_j|m\rangle = a_m^{(j)}|m\rangle\ . \tag{2.27}$$

Since the set of operators $A_1, A_2, \ldots$ is complete, m is the only label needed for the basis of eigenvectors.

Let $V(q)$ be unitary operators that depend on a real variable q which runs from 0 to Q, with $V(q)$ differentiable as a function of q.

Suppose $V(0)$ is 1 and

$$V(Q)A_jV(Q)^{-1} = A_j \tag{2.28}$$

for each of the operators $A_1, A_2, \ldots$. For each q and each of the operators $A_1, A_2, \ldots$ let

$$A_j(q) = V(q)A_jV(q)^{-1}\ . \tag{2.29}$$

In particular

$$A_j(0) = A_j = A_j(Q)\ . \tag{2.30}$$

The set of operators $A_1(q), A_2(q), \ldots$ is the basis of our construction. It forms a loop as q runs from 0 to Q. This is the generalization of the loop made by $\hat{k}(q)$ or $\hat{k}(q)\cdot\mathbf{S}$ for spin. We could introduce the operators $A_j(q)$ directly, without using the $V(q)$. We just need to describe the properties the $A_j(q)$ should have. We could do that in terms of their spectra and differentiability as functions of q. It is simpler to state it this way. As q runs from 0 to Q, the changes in the operators $A_j(q)$ have to be changes that can be made with unitary operators $V(q)$ that are differentiable as a function of q. Besides that we require only that the $A_j(q)$ satisfy (2.30) with a complete set of commuting Hermitian operators A_j that have a basis of eigenvectors.

Let $U(q)$ be unitary operators that are differentiable as a function of q between 0 and Q and let

$$i\frac{d}{dq}U(q) = G(q)U(q)\ . \tag{2.31}$$

Then $G(q)$ is Hermitian and

$$U(q+\Delta q) = e^{-iG(q)\Delta q}U(q)\ . \tag{2.32}$$

Suppose $U(0)$ is 1 and, for each q from 0 to Q,

$$U(q)A_jU(q)^{-1} = A_j(q) \tag{2.33}$$

for each of the operators $A_1, A_2, \ldots$ and

$$\langle m|U(q)^{-1}G(q)U(q)|m\rangle = 0 \tag{2.34}$$

for all m. There are operators $U(q)$ with these properties; they do exist. They are unique. Thus the $U(q)$ are obtained from the $A_j(q)$. That is proved in the Appendix. If the $V(q)$ satisfy the last requirement (2.34), then they are the $U(q)$; otherwise they are replaced with new operators $U(q)$.

From (2.32) we see that since $U(0)$ is 1 we get $U(Q)$ by multiplying the $e^{-iG(q)\Delta q}$ in a sequence from 0 to Q. From (2.30) and (2.33) we see that $U(Q)$ commutes with all the operators A_j. Since these operators make a complete set, $U(Q)$ must be a function of them, so

$$U(Q) = e^{-i\phi(A_1,A_2,\ldots)}\ , \tag{2.35}$$

with ϕ a real function of real variables.

For each m we have

$$U(Q)|m\rangle = e^{-i\phi(a_m^{(1)},a_m^{(2)},\ldots)}|m\rangle\ . \tag{2.36}$$

The Berry phases $\gamma_m(Q)$ are defined by

$$e^{i\gamma_m(Q)} = e^{-i\phi(a_m^{(1)},a_m^{(2)},\ldots)}\ . \tag{2.37}$$

They are determined by the loop of operators $A_j(q)$, because it determines the $U(q)$.

Condition (2.34) is a generalization of our assumption for spin that each rotation from a $\hat{k}(q)$ to $\hat{k}(q+\Delta q)$ is around the $\mathbf{k}\times\Delta\mathbf{k}$ axis, as explained at the end of Sec. II A. It is essentially the same as the condition used by Anandan and Stodolsky.[7] The rationale for it in terms of dynamics is described in Sec. II E. To see the role it plays, consider what we would have without it. In the Appendix we show that if condition (2.34) were removed $G(q)$ could be changed to

$$G'(q) = G(q) - \sum_m d_m(q)U(q)|m\rangle\langle m|U(q)^{-1}\ , \tag{2.38}$$

with the d_m real functions of q. In

$$e^{-iG'(q)\Delta q}U(q)|m\rangle = e^{-iG(q)\Delta q}e^{id_m(q)\Delta q}U(q)|m\rangle \tag{2.39}$$

the phase $d_m(q)\Delta q$ would be arbitrary. It is not needed. It is not part of the change of the state represented by $U(q)|m\rangle$. If $G(q)$ satisfies (2.34), it does not produce such a phase. It does not produce a change of $U(q)|m\rangle$ proportional to $U(q)|m\rangle$. The change

$$\Delta U(q)|m\rangle = -i\Delta qG(q)U(q)|m\rangle \tag{2.40}$$

is orthogonal to $U(q)|m\rangle$. It is a change of the state. The Berry phase comes from these changes of states around the loop.

The phase $d_m(q)\Delta q$ from each interval Δq would add to the Berry phase. Consider the effect of changing $G(q)$ to $G'(q)$ in an interval Δq at a particular value of q. A new factor would be put into the sequence of unitary operators that are multiplied together to make $U(Q)$, so $U(Q)$ would be changed to

$$\begin{aligned} U'(Q) &= U(Q)U(q)^{-1} \\ &\quad\times\exp\Big(i\sum_m d_m(q)\Delta qU(q)|m\rangle\langle m|U(q)^{-1}\Big) \\ &\quad\times U(q) \\ &= U(Q)\exp\Big(i\sum_m d_m(q)\Delta q|m\rangle\langle m|\Big)\ . \end{aligned} \tag{2.41}$$

Contributions from other intervals Δq would add to this one. Altogether $U(Q)$ would be changed to

$$U'(Q) = U(Q)\exp\Big(i\sum_m\int_0^Q d_m(q)dq|m\rangle\langle m|\Big)\ . \tag{2.42}$$

The Berry phase $\gamma_m(Q)$ defined by (2.36) and (2.37) would be changed to

$$\gamma'_m(Q) = \gamma_m(Q) + \int_0^Q d_m(q)\,dq\ . \tag{2.43}$$

This is excluded by condition (2.34). It implies that the d_m are all zero. That makes $G(q)$ unique. It gives us unique unitary operators $U(q)$ and Berry phases $\gamma_m(Q)$.

D. The integral formula

To make another bridge to familiar ground we show that Berry's integral formula[1] for the Berry phase follows from our definition. We do this separately because it requires different language.

For each q we have a complete set of commuting Hermitian operators $A_1(q),A_2(q),\ldots$. For each m let $|m;q\rangle$ be a normalized eigenvector of the operators $A_1(q),A_2(q),\ldots$ for the set of eigenvalues $a_m^{(1)},a_m^{(2)},\ldots$, so

$$A_j(q)|m;q\rangle = a_m^{(j)}|m;q\rangle \,. \tag{2.44}$$

In particular, since

$$A_j(0) = A_j = A_j(Q) \,, \tag{2.30}$$

let

$$|m;0\rangle = |m\rangle = |m;Q\rangle \,. \tag{2.45}$$

For each q we have an orthonormal basis of eigenvectors $|m;q\rangle$. We assume the phases are chosen to make the vectors differentiable as functions of q.

We can see that

$$A_j(q)U(q)|m\rangle = U(q)A_j|m\rangle = a_m^{(j)}U(q)|m\rangle \,, \tag{2.46}$$

so $U(q)|m\rangle$ is an eigenvector of the operators $A_1(q),A_2(q),\ldots$ for the set of eigenvalues $a_m^{(1)},a_m^{(2)},\ldots$. Therefore

$$|m;q\rangle = e^{-i\gamma_m(q)}U(q)|m\rangle \,, \tag{2.47}$$

with γ_m a real function of q for each m. Let $\gamma_m(0)$ be 0. We have

$$|m\rangle = |m;Q\rangle = e^{-i\gamma_m(Q)}U(Q)|m\rangle \,, \tag{2.48}$$

so

$$U(Q)|m\rangle = e^{i\gamma_m(Q)}|m\rangle \,. \tag{2.49}$$

Comparing with Eqs. (2.36) and (2.37) we see that the $\gamma_m(Q)$ defined here is the same as the $\gamma_m(Q)$ defined there. It is the Berry phase.

From (2.47), (2.31), and condition (2.34) we get

$$i\left\langle m;q\left|\frac{d}{dq}\right|m;q\right\rangle = \frac{d}{dq}\gamma_m(q) \,, \tag{2.50}$$

so

$$\gamma_m(Q) = i\int_0^Q \left\langle m;q\left|\frac{d}{dq}\right|m;q\right\rangle dq \,. \tag{2.51}$$

This is Berry's integral formula for the Berry phase.[1] We know the Berry phase does not depend on the phases of the vectors $|m;q\rangle$ because we obtained the Berry phase before we introduced these vectors. The value (2.4) of the Berry phase for spin can be obtained directly[6] from this integral formula (2.51) by using (2.9) and (2.10) with (2.47).

E. Dynamics

All this has been formulated without any reference to dynamics, without considering time or a Hamiltonian, but applications that bring in dynamics are close at hand. Here we describe alternatives for dynamics that provide different applications and interpretations of the same mathematics.

Let the time coordinate t be q/g and let the Hamiltonian be

$$H(t) = gG(q) + \sum_m E_m(t)U(q)|m\rangle\langle m|U(q)^{-1} \,, \tag{2.52}$$

with g a real parameter and the E_m real functions of t. In particular, we will be interested in the case where g is 1 and the limit where g goes to 0.

If the state at time 0 is represented by

$$|\psi_m(0)\rangle = |m\rangle \,, \tag{2.53}$$

the state at time t is represented by

$$|\psi_m(t)\rangle = e^{-i\omega_m(t)}U(q)|m\rangle \,, \tag{2.54}$$

where $\omega_m(0)$ is 0 and

$$\frac{d}{dt}\omega_m(t) = E_m(t) \,. \tag{2.55}$$

This is the solution of the Schrödinger equation,

$$i\frac{d}{dt}|\psi_m(t)\rangle = H(t)|\psi_m(t)\rangle \,, \tag{2.56}$$

as one can easily see from (2.31). From (2.32) we have

$$\begin{aligned}|\psi_m(t+\Delta t)\rangle &= e^{-iE_m(t)\Delta t}e^{-iG(q)\Delta q}|\psi_m(t)\rangle \\ &= e^{-iH(t)\Delta t}|\psi_m(t)\rangle \,.\end{aligned} \tag{2.57}$$

From (2.32) and (2.33) we get

$$\begin{aligned}A_j(q+\Delta q) &= e^{-iH(t)\Delta t}A_j(q)e^{iH(t)\Delta t} \,, \\ i\frac{d}{dt}A_j(q) &= -[A_j(q),H(t)] \,.\end{aligned} \tag{2.58}$$

At time $T = Q/g$ the loop is complete. Then

$$\begin{aligned}|\psi_m(T)\rangle &= e^{-i\omega_m(T)}U(Q)|m\rangle \\ &= e^{-i\omega_m(T)}e^{-i\phi(A_1,A_2,\ldots)}|m\rangle \\ &= e^{-i\omega_m(T)}e^{-i\phi(a_m^{(1)},a_m^{(2)},\ldots)}|m\rangle \\ &= e^{-i\omega_m(T)}e^{i\gamma_m(Q)}|\psi_m(0)\rangle \,.\end{aligned} \tag{2.59}$$

The motion is cyclic for the states represented by the vectors $|m\rangle$ because they are eigenvectors of $U(Q)$. In general it is not cyclic for other states. In the Heisenberg picture the motion is cyclic for the operators $A_1,A_2,\ldots$. In general it is not cyclic for other operators.

For spin we have

$$|\psi_m(t+\Delta t)\rangle = e^{-iE_m(t)\Delta t}e^{-ik^{-2}\mathbf{k}\times\Delta\mathbf{k}\cdot\mathbf{S}}|\psi_m(t)\rangle \tag{2.60}$$

and

$$\begin{aligned}|\psi_m(T)\rangle &= e^{-i\omega_m(T)}e^{-i\phi\hat{k}(0)\cdot\mathbf{S}}|m\rangle \\ &= e^{-i\omega_m(T)}e^{-im\Omega}|\psi_m(0)\rangle \,.\end{aligned} \tag{2.61}$$

The motion is cyclic for the states represented by the vectors $|m\rangle$ because they are eigenvectors of the unitary operator for the product rotation. The motion of the spin from time 0 to time T is not entirely cyclic. The spin matrices are rotated. Only $\hat{k}(0)\cdot\mathbf{S}$ is unchanged.

Suppose g is 1 and all the E_m are zero. Then q is time and $G(q)$ is the Hamiltonian so

$$|\psi(t+\Delta t)\rangle = e^{-iG(q)\Delta t}|\psi(t)\rangle \tag{2.62}$$

and

$$|\psi(t)\rangle = U(q)|\psi(0)\rangle \tag{2.63}$$

for any state. In particular, $\omega_m(T)$ is 0 in (2.59) so the Berry phase stands out by itself. For spin

$$|\psi(t+\Delta t)\rangle = e^{-ik^{-2}\mathbf{k}\times\Delta\mathbf{k}\cdot\mathbf{S}}|\psi(t)\rangle \tag{2.64}$$

for any state; the motion is just rotation around the $\mathbf{k}\times\Delta\mathbf{k}$ axis from $\hat{k}(t)$ to $\hat{k}(t+\Delta t)$. The Berry phase stands out alone in (2.61) because $\omega_m(T)$ is zero.

Suppose terms with nonzero E_m are added to the Hamiltonian. They change the dynamics but do not change the Berry phase. These terms commute with $A_1(q), A_2(q), \ldots$ so they have no effect on the loop made by $A_1(q), A_2(q), \ldots$; they cancel out of Eqs. (2.58).

The E_m terms may make the Hamiltonian simpler. A nice example of this is discussed by Aharonov and Anandan.[8] They consider a spin and show that the Berry phase of the loop made by

$$\hat{k}(t)\cdot\mathbf{S} = S_1 \sin\theta\cos\omega t + S_2 \sin\theta\sin\omega t + S_3\cos\theta \tag{2.65}$$

is a result of the dynamics generated by the Hamiltonian

$$H = \omega S_3 . \tag{2.66}$$

To see this we only need to check that

$$i\frac{d}{dt}\hat{k}(t)\cdot\mathbf{S} = -[\hat{k}(t)\cdot\mathbf{S}, H] , \tag{2.67}$$

which is Eq. (2.58) for a spin when q is t. By itself $G(q)$ is not so simple. From (2.8) we get

$$G(q) = \omega\sin\theta(-S_1\cos\theta\cos\omega t - S_2\cos\theta\sin\omega t + S_3\sin\theta) . \tag{2.68}$$

In the limit as g goes to zero the state vectors $|\psi_m(t)\rangle$ become eigenvectors of the Hamiltonian $H(t)$ with eigenvalues $E_m(t)$. They could be realized in adiabatic evolution for a slowly changing Hamiltonian. Thus we recover the dynamics originally used to calculate Berry phases.[1,2] In the limit, when $gG(q)$ is gone, the Hamiltonian only produces the phases ω_m. It does not change the state represented by $|\psi_m(t)\rangle$; it only multiplies the state vector by a phase factor. The terms that remain in the Hamiltonian have no effect on the Berry phase; it does not depend on the E_m. The Berry phase is produced by the changes in the state described by $e^{-iG(q)\Delta q}$ and $U(q)$. Before the limit is reached, the Hamiltonian provides a model for these changes: they are made by a weak perturbation $gG(q)$ in $H(t)$ that acts over long times q/g. The Berry phase is the same whether g is large or small. The limit is needed only to get adiabatic evolution.

A rationale for our definition of Berry phases is obtained from any of our alternatives for dynamics. The definition fits the dynamics, whichever the dynamics may be. To define Berry phases we assume $G(q)$ satisfies condition (2.34). That excludes the additional d_m terms of $G'(q)$ in (2.38). These terms can be included as E_m terms in the Hamiltonian. They contribute to the dynamical phases ω_m, not the Berry phases γ_m. The definition of Berry phases is a convention that splits the Hamiltonian into two parts. Condition (2.34) is equivalent to

$$\langle m|U(q)^{-1}H(t)U(q)|m\rangle = E_m(t) . \tag{2.69}$$

It implies the E_m are mean values of the Hamiltonian. This identifies the E_m and ω_m with dynamics.[8] It distinguishes the E_m terms of the Hamiltonian from the part that generates Berry phases.

This is particularly clear for two kinds of dynamics. One is adiabatic evolution. In the limit as g goes to zero, the $E_m(t)$ become eigenvalues of $H(t)$. The Hamiltonian has only E_m terms. There is no need to split it. It does not contain the operator $G(q)$ that generates Berry phases. The ω_m are easily identified because they are the only phases produced by the Hamiltonian.

The rationale for our definition of Berry phases is also particularly clear when there are no E_m terms in the Hamiltonian. Suppose t is q. Then the Hamiltonian is just $G(q)$. Again, there is no need to split it. The Berry phases do not have to be separated from other phases. They stand out alone in (2.59), for example, in (2.61) for spin. We have an example of this kind of dynamics for spin. Let t be q and let $H(t)$ be the $G(q)$ given by (2.68). It produces the Berry phases of the loop defined by (2.65). This is a particularly clean and clear example of Berry phases produced in evolution that is not adiabatic with states that are not represented by eigenvectors of the Hamiltonian.

III. BERRY PHASES FROM LORENTZ TRANSFORMATIONS

Here our general definition of Berry phases and our method of calculating them are illustrated by obtaining Berry phases from Lorentz transformations. They are similar to those obtained from rotations. Instead of the three spin matrices that are generators for rotations, we use two generators for Lorentz transformations and one for rotations. Instead of a vector making a closed loop on the unit sphere, we consider a vector making a closed loop on a unit hyperboloid. In place of the projection of the spin in the direction of the vector, we have a linear combination of the three generators specified by the vector. This reduces to the rotation generator at the beginning and end of the loop.

We find the unitary operators required by the general definition. They represent transformations of the vector around the loop. They transform the rotation generator to a linear combination of the generators at each point on the loop and back to the rotation generator at the end of the loop. When the loop is complete, the unitary operators have not changed the rotation generator, so they have not changed a state represented by an eigenvector of the rotation generator. The state vector is changed by a Berry phase.

We calculate the Berry phases by combining infinitesimal loops the same as for rotations. Instead of a solid angle, which measures the area of a loop on the unit sphere, we get the analogous measure of the area enclosed by the loop on the unit hyperboloid. The sign is opposite what it is for rotations. All this is described in Sec. III A. The formula for the values of the Berry phases is proved in Sec. III B. The same Berry phases have been considered by other methods in various contexts.[9–14]

A. Description

Consider an irreducible unitary representation of the two-dimensional Lorentz group. The generators are Hermitian operators K_1, K_2, J_3 that satisfy the commutation relations

$$[K_2,J_3] = iK_1\,, \quad [J_3,K_1] = iK_2\,, \quad [K_1,K_2] = -iJ_3\,. \tag{3.1}$$

There is an orthonormal basis of eigenvectors $|m\rangle$ of J_3 labeled by the eigenvalues m, so

$$J_3|m\rangle = m|m\rangle\,. \tag{3.2}$$

Let $\mathbf{k}(q)$ be a real three-vector that depends on a real variable q which runs from 0 to Q with $\mathbf{k}(q)$ differentiable as a function of q. Suppose

$$-k_1(q)^2 - k_2(q)^2 + k_3(q)^2 = 1 \tag{3.3}$$

and

$$\mathbf{k}(0) = (0,0,1) = \mathbf{k}(Q)\,. \tag{3.4}$$

As q goes from 0 to Q the vector $\mathbf{k}(q)$ goes around a closed loop C on the unit hyperboloid described by (3.3).

Let

$$A(q) = k_1(q)K_1 + k_2(q)K_2 + k_3(q)J_3\,. \tag{3.5}$$

Then

$$A(0) = J_3 = A(Q)\,. \tag{3.6}$$

We write $\mathbf{k}$ for $\mathbf{k}(q)$ and $\Delta\mathbf{k}$ for $(d\,\mathbf{k}/dq)\Delta q$. Then $\mathbf{k}+\Delta\mathbf{k}$ is $\mathbf{k}(q+\Delta q)$. Let

$$\begin{aligned} G(q)\Delta q = {} & (k_2\Delta k_3 - k_3\Delta k_2)K_1 \\ & + (k_3\Delta k_1 - k_1\Delta k_3)K_2 \\ & - (k_1\Delta k_2 - k_2\Delta k_1)J_3\,. \end{aligned} \tag{3.7}$$

Then

$$A(q+\Delta q) = e^{-iG(q)\Delta q}A(q)e^{iG(q)\Delta q}\,, \tag{3.8}$$

because

$$i[A(q),G(q)\Delta q] = (\Delta k_1)K_1 + (\Delta k_2)K_2 + (\Delta k_3)J_3\,. \tag{3.9}$$

You can easily check this and check that also

$$\begin{aligned} &(1/i)[A(q),(\Delta k_1)K_1 + (\Delta k_2)K_2 + (\Delta k_3)J_3] \\ &\quad = G(q)\Delta q\,. \end{aligned} \tag{3.10}$$

Let $U(q)$ be the product of the operators $e^{-iG\Delta q}$ from 0 to q, so $U(0)$ is 1 and

$$U(q+\Delta q) = e^{-iG(q)\Delta q}U(q)\,. \tag{3.11}$$

Then $U(q)$ is unitary and

$$A(q) = U(q)A(0)U(q)^{-1}\,. \tag{3.12}$$

Since $A(0)$ is J_3 we can see that $U(q)|m\rangle$ is an eigenvector of $A(q)$ with eigenvalue m. Then from (3.10) we can see that

$$\langle m|U(q)^{-1}G(q)U(q)|m\rangle = 0\,. \tag{3.13}$$

Since both $A(0)$ and $A(Q)$ are J_3, we have

$$J_3 = U(Q)J_3U(Q)^{-1}\,. \tag{3.14}$$

This implies $U(Q)$ is a function of J_3. Since $U(Q)$ is a product of unitary operators generated by K_1, K_2, and J_3, we must have

$$U(Q) = e^{-i\phi J_3}\,, \tag{3.15}$$

with ϕ a real number. For an eigenvector $|m\rangle$ of J_3 we have

$$U(Q)|m\rangle = e^{-im\phi}|m\rangle\,. \tag{3.16}$$

The Berry phase $\gamma_m(Q)$ is defined by

$$e^{i\gamma_m(Q)} = e^{-im\phi}\,. \tag{3.17}$$

We will show that ϕ is determined by a measure of the area circumscribed by $\mathbf{k}(q)$ on the unit hyperboloid as $\mathbf{k}(q)$ moves around the loop C. Let

$$k_1 = \sinh r\cos\psi\,, \quad k_2 = \sinh r\sin\psi\,, \quad k_3 = \cosh r\,, \tag{3.18}$$

let $\Delta\mu$ be $\sinh r(\Delta r)(\Delta\psi)$ and let μ be the integral of $\Delta\mu$ over the area enclosed by C on the unit hyperboloid described by (3.3). We will show that ϕ is $-\mu$ if the area is on your left as you follow $\mathbf{k}(q)$ around the loop and ϕ is μ if the area is on your right. The sign is opposite what it is for rotations. Otherwise the only change is that the measure $d\mu$ on the unit hyperboloid replaces the measure $\sin\theta(d\theta)(d\phi)$ on the unit sphere that gives the solid-angle factor in Berry phases for rotations.

B. Calculation

The proof will be the same as it was for rotations once we calculate the contribution of an infinitesimal area of the loop. We have to consider an infinitesimal loop starting at an arbitrary vector $\mathbf{k}$ that satisfies (3.3). Since the 1 and 2 directions are related by axial symmetry, it is sufficient to consider an infinitesimal loop starting at

$$\mathbf{k}(P) = (k_1,0,k_3)\,. \tag{3.19}$$

Let the loop have four sides with

$$\begin{aligned} &\Delta\mathbf{k} = \left(\epsilon,0,\frac{k_1+\frac{1}{2}\epsilon}{k_3+\frac{1}{2}(k_1/k_3)\epsilon}\epsilon\right), \\ &\Delta\mathbf{k} = (0,\eta,(1/k_3)\tfrac{1}{2}\eta^2)\,, \\ &\Delta\mathbf{k} = \left(-\epsilon,0,-\frac{k_1+\frac{1}{2}\epsilon}{k_3+\frac{1}{2}(k_1/k_3)\epsilon}\epsilon\right), \\ &\Delta\mathbf{k} = (0,-\eta,-(1/k_3)\tfrac{1}{2}\eta^2)\,, \end{aligned} \tag{3.20}$$

where ϵ and η are small real numbers. The vectors $\mathbf{k}$ at the centers of the sides are

$$\begin{aligned} &(k_1+\tfrac{1}{2}\epsilon,0,k_3+\tfrac{1}{2}(k_1/k_3)\epsilon)\,, \\ &(k_1+\epsilon,\tfrac{1}{2}\eta,k_3+(k_1/k_3)\epsilon)\,, \\ &(k_1+\tfrac{1}{2}\epsilon,\eta,k_3+\tfrac{1}{2}(k_1/k_3)\epsilon)\,, \\ &(k_1,\tfrac{1}{2}\eta,k_3)\,. \end{aligned} \tag{3.21}$$

We calculate Δk_3 from Δk_1 and Δk_2 to maintain the hyperboloid condition (3.3). We use vectors at the centers of the sides and include the terms with a factor $\frac{1}{2}$ to be sure everything is correct to second order in ϵ and η. It is easy to see that all the terms containing $\frac{1}{2}$ cancel out. Hence we omit them. The $G(q)\Delta q$ for the four sides can be calculated from (3.7). With the terms containing $\frac{1}{2}$ omitted, they are

$$k_3\epsilon K_2 - \frac{k_1^2}{k_3}\epsilon K_2 = \frac{\epsilon}{k_3}K_2\,,$$
$$-\left(k_3 + \frac{k_1}{k_3}\epsilon\right)\eta K_1 - (k_1+\epsilon)\eta J_3\,,$$
$$-\eta\frac{k_1}{k_3}\epsilon K_1 - k_3\epsilon K_2 + \frac{k_1^2}{k_3}\epsilon K_2 - \eta\epsilon J_3 \tag{3.22}$$
$$= -\frac{k_1}{k_3}\epsilon\eta K_1 - \frac{\epsilon}{k_3}K_2 - \epsilon\eta J_3\,,$$
$$k_3\eta K_1 + k_1\eta J_3\,.$$

The product of the four $e^{-iG(q)\Delta q}$ around the loop is

$$1 + i2\epsilon\eta((k_1/k_3)K_1 + J_3) - \epsilon\eta[(1/k_3)K_2, k_3K_1 + k_1J_3]$$
$$= 1 + i2\epsilon\eta(1/k_3)(k_1K_1 + k_3J_3)$$
$$- i\epsilon\eta(1/k_3)(k_1K_1 + k_3J_3)$$
$$= 1 + i(\epsilon\eta/k_3)(k_1K_1 + k_3J_3) \tag{3.23}$$

to second order in ϵ and η. We recognize the factor

$$\frac{\epsilon\eta}{k_3} = \frac{(\Delta k_1)(\Delta k_2)}{k_3} = \frac{\cosh r(\Delta r)\sinh r(\Delta\psi)}{\cosh r}$$
$$= \sinh r(\Delta r)(\Delta\psi) = \Delta\mu\,; \tag{3.24}$$

it is the measure $\Delta\mu$ of the area of the infinitesimal loop on the unit hyperboloid. Let $\Delta\phi$ be $-\Delta\mu$. For an infinitesimal loop going around the opposite way let $\Delta\phi$ be $\Delta\mu$. Then the product of the $e^{-iG(q)\Delta q}$ around an infinitesimal loop starting at $\mathbf{k}(P)$ is

$$e^{-i(\Delta\phi)(k_1K_1 + k_3J_3)} = e^{-i(\Delta\phi)A(P)}\,, \tag{3.25}$$

where $A(P)$ is the $A(q)$ made from $\mathbf{k}(P)$.

To get $U(Q)$ and the Berry phases we have to consider a loop starting at $\mathbf{k}(0)$, which is $(0,0,1)$. This loop can be divided into infinitesimal loops the same as for rotations. All the infinitesimal loops will start at $\mathbf{k}(0)$. Some will have two parts. One part will be an infinitesimal loop starting at a vector $\mathbf{k}(P)$. That will be like the infinitesimal loop we have considered. The other part will be a path connecting $\mathbf{k}(P)$ to $\mathbf{k}(0)$. The contribution from the infinitesimal loop to $U(Q)$ will be a factor

$$e^{-i(\Delta\phi)A(0)} = e^{-i(\Delta\phi)J_3}\,. \tag{3.26}$$

The transformation from $A(P)$ to $A(0)$ is obtained the same as (2.25) for rotations. The proof is completed by multiplying these factors.

C. Comment

These Berry phases will be obtained wherever the sequence of Lorentz transformations occurs. Any state represented by an eigenvector of J_3 will be unchanged by the sequence of unitary operators $e^{-iG(q)\Delta q}$. The state vector will be changed only by a Berry phase. In particular, if J_3 is one of a complete set of commuting operators, they will be unchanged and an eigenvector of them will be changed only by a Berry phase.

IV. BERRY PHASES FOR ANY SEMISIMPLE LIE GROUP

Here we show that our general definition of Berry phases fits any unitary representation of a semisimple Lie group. We choose a complete set of commuting operators that contains a basis for a Cartan subalgebra of the Lie algebra of generators. We consider a sequence of transformations by unitary operators in the group representation that takes each operator in the Cartan subalgebra around a loop in the Lie algebra back to the same operator. We show it can be made by unitary operators in the group representation that have all the properties required for the definition of Berry phases. These unitary operators are determined by the transformations.

The loops are the basic structures that determine the Berry phases. They are made when operators from the Cartan subalgebra are transformed around the Lie algebra. Since the Lie algebra is a real vector space, the loops are made by vectors moving in a real space. The motion of the vectors is restricted by the structure of the Lie algebra.

One restriction is easy to describe: the motion of the vectors must maintain the lengths and angles defined by the Cartan metric of the Lie algebra. This is the generalization of vectors tracing loops on the surface of a sphere for rotations and on a hyperboloid for Lorentz transformations.

The Berry phases are determined by one loop made by one suitably chosen operator from the Cartan subalgebra. The loops made by other operators from the Cartan subalgebra can be found from this one. The only restriction on this loop is that its vector must always move in a direction that is orthogonal in the Cartan metric to the vectors of all the loops.

Everything can be done separately for each subalgebra of generators. Since a semisimple Lie algebra is a direct sum of simple subalgebras, it is necessary to consider only irreducible unitary representations of simple groups. The Berry phases are sums of Berry phases for the simple subgroups.

A. Definition

Consider a unitary representation of a semisimple Lie group. The generators are Hermitian operators that form a semisimple Lie algebra. Let operators $H_1, H_2, \ldots$ be a basis for a Cartan subalgebra. They commute with each other; the only operators in the Lie algebra that commute with all the H_j are linear combinations of the H_j. Let operators $G_1, G_2, \ldots$ and the $H_1, H_2, \ldots$ be a basis for the whole Lie algebra. We choose the basis so the operators

$$F_a = (1/i)[H_1, G_a] \tag{4.1}$$

also are a basis for the subspace spanned by the G_a. This can be arranged by using a Cartan–Weyl basis and choosing H_1 in a direction in the Cartan subalgebra that catches every root.[15,16]

Let $V(q)$ be unitary operators that depend on a real parameter q which runs from 0 to Q with $V(q)$ differentiable as a function of q. Let

$$i\frac{d}{dq}V(q) = K(q)V(q)\,. \tag{4.2}$$

Then $K(q)$ is Hermitian and

$$V(q + \Delta q) = e^{-iK(q)\Delta q}V(q)\,. \tag{4.3}$$

Suppose the $V(q)$ are in the group representation. Then $K(q)$ is in the Lie algebra. Suppose $V(0)$ is 1 and

$$V(Q)H_jV(Q)^{-1} = H_j \tag{4.4}$$

for each of the operators $H_1,H_2,\ldots$.

For each q let

$$\begin{aligned} H_j(q) &= V(q)H_jV(q)^{-1}, \\ G_a(q) &= V(q)G_aV(q)^{-1}, \\ F_a(q) &= V(q)F_aV(q)^{-1}. \end{aligned} \tag{4.5}$$

The $H_j(q)$ and $G_a(q)$ are a basis for the Lie algebra and the $F_a(q)$ are a basis for the subspace spanned by the $G_a(q)$. We can write

$$\begin{aligned} K(q) &= \sum_j h_j(q)H_j(q) + \sum_a g_a(q)G_a(q) \\ &= \sum_j h_j(q)H_j(q) + \sum_a f_a(q)F_a(q), \end{aligned} \tag{4.6}$$

with the h_j, g_a, and f_a real functions of q. Let

$$G(q) = \sum_a g_a(q)G_a(q) = \sum_a f_a(q)F_a(q). \tag{4.7}$$

Let $U(q)$ be the product of the $e^{-iG\Delta q}$ from 0 to q, so $U(0)$ is 1 and

$$U(q+\Delta q) = e^{-iG(q)\Delta q}U(q). \tag{4.8}$$

Then $U(q)$ is unitary and

$$i\frac{d}{dq}U(q) = G(q)U(q). \tag{4.9}$$

We have made it so

$$\begin{aligned} H_j(q+\Delta q) &= e^{-iK(q)\Delta q}H_j(q)e^{iK(q)\Delta q} \\ &= H_j(q) + i[H_j(q),K(q)]\Delta q \\ &= H_j(q) + i[H_j(q),G(q)]\Delta q \\ &= e^{-iG(q)\Delta q}H_j(q)e^{iG(q)\Delta q}, \end{aligned} \tag{4.10}$$

$$H_j(q) = U(q)H_jU(q)^{-1}. \tag{4.11}$$

In particular, from (4.4) we have

$$H_j = H_j(Q) = U(Q)H_jU(Q)^{-1}. \tag{4.12}$$

Therefore

$$U(Q) = \exp\left(-i\sum_j \phi_j H_j\right) \tag{4.13}$$

with the ϕ_j real numbers.

As q runs from 0 to Q the operator $H_j(q)$ goes around a loop in the Lie algebra from H_j back to H_j. The sequence of transformations of the H_j is made by the sequence of unitary operators $e^{-iG(q)\Delta q}$. The product of these unitary operators for the whole sequence is $U(Q)$. The product of the $e^{-iG\Delta q}$ up to q is $U(q)$.

The $U(q)$ have all the properties required for the definition of Berry phases. Suppose we have a complete set of commuting operators $A_1,A_2,\ldots$, some of which are $H_1,H_2,\ldots$. Then

$$U(Q)A_jU(Q)^{-1} = A_j \tag{4.14}$$

for each of these operators $A_1,A_2,\ldots$. Let m label the different sets of eigenvalues $a_m^{(1)},a_m^{(2)}$ of the operators $A_1,A_2,\ldots$ and for each m let $|m\rangle$ be a normalized eigenvector of the $A_1,A_2,\ldots$ for the eigenvalues $a_m^{(1)},a_m^{(2)},\ldots$, so

$$A_j|m\rangle = a_m^{(j)}|m\rangle. \tag{4.15}$$

From (4.1) we can see that

$$\begin{aligned} &\langle m|U(q)^{-1}G(q)U(q)|m\rangle \\ &= \sum_a f_a(q)\langle m|U(q)^{-1}F_a(q)U(q)|m\rangle \\ &= \sum_a f_a(q)\langle m|U(q)^{-1}V(q)F_aV(q)^{-1}U(q)|m\rangle \\ &= \sum_a f_a(q)\langle m|U(q)^{-1}V(q) \\ &\quad\times\frac{1}{i}[H_1,G_a]V(q)^{-1}U(q)|m\rangle \\ &= \sum_a f_a(q)\langle m| \\ &\quad\times\frac{1}{i}[H_1,U(q)^{-1}V(q)G_aV(q)^{-1}U(q)]|m\rangle \\ &= 0, \end{aligned} \tag{4.16}$$

for all m. The Berry phases $\gamma_m(Q)$ are defined by

$$U(Q)|m\rangle = e^{-i\gamma_m(Q)}|m\rangle. \tag{4.17}$$

Since the Lie algebra we are considering is semisimple, it is a direct sum of simple subalgebras.[15,16] Operators in different subalgebras commute with each other. Each $G(q)$ is a sum of operators from different simple subalgebras and each $U(q)$ is a product of unitary operators generated by different simple subalgebras. In particular, $U(Q)$ is, so the Berry phase $\gamma_m(Q)$ is a sum of Berry phases from unitary representations of simple groups. If each H_j were chosen to be in one of the simple subalgebras then $H_j(q)$ would be in the same subalgebra for all q. Everything could be done separately for each of the simple subalgebras. It is necessary to consider only irreducible unitary representations of simple Lie groups.

B. Determination

What determines the Berry phases? The $G(q)$ are determined by the changes

$$\begin{aligned} \Delta H_j(q) &= H_j(q+\Delta q) - H_j(q) \\ &= i[H_j(q),G(q)]\Delta q. \end{aligned} \tag{4.18}$$

If $G(q)$ and $G'(q)$ give the same changes $\Delta H_j(q)$, the difference between them must commute with the $H_j(q)$. If both $G(q)$ and $G'(q)$ are in the Lie algebra, they must be linear combinations of the $H_j(q)$ and $G_a(q)$ like $K(q)$ in (4.6). Therefore the difference between $G(q)$ and $G'(q)$ must be a linear combination of the $H_j(q)$. This is excluded by the condition (4.16); if $G(q)$ and $G'(q)$ both satisfy (4.16) they must be the same.

Thus the Berry phases are determined by the loops made by the operators $H_j(q)$ as q runs from 0 to Q. Since the Lie algebra is a real vector space, we can think of $H_j(q)$ for each j as a vector that goes around a loop in a real vector space. Since the changes $\Delta H_j(q)$ come from the commutators (4.18), the loops are restricted by the structure of the Lie algebra.

One restriction is easy to describe. As q runs from 0 to Q the motion of the vectors $H_j(q)$ must maintain the lengths and angles defined by the Cartan metric of the Lie algebra. The Killing form of the Cartan metric $(H_j(q),H_k(q))$ must not change; it is not changed by an automorphism of the Lie algebra, and the changes of the $H_j(q)$ as functions of q are made by automorphisms.[15]

The loop made by one suitably chosen operator from the Cartan subalgebra is enough to determine the Berry phases. From (4.18), (4.7), and (4.1) we have

$$\Delta H_1(q) = -\sum_a g_a(q)F_a(q)\Delta q\,. \tag{4.19}$$

We see that $G(q)$ is determined by $\Delta H_1(q)$. The $U(q)$ and therefore $U(Q)$ and the Berry phases are all determined by the loop made by $H_1(q)$ as q runs from 0 to Q. The loops made by the other $H_j(q)$ can be found from this one.

In the loop made by $H_1(q)$ each $\Delta H_1(q)$ can be any linear combination of the $F_a(q)$. The only restriction on this loop is that at each q the change in $H_1(q)$ must be in the directions of the $F_a(q)$. In the Cartan metric all the F_a are orthogonal to all the H_j and no operator in the Cartan subalgebra is orthogonal to all the H_j, so the only restriction on the loop made by $H_1(q)$ is that each $\Delta H_1(q)$ must be orthogonal to all the $H_j(q)$.

APPENDIX: EXISTENCE AND UNIQUENESS

Here we show that there are operators $U(q)$ with all the properties we assumed in Sec. II C. We also show that they are unique and that if condition (2.34) were removed the $G(q)$ could be changed to the $G'(q)$ described by (2.38).

We assume there are operators $V(q)$ with all but one of these properties; the $V(q)$ may fail to satisfy condition (2.34). Let

$$i\frac{d}{dq}V(q) = K(q)V(q)\,. \tag{A1}$$

Then $K(q)$ is Hermitian and

$$V(q+\Delta q) = e^{-iK(q)\Delta q}V(q)\,. \tag{A2}$$

From this, (2.29), (2.32), and (2.33), we get

$$e^{-iK(q)\Delta q}A_j(q)e^{iK(q)\Delta q} = A_j(q+\Delta q) = e^{-iG(q)\Delta q}A_j(q)e^{iG(q)\Delta q}\,, \tag{A3}$$

$$[A_j(q),K(q)] = [A_j(q),G(q)]\,. \tag{A4}$$

Let

$$G(q) = K(q) + D(q)\,. \tag{A5}$$

The difference $D(q)$ between $G(q)$ and $K(q)$ must commute with all the $A_j(q)$. Then $V(q)^{-1}D(q)V(q)$ must commute with all the A_j; it must be diagonal in the basis of eigenvectors $|m\rangle$ of the complete set of commuting operators A_j. This means

$$G(q) = K(q) + \sum_m d_m(q)V(q)|m\rangle\langle m|V(q)^{-1}\,, \tag{A6}$$

with the d_m real functions of q. Let

$$d_m(q) = -\langle m|V(q)^{-1}K(q)V(q)|m\rangle\,. \tag{A7}$$

This defines $G(q)$. We have

$$\langle m|V(q)^{-1}G(q)V(q)|m\rangle = 0\,. \tag{A8}$$

Let $U(0)$ be 1. Let $U(q)$ be the operator made from the $G(q)$ by following (2.32); it is the product of the $e^{-iG\Delta q}$ from 0 to q. We have made it so

$$U(q)A_jU(q)^{-1} = A_j(q)\,. \tag{A9}$$

Then

$$A_j(q)U(q)|m\rangle = U(q)A_j|m\rangle = a_m^{(j)}U(q)|m\rangle \tag{A10}$$

and

$$A_j(q)V(q)|m\rangle = V(q)A_j|m\rangle = a_m^{(j)}V(q)|m\rangle\,, \tag{A11}$$

so $U(q)|m\rangle$ and $V(q)|m\rangle$ are eigenvectors of the same complete set of commuting operators $A_j(q)$ for the same eigenvalues $a_m^{(j)}$. They can differ only by a phase factor. We can replace $V(q)|m\rangle$ with $U(q)|m\rangle$in (A6) and (A8). Then (A8) becomes condition (2.34). We can see that the $U(q)$ have all the required properties. We can also see that if condition (2.34) were removed, the $G(q)$ could be changed to operators like $K(q)$. From (A6) we see they are the $G'(q)$ in (2.38).

[1] M. V. Berry, Proc. R. Soc. London Ser. A **392**, 45 (1984); Nature **326**, 277 (1987).
[2] C. A. Mead and D. G. Truhlar, J. Chem. Phys. **70**, 2284 (1979).
[3] R. Y. Chiao and Y. S. Wu, Phys. Rev. Lett. **57**, 933 (1986).
[4] A. Tomita and R. Y. Chiao, Phys. Rev. Lett. **57**, 937 (1986).
[5] I. Bialynicki-Birula and Z. Bialynicki-Birula, Phys. Rev. D **35**, 2383 (1987).
[6] T. F. Jordan, J. Math. Phys. **28**, 1759 (1987).
[7] J. Anandan and L. Stodolsky, Phys. Rev. D **35**, 2597 (1987).
[8] Y. Aharonov and J. Anandan, Phys. Rev. Lett. **58**, 1593 (1987).
[9] M. V. Berry, J. Phys. A **18**, 15 (1985).
[10] J. H. Hannay, J. Phys. A **18**, 221 (1985).
[11] L. Vinet, Phys. Rev. D **37**, 2369 (1988).
[12] Ph. de Sousa Gerbert, "A systematic derivative expansion of the adiabatic phase," Massachusetts Institute of Technology preprint.
[13] R. Jackiw, "Three elaborations of Berry's connection, curvature and phase," in the workshop on nonintegrable phases in dynamical systems at the University of Minnesota, 1987.
[14] R. G. Littlejohn, "Phase anholonomy in the classical adiabatic motion of charged particles," Los Alamos National Laboratory preprint.
[15] A. O. Barut and R. Raczka, *Theory of Group Representations and Applications* (Polish Scientific, Warsaw, 1977).
[16] H. Bacry, *Lectures on Group Theory and Particle Theory* (Gordon and Breach, London, 1977).

Anharmonic oscillator: Constructing the strong coupling expansions

A. V. Turbiner and A. G. Ushveridze[a)]
Institute for Theoretical and Experimental Physics, Moscow 117259, USSR

(Received 20 November 1986; accepted for publication 24 February 1988)

Two novel approaches to construction of the strong coupling expansion for the anharmonic oscillator with the potential $V(x) = \frac{1}{2}x^2 + (g/4)x^4$ are proposed. The first one is simply a straightforward solution of the Schrödinger equation via the "nonlinearization" technique, resulting in the rapidly convergent perturbation series. The second one is a version of the path integral perturbation theory, but with an unconventional choice of the zeroth approximation action. Nine leading coefficients of the strong coupling expansion are computed. They decrease rapidly, the ninth one being of the order of 10^{-9}. Three leading corrections of the nonlinearization approach provide the ground-state energy within a relative accuracy of 10^{-7}–10^{-9}, at an arbitrary coupling g. The explicit formulas for corrections enable one to study the analyticity properties of the energy as a function of the coupling g. In the second approach the strong coupling expansion coefficients are computed as an infinite linear series of the weak coupling expansion coefficients. This greatly simplifies calculations, though at the price of slower and slower convergence of the series for the higher-order coefficients. If 41 terms of the weak coupling expansion are included, a relative accuracy of 10^{-6} and 10^{-2} is achieved for the second and ninth coefficients of the strong coupling expansion, respectively. In the Appendices the weak coupling expansion coefficients up to the 41st order are tabulated. In addition, both a proof of the convergence property of the nonstandard perturbation theory and also a derivation of certain functional inequalities are given, which make clear already at a qualitative level why and how the nonstandard perturbation theory acquires the convergence property.

I. INTRODUCTION

The strong coupling problem is a major one in both quantum mechanics and quantum field theory, as practically all the interesting phenomena belong to the strong coupling regime, beyond the applicability of the weak perturbation approximations. The strong coupling problem proved to be so involved that until very recently there were very few attempts to formulate its general solution.

The principal trouble with perturbation theory, which has long remained a basic tool in studying quantum mechanical problems and quantum field theory models, is that the resulting perturbation series has zero radius of convergence, which blocks extension into the domain of couplings of the order of unity. The origin of this divergence was studied in detail by Dyson for the case of QED[1]: the perturbation Lagrangian (perturbation potential) at strong fields becomes "larger" than the nonperturbed one, irrespective of how weak the coupling is. This observation is known under the generic name of "Dyson's instability argument." (For a discussion of its application to quantum mechanics see Ref. 2. The methods of computing the higher-order perturbation corrections in different quantum mechanical problems and in quantum field theory are reviewed in Ref. 3.)

It is sensible, then, to try to split the Lagrangian of interest into perturbed and unperturbed pieces in such a way as to circumvent Dyson's arguments. If so, the resulting perturbation series might be a convergent one. However, a straightforward realization of this idea requires the nonperturbed problem to be exactly soluble, and we know of few exactly soluble problems in either quantum mechanisms or quantum field theory. In this paper we shall demonstrate that this is a fictitious limitation. One could well solve the quantum mechanical problem without imposing a simpleminded exact solubility, and quantum field theory without requiring that the nonperturbed theory be of particular physical meaning.

In this paper we formulate two practical realizations of such an approach. The first one is specifically quantum mechanics oriented and leads to the convergent perturbation expansion for the Schrödinger equation. It is based on the so-called nonlinearization technique.[2] The second one will be referred to as a "nonstandard perturbation theory" (NSPT),[4] and its tool is a path integral. Here we apply both methods to the anharmonic oscillator

$$H = -\frac{1}{2}\frac{d^2}{dx^2} + \frac{1}{2}x^2 + \frac{g}{4}x^4 \tag{1}$$

in the strong coupling regime. Our purpose is a computation of coefficients of the strong coupling expansion for the ground-state energy

$$E(g) = g^{1/3}\sum_{n=0}^{\infty} b_n g^{-2n/3}. \tag{2}$$

We emphasize that despite its apparent simplicity, this problem possesses all the features of infinite-dimensional field theories, barring, perhaps, the complications with renormalization.

The paper is organized as follows. In Sec. II we review briefly the current understanding of the anharmonic oscillator problem. The subject of Sec. III is a construction of expansion (2) in the framework of the nonlinearization method. In Sec. IV we discuss NSPT, present a calculation of certain functional integrals, and derive the relationship between the coefficients of the weak coupling expansion

[a)] Permanent address: Institute of Physics, Georgian Academy of Sciences, Tbilisi, USSR.

 0022-2488/88/092053-11$02.50

$$E(g) = \sum_{n=0}^{\infty} a_n g^n \tag{3}$$

and the strong coupling expansion coefficients. In the Conclusions, Sec. V, we summarize and discuss our main results.

In Appendix A we cite 41 terms of the weak coupling expansion within absolute accuracy. The convergence property of NSPT is proved in Appendix B. In Appendix C we also derive functional inequalities, generalized Sobolev inequalities, which are useful for a qualitative analysis of the conditions for the convergence of NSPT.

II. THE ANHARMONIC OSCILLATOR PROBLEM: WHERE ARE WE?

The anharmonic oscillator problem (1) is a very old one in quantum mechanics. Its many facets have been discussed in countless publications and there is no sign of their flow running short, as on one hand the problem has numerous important applications in many fields, and on the other hand it is a perfect testing ground for any novel methods. Despite these efforts the subject is far from being exhausted.

Where are we? By means of various numerical methods a few of the lowest-level energies have been studied in the whole domain $0 \leqslant g \leqslant \infty$ (for reviews see Refs. 2, 5, and 6). For the ground-state energy 75 coefficients of the weak coupling (3) were computed[7]; the series (3) is an asymptotic one, with zero radius of convergence. The coefficients have the asymptotic behavior[8]

$$a_n = (-1)^{n+1} n!\, a^n n^{\alpha} A\,(1 + O(1/n)), \tag{4}$$

where A, a, and α are well-known parameters. The Borel summation was proved to be a correct summation of the series (3)[9] (see also Ref. 10). In an important paper by Simon[10] it was proved that the Riemann surface of $E(g)$ consists of three sheets. On the first sheet the only singularity is a cut from zero to $-\infty$. On the second and third sheets there are branch points, which converge to zero, forming the "horn" structure.[7,11] In the quasiclassical approximation these branch points are the root ones. In the limit of $g \to -0$ the discontinuity on the cut of the first sheet is of the form[8]

$$\operatorname{Im} E(g) \underset{g \to -0}{\to} \frac{1}{2^{5/2}\pi^{1/2} g^{1/2}}\, e^{-4/3g}(1 + c_1 g + c_2 g^2 + \cdots). \tag{5}$$

The coefficients of a few leading corrections $c_1, c_2, \ldots$ are known.[7]

Several approximate analytic expressions for the ground-state energy as a function of coupling have been proposed.[2,12,13] However, only in Ref. 2 were these expressions constructed in an effort to reproduce the correct analyticity properties.

Until recently the strong coupling expansion (2) computations were limited to calculation of the coefficient b_0. Hioe *et al.*[5] were apparently the first to calculate the next coefficients, b_1 and b_2, using the brute force numerical method. The numerical results for the energy as a function of g were fitted to partial sums of (2). [The functional form of expansion (2) follows from the dimensional considerations.] In Ref. 14 these coefficients were computed using the $1/N$ expansion for the $O(N)$ symmetric anharmonic oscillator (1). However, the accuracy of such an approximation could not be easily controlled, and these results are not particularly useful.

The above is practically all we know of the anharmonic oscillator (1).

III. STRONG COUPLING EXPANSION AND THE "NONLINEARIZATION" PROCEDURE

We recall briefly the nonlinearization technique. For the sake of brevity we confine ourselves to the ground-state problem.

(i) Following Price,[15] transform the Schrödinger equation into the Ricatti equation and develop the perturbation theory directly from the Ricatti equation. Such an approach has the advantage of a solution at each step of the first-order differential equation.

Specifically, in terms of the logarithmic derivative of the wave function $y = -\psi'/\psi$, the Schrödinger equation takes the form

$$y' - y^2 = 2(E - \mathscr{V}). \tag{6}$$

Let now the potential (x) be of the form

$$\mathscr{V}(x) = \mathscr{V}_0(x) + \lambda \mathscr{V}_1(x). \tag{7}$$

Here $\lambda = 1$ is a formal perturbation expansion parameter.

In terms of the perturbation expansions

$$y(x) = \sum_{n=0}^{\infty} \lambda^n y_n(x), \tag{8}$$

$$E = \sum_{n=0}^{\infty} \lambda^n E_n, \tag{9}$$

one obtains a set of equations

$$y_n' - 2y_0 y_n = 2(E_n - Q_n), \tag{10}$$

where $Q_1 = \mathscr{V}_1$, $Q_n = -\frac{1}{2}\sum_{i=1}^{n-1} y_i y_{n-i}$, and the nth order correction to the energy is given by[15,16]

$$E_n = \int_{-\infty}^{\infty} Q_n \psi_0^2\, dx \left(\int_{-\infty}^{\infty} \psi_0^2\, dx\right)^{-1}, \tag{11}$$

where $\psi_0(x)$ is the wave function of the ground state in the potential $\mathscr{V}_0(x)$ ($\mathscr{V}_0 = E_0 + \frac{1}{2}\psi_0''/\psi_0$). The solution of Eq. (10) is of the form

$$\psi_0^2 y_n(x) = 2\int_{-\infty}^{x} (E_n - Q_n)\psi_0^2(x')dx', \tag{12}$$

and satisfies the condition that the flux of particles vanishes at infinity:

$$|y_n \psi_0^2| \to 0 \quad \text{at} \quad |x| \to \infty. \tag{13}$$

The significant observation is that the perturbation theory can be formulated without examining the whole spectrum of the unperturbed problem; only the ground-state wave function of the unperturbed problem enters into all the formulas. Hence one has the freedom of choosing the unperturbed problem.

(ii) The principal criterion in making this choice must be convergence of the resulting perturbation expansion. The only convergence criterion we know of presently is Dyson's argument,[1] which suggests the following rule for constructing the convergent expansions[2]: the wave function $\psi_0(x)$ one

starts with must be such that the corresponding potential ($\mathscr{V}_0 = E_0 + \frac{1}{2}\psi_0''/\psi_0$) reproduces all the singularities and the asymptotic behavior of the original potential $\mathscr{V}(x)$.

(iii) Once the unperturbed problem is so chosen that the perturbation expansion converges, one would like to make its convergence as fast as possible. In doing so one can use the correspondence between perturbation theory and the variational principle: any variational calculation corresponds to the first two terms of a certain perturbation theory. Let us elucidate this statement. Let $\psi_0(x)$ be a certain trial function. Corresponding to this trial function is a certain Hamiltonian $H_0 = \frac{1}{2}p^2 + \mathscr{V}_0(x)$. Then the variational estimate for the energy is

$$\begin{aligned} E_{\text{var}} &= \int \psi_0 H \psi_0 \, dx \\ &= \int \psi_0 H_0 \psi_0 \, dx + \int \psi_0 (H - H_0)\psi_0 \, dx \\ &= E_0 + E_1, \end{aligned} \tag{14}$$

where E_1 is the first perturbation correction for the perturbation potential ($\mathscr{V} - \mathscr{V}_0$). If further corrections E_2, E_3, etc., are included, one could control the accuracy. On the other hand, further iterations could improve the accuracy.

After this brief summary of the nonlinearization technique (for more details see the review in Ref. 2), we are in a position to construct the zeroth-order approximation for the anharmonic oscillator (1). The main attention must obviously be paid to reproduction of the large-distance behavior of the initial potential. Hence our choice of the zeroth-order approximation is

$$y_0 = x(a^2 + \tfrac{1}{2}gx^2)^{1/2}. \tag{15}$$

Here a is the variational parameter. The function (15) corresponds to a potential ($E_0 = 0$)

$$\begin{aligned} \mathscr{V}_0(x) = \frac{a^2}{2}x^2 + \frac{g}{4}x^4 - \frac{1}{2}\left(a^2 + \frac{g}{2}x^2\right)^{1/2} \\ - \frac{1}{4}\frac{gx^2}{(a^2 + (g/2)x^2)^{1/2}}, \end{aligned} \tag{16}$$

which reproduces the initial potential at large distances and is quadratic at small distances. Hence it approximates the initial potential in the whole domain of variation of the coupling g. This choice has already been considered twice: at an intuitive level in Ref. 17, and in Ref. 2 in the course of constructing the quantum mechanical analog of the leading-log approximation of the quantum field theory. In the above papers it was shown that this approximation provides a good description of the energy as a function of g. In this paper we shall employ it to compute the coefficients b_i of the series (2).

Start with a standard variation minimization of the first correction to the energy E_1 ($E_0 = 0$ by construction):

$$\begin{aligned} E_1(a,g) &= \frac{1}{2}\int_0^\infty \left[(1-a^2)x^2 + \frac{a^2 + gx^2}{(a^2 + (g/2)x^2)^{1/2}}\right] \\ &\times \exp\left\{-\frac{4}{3g}\left[\left(a^2 + \frac{g}{2}x^2\right)^{3/2} - a^3\right]\right\} dx \\ &\times \left(\int_0^\infty \exp\left\{-\frac{4}{3g}\left[\left(a^2 + \frac{g}{2}x^2\right)^{3/2} - a^3\right]\right\} dx\right)^{-1}. \end{aligned} \tag{17}$$

The results are quoted in Table I. Already the first-order calculation gives the absolute accuracy of 10^{-6}–10^{-5} and 10^{-3}–10^{-2} in the weak and strong coupling regions, respectively. A calculation of corrections E_2 and E_3 indicates a rapid convergence. The accuracy increases to 10^{-9} and 10^{-6} at couplings 0.1 and 10^3, respectively, covering the results of Ref. 5 but disagreeing with those of Ref. 6 in the last decimal digits. The variational parameter a^2 has the following large coupling behavior:

$$a^2 = 1.003\,96\left(\frac{g}{2}\right)^{2/3} + 0.881 + \cdots. \tag{18}$$

A simple analysis of (17) shows that $E_1(g)$ has many features of the exact function $E(g)$: it has a cut from zero to $-\infty$, with the exponentially small discontinuity at $g \to -0$; also at $g = 0$ the expansion contains only integer powers of g and at $g = \infty$ it has a form (2). All these properties are preserved by the relevant formulas for E_2 and E_3.

We now make use of (18) in Eq. (17) and the corresponding counterparts for E_2 and E_3, and expand the resulting formulas in inverse powers of g. The coefficients thus found are precisely the strong coupling expansion coeffi-

TABLE I. Ground-state energy for the anharmonic oscillator (1) versus the coupling constant within the framework of the nonlinearization method. The first three corrections are taken into account. The normalizations follow the conventions of Ref. 5 ($E' = E, g' = 4g$). The results obtained are compared with those of Refs. 5 and 6.

g'	$a_{\min}$	E_1'	$-\Delta E_2'$	E_2'	$-\Delta E_3'$	E_3'	E'_{exact} [5]
0.1	1.238 0006	0.559 149 114 3	0.000 002 775 9	0.559 146 338 4	0.000 000 013 9	0.559 146 324 4	0.559 146
0.5	1.862 5	0.696 213 718	0.000 037 476	0.696 176 242	0.000 000 409	0.696 175 833	0.696 176 0.696 175 8
1.0	2.441 6	0.803 848 107	0.000 076 438	0.803 771 669	0.000 000 982	0.803 770 686	0.803 771
10.0	8.23	1.505 310 443	0.000 332 248	1.504 978 194	0.000 005 530	1.504 972 665	1.504 97
50.0	22.46	2.500 370 988	0.000 650 339	2.499 720 649	0.000 011 32	2.499 709 33	2.499 71 2.499 708 8
100.0	35.145 95	3.132 240 9	0.000 841 4	3.131 399 4	0.000 014 54	3.131 384 88	3.133 8
500.0	101.330 013	5.321 403 61	0.001 481 74	5.319 921 9	0.000 026 20	5.319 895 66	5.319 89 5.319 894 35
1 000.0	160.35	6.696 133 9	0.001 877 1	6.694 256 78	0.000 033 90	6.694 222 88	6.694 22

cients. They are quoted in Table II. One can see how the accuracy does improve with the number of corrections included. We notice that b_0 is determined to an accuracy of 10^{-8}.

Hence in the framework of the nonlinearization technique one first constructs the convergent perturbation expansion, and by reexpanding the results in inverse powers of the coupling constant one finds the strong coupling expansion coefficients.

IV. STRONG COUPLING EXPANSION AND THE NONSTANDARD PERTURBATION THEORY

In this section we shall construct the strong coupling expansion using the path integral approach. Bender *et al.*[18] have tried to do so starting with the lattice version of the path integral. Here one encounters insurmountable difficulties with the transition to the continuum limit, since for a finite lattice parameter the strong coupling expansion has a form different from (2). We know of no real calculations in the framework of this method. Here we shall rely upon another method, based on NSPT.[4] Its essence is a nontrivial splitting of the action into the perturbation and the unperturbed part in the process of computing the path integral. The result is an expression for the strong coupling expansion coefficients in terms of a series of the weak coupling expansion coefficients. The latter can be computed without any significant complications.

A. Introductory comments

The anharmonic oscillator problem is well-known to be equivalent to the one-dimensional Euclidian field theory with an action[7]

$$S[\varphi,g]=\int \dot{\varphi}\,dx+\int \varphi^2\,dx+g\int \varphi^4\,dx. \tag{19}$$

Therefore the ground-state energy can be computed in terms of the grand partition function, using the Feynman–Kac formula

$$E(g)=-\frac{1}{\int dx}\ln\int \mathscr{D}\varphi\cdot\exp\{-S[\varphi,g]\}. \tag{20}$$

We introduce the following correlation functions:

$$A_n^k=\frac{(-1)^n}{n!}\int \mathscr{D}\varphi\varphi^{2k}(0)\left(\int \varphi^4\,dx\right)^n \times\exp\left\{-\int \dot{\varphi}^2\,dx-\int \varphi^2\,dx\right\}, \tag{21}$$

$$B_n^k=\frac{(-1)^n}{n!}\int \mathscr{D}\varphi\varphi^{2k}(0)\left(\int \varphi^2\,dx\right)^n \times\exp\left\{-\int \dot{\varphi}^2\,dx-\int \varphi^4\,dx\right\}. \tag{22}$$

(All the functional integrals are supposed to be computed with dimensional regularization, so that all the vacuum loops vanish and only connected terms are left in the integrals. An important feature of dimensional regularization is invariance of the measure under arbitrary scale transformation: $\mathscr{D}[\alpha\varphi(\beta x)]=\mathscr{D}\varphi(x)$. A derivation of this property and more detailed discussion of this scheme can be found in Ref. 19.) From the definition (20) one can show [making use of the direct and inverse power expansions of (20) in g] that the correlators (21) and (22) are related to the strong and weak coupling expansion coefficients via[19]

$$a_n=[2/(1-3n)]A_n^1,\quad a_{n+1}=[1/(n+1)]A_n^2, \tag{23}$$

$$b_n=[3/(1-2n)]B_n^2,\quad b_{n+1}=[1/(n+1)]B_n^1. \tag{24}$$

It is convenient to introduce into (21) and (22) auxiliary Gaussian integrations over the $2s$-dimensional space:

$$A_n^k=\frac{(-1)^n}{n!}\int d^{2s}\rho\int \mathscr{D}\varphi\varphi^{2k}(0)\left(\int \varphi^4\,dx\right)^n \times\exp\left\{-\pi\rho^2-\int \dot{\varphi}\,dx-\int \varphi^2\,dx\right\}, \tag{25}$$

$$B_n^k=\frac{(-1)^n}{n!}\int d^{2s}\rho\int \mathscr{D}\varphi\varphi^{2k}(0)\left(\int \varphi^2\,dx\right)^n \times\exp\left\{-\pi\rho^2-\int \dot{\varphi}^2\,dx-\int \varphi^4\,dx\right\}. \tag{26}$$

The resulting generalized action is of the form

$$S[\varphi,\rho]=\pi\rho^2+\int \dot{\varphi}^2\,dx+\int \varphi^4\,dx. \tag{27}$$

To construct the NSPT let us divide the generalized action (27) into two parts:

TABLE II. The first coefficients of the strong coupling expansion $2^{2m/3-4/3}b_m$ obtained taking into account the first two corrections of the nonlinearization method. They are compared with the calculation based on NSPT, including 41 terms of the weak coupling expansion coefficients a_n, and with the previous computations of Hioe *et al.*[5] and Smondyrev–Kudinov.[14]

			Nonlinearization method[c]					
m	Hioe *et al.*[a]	Kudinov and Smondyrev[b]	First approx.	Second correct.	Second approx.	Third correct.	Final	NSPT (final)[d]
0	0.667 986 26	0.667 982	0.668 179	−1.896(−4)	0.667 989 8	−3.37(−6)	0.667 986 43	—
1	0.143 67	0.143 674	0.143 480	2.872(−4)	0.143 667 5	—	—	0.143 67
2	−0.008 8	−0.008 634	−0.008 549	−1.01(−4)	−0.008 650	—	—	−0.008 628
3	—	0.000 824	—	—	—	—	—	0.000 818

[a] See Ref. 5.
[b] See Ref. 14.
[c] See Table I.
[d] See Table IV.

$$S[\varphi,\rho] = S_0[\varphi,\rho] + \lambda S_1[\varphi,\rho], \tag{28}$$

where $\lambda = 1$ is a formal expansion parameter, and expand (26) in λ:

$$B_n^k = \sum_{m=0}^{\infty} \lambda^m G_{n,m}^k, \tag{29}$$

where

$$G_{n,m}^k = \frac{(-1)^{n+m}}{n!m!} \int d^{2s}\rho \int \mathscr{D}\varphi\varphi^{2k}(0) \times \left(\int \varphi^2\,dx\right)^n \{S_1[\varphi,\rho]\}^m e^{-S_0[\varphi,\rho]}. \tag{30}$$

Following a conventional formulation of perturbation theory, we shall refer to $S_0[\varphi,\rho]$ as an unperturbed action and to $S_1[\varphi,\rho]$ as a perturbation.

Different decompositions (28) lead to different perturbation expansions. Those of practical value are the ones for which the unperturbed problem is exactly soluble. Until recently it was common to start with actions quadratic in the fields, i.e., with free-field theory. This has usually led to divergent perturbation expansions (see the above discussion). In Refs. 4 and 19–21 it was observed that apart from the Gaussian functional integrals there do exist other exactly calculable integrals. It is not easy to find sensible field theory interpretations for the latter integrals, but that does not prevent one from using them for the formulation of a *convergent* perturbation procedure for computing the non-Gaussian path integrals. Such a procedure will be referred to as NSPT. Its application to the anharmonic oscillator for a certain special decomposition (28) was described in Ref. 19. In that paper the energy was studied as a function of g, and it was found that the convergence gradually slows down as the higher-order terms are included in the series (3). For instance, up to 30 terms must be included to reach an accuracy of 10^{-3} at $g = 10$. One of the main conclusions of this study was that the convergence property of NSPT must be improved.

In this paper we start with a decomposition different from that of Ref. 19. Specifically,

$$S_0[\varphi,\rho] = \pi\rho^2 + \int \dot{\varphi}^2\,dx + \frac{r}{C}\left\{\pi\rho^2 + \int \dot{\varphi}^2\,dx\right\}^{1/2} \times \left\{\int \varphi^2\,dx\right\}^{3/2}, \tag{31}$$

$$S_1[\varphi,\rho] = \int \varphi^4\,dx - \frac{r}{C}\left\{\pi\rho^2 + \int \dot{\varphi}^2\,dx\right\}^{1/2} \times \left\{\int \varphi^2\,dx\right\}^{3/2}, \tag{32}$$

where r is a parameter, and C is a constant defined by

$$C = 2\cdot 3^{3/2} B(2,2) = \sqrt{3}. \tag{33}$$

Here $B(x,y)$ is the Euler beta function. The origin of such a choice will be obvious from what follows.

B. Constructing the nonstandard perturbation expansions (calculation of the nonstandard functional integrals)

One has to compute different functional integrals with the weight function $\exp\{-S_0[\varphi,\rho]\}$ [see (31)]. In doing so we shall rely upon the method developed by one of the authors.[19]

Let us rewrite (30) in the following equivalent form:

$$G_{n,m}^k = \frac{(-1)^n}{n!m!} \int_0^\infty dt \int_0^\infty du\, u^n \times \exp\left\{-t - \frac{r}{C} t^{1/2}u^{3/2}\right\} \int d^{2s}\rho \int \mathscr{D}\varphi\varphi^{2k}(0) \times \left[\frac{r}{C} t^{1/2}u^{3/2} - \int \varphi^4\,dx\right]^m \times \delta\left(t - \pi\rho^2 - \int \dot{\varphi}^2\,dx\right)\delta\left(u - \int \varphi^2\,dx\right). \tag{34}$$

The interchange of the order of integrations is legal as all the integrals are finite. We now make use of invariance of the measure under the scale transformations

$$\rho^2 \to t\rho^2, \quad \varphi^2 \to t^{1/2}u^{1/2}\varphi^2, \quad x \to t^{-1/2}u^{1/2}x. \tag{35}$$

[Recall the discussion above, after (22).] Then (34) takes the form

$$G_{n,m}^k = \frac{(-1)^n}{n!m!} \int_0^\infty dt \int_0^\infty du\, t^{s+k/2-1}u^{n+k/2-1} \times \left\{\frac{r}{C} t^{1/2}u^{3/2}\right\}^m \exp\left\{-t - \frac{r}{C} t^{1/2}u^{3/2}\right\} \times \int d^{2s}\rho \int \mathscr{D}\varphi\varphi^{2k}(0)\left[1 - \frac{C}{r}\int \varphi^4\,dx\right]^m \times \delta\left(1 - \pi\rho^2 - \int \dot{\varphi}^2\,dx\right)\delta\left(1 - \int \varphi^2\,dx\right). \tag{36}$$

We now perform integrations in u and t and make use of a binomial formula for the expression within the square brackets. Finally, one obtains

$$G_{n,m}^k = \frac{(-1)^n}{n!}\Gamma\left(m + \frac{k+2n}{3}\right) \times \Gamma\left(s - m + \frac{k+2n}{3}\right)\frac{2}{3}\left(\frac{C}{r}\right)^{(k+2n)/3} \times \sum_{l=0}^{m}\left(-\frac{C}{r}\right)^l \frac{I_l}{l!(m-l)!}, \tag{37}$$

where

$$I_l = \int d^{2s}\rho \int \mathscr{D}\varphi\varphi^{2k}(0)\left\{\int \varphi^4\,dx\right\}^l \times \delta\left(1 - \pi\rho^2 - \int \dot{\varphi}^2\,dx\right)\delta\left(1 - \int \varphi^2\,dx\right). \tag{38}$$

Hence one must compute the functional integral (38).

It is convenient to employ once more the scale invariance of the measure and to make the scale transformations

$$\rho^2 \to z^{-1}\rho^2, \quad \varphi^2 \to z^{-1/2}y^{-1/2}\varphi^2, \quad x \to z^{1/2}y^{-1/2}x. \tag{39}$$

This leads to

$$I_l = z^{-s-k/2-l/2+1}y^{-k/2-3l/2+1} \times \int d^{2s}\rho \int \mathscr{D}\varphi\varphi^{2k}(0)$$

$$\times\left\{\int \varphi^4\,dx\right\}^l \delta\left(z-\pi\rho^2-\int \dot{\varphi}^2\,dx\right)$$
$$\times\delta\left(y-\int \varphi^2\,dx\right). \tag{40}$$

Evidently neither the rhs nor the lhs of (40) depends on z and y. We then act on the rhs of (40) by the unity operator

$$1=\Gamma^{-1}(s+k/2+l/2)\Gamma^{-1}(k/2+3l/2)$$
$$\times\int_0^\infty dz\, z^{s+k/2+l/2-1}e^{-z}$$
$$\times\int_0^\infty dy\, y^{k/2+3l/2-1}e^{-y}, \tag{41}$$

and perform integrations over z and y. The result is

$$I_l=\Gamma^{-1}(s+k/2+l/2)\Gamma^{-1}(k/2+3l/2)$$
$$\times\int d^{2s}\rho\int \mathscr{D}\varphi\varphi^{2k}(0)\left\{\int \varphi^4\,dx\right\}^l$$
$$\times\exp\left\{-\pi\rho^2-\int \dot{\varphi}^2\,dx-\int \varphi^2\,dx\right\}. \tag{42}$$

Recalling the definition of the correlator (25), we obtain the final form of the functional integral (38),

$$I_l=\frac{(-1)^l l!\,A_l^k}{\Gamma(s+k/2+l/2)\Gamma(k/2+3l/2)}. \tag{43}$$

Making use of (43) in (37), we obtain the final formula for the coefficients of the NSPT series:

$$G_{n,m}^k=\frac{(-1)^n}{n!}\left(\frac{C}{r}\right)^{(k+2n)/3}\frac{2}{3}$$
$$\times\sum_{l=0}^m \frac{\Gamma(m+(k+2n)/3)\Gamma(s-n+(k+2n)/3)}{\Gamma(k/2+3l/2)\Gamma(s+k/2+l/2)}$$
$$\times\left(\frac{C}{r}\right)^l. \tag{44}$$

Thus after some algebra the nontrivial functional integrals (30) with the action (31) have been reduced to conventional Gaussian integrals for the correlators A_l^k of the free-field theory. The latter, by virtue of the relationships (23), can be found in terms of the coefficients of the weak coupling expansion for the ground-state energy. We emphasize that the NSPT terms are obviously finite and that the parameter s satisfies the inequality

$$s>n-(k+2n)/3. \tag{45}$$

One of the most important features of the NSPT thus constructed is its convergence at

$$r>\tfrac{1}{2}. \tag{46}$$

A rigorous proof of this is given in Appendix C. It can be understood qualitatively as follows. In Appendix B we shall prove that for any vector ρ and for all the fields φ for which the integrands of the considered functional integrals exist, the following inequality holds:

$$\left\{\pi\rho^2+\int \dot{\varphi}^2\,dx\right\}^{1/2}\left\{\int \varphi^2\,dx\right\}^{3/2}\geqslant C\int \varphi^4\,dx \tag{47}$$

($C=\sqrt{3}$). It is a generalization of the well-known Sobolev inequality (see, for example, Ref. 22). Now we apply it to (31) and (32). Since (32) is non-negative valued, one can easily see that

$$S_0[\varphi,\rho]>|S_1[\varphi,\rho]|, \tag{48}$$

for any fields φ and vectors ρ, provided that r satisfies (46). This insures the convergence property of the NSPT (29) with the unperturbed action (31).

As the NSPT series are convergent ones, one can combine formulas (29) and (44):

$$B_m^k=\frac{(-1)^m}{m!}\left(\frac{C}{r}\right)^{(k+2m)/3}\frac{2}{3}\sum_{n=0}^\infty\sum_{l=0}^n \frac{\Gamma(n+(k+2m)/3)\Gamma(s-m+(k+2m)/3)}{\Gamma(k/2+3l/2)\Gamma(s+k/2+l/2)}\left(\frac{C}{r}\right)^l\frac{A_l^k}{(n-l)!}, \tag{49}$$

thus rewriting the correlators B_n^k in terms of A_n^k. Now we set $k=1,2$ and take into account (23) and (24). We find two sets of formulas relating the strong and weak coupling expansion coefficients:

$$b_{m+1}=\frac{(-1)^m}{m!}\left(\frac{C}{r}\right)^{(2m+1)/3}\frac{2}{3}\sum_{n=0}^\infty\sum_{l=0}^n \frac{\Gamma(n+(2m+1)/3)\Gamma(s-m+(2m+1)/3)}{\Gamma(\tfrac{1}{2}+3l/2)\Gamma(s+\tfrac{1}{2}+l/2)}\left(\frac{C}{r}\right)^l\frac{1-3l}{2(n-l)!}a_l, \tag{50}$$

and

$$b_m=\frac{(-1)^m}{m!}\left(\frac{C}{r}\right)^{(2m+2)/3}\frac{2}{1-2m}\sum_{n=0}^\infty\sum_{l=0}^n \frac{\Gamma(n+(2m+2)/3)\Gamma(s-m+(2m+2)/3)}{\Gamma(1+3l/2)\Gamma(s+1+l/2)}\left(\frac{C}{r}\right)^l\frac{l+1}{(n-l)!}a_{l+1}. \tag{51}$$

Thus we have reduced a calculation of the strong coupling expansion coefficients to calculation of the weak coupling expansion coefficients and to calculation of the partial sums (50) and (51). The convergence property of NSPT guarantees the gradual increase of the accuracy when a larger and larger number of weak coupling expansion terms are included. The technique of calculating the latter coefficients for the anharmonic oscillator (1) is developed in detail.[7] (See also a review in Ref. 23.) We have applied the method of Dalgarno's F functions (see Ref. 24), used previously also by Bender and Wu,[7] and have computed the leading 41 coefficients with absolute accuracy. (See Table III and Appendix A. The numerical calculations were performed using the symbolic integration routine REDUCE-2.) The eight leading coefficients b_m are tabulated in Table IV for the cases of 10, 20, 30, and 40 coefficients a_m included. The higher the order

of b_m, the slower the convergence; whereas 40 terms of an expansion (3) provide an accuracy of 10^{-5} in b_1, in the coefficient b_8 the accuracy is as poor as 10^{-2}. Unfortunately, improvement of the accuracy with an increase in the number of included coefficients a_m proved to be very slow.

We conclude with one important observation. Since the series (50) and (51) do converge at any positive-valued and real m, an analytic continuation into the complex m plane is possible. One finds the following elegant relationship between B_z at $z<0$ and A_z at $z>0$:

$$A_z 2/(1-e^{2\pi i z}) = B_{z^*} 3/(1-e^{2\pi i z^*}), \tag{52}$$

where

$$z^* = -\tfrac{3}{2}z + \tfrac{1}{2}.$$

From (52) it follows that a_z and b_z, as functions of a parameter z, have simple poles at $z^* = n > 0$ and at $z = n > 0$, respectively.

V. CONCLUSIONS

We have presented here practical methods for constructing the strong coupling expansion in nonrelativistic quantum mechanics. These methods are based on the construction of convergent perturbation expansions, the essence of which is a suitable decomposition of the initial Hamiltonian into the perturbation and the unperturbed components. These methods enabled us to compute fairly easily a few leading terms in the strong coupling expansion (2). Unfortunately, neither realization of this approach is free of limitations. The quantum mechanical nonlinearization technique does not allow a straightforward extension into field theory. The second formulation is a field-theoretic one, but in its present form its rate of convergence with the number of included orders of the weak coupling expansion is too low. The decomposition [(31) and (32)] has lead to a sharp acceleration of the convergence compared to that achieved in Ref. 19, but still a few dozen weak coupling expansion coefficients were needed to get an accuracy acceptable in the quan-

TABLE III. The first 41 coefficients of the weak coupling expansion a_n [$a_n = (-1)^{n+1}2^{-k}\hat{a}_n$].

n	k	$\hat{a}_n$
0	1	1
1	2	3
2	3	21
3	4	333
4	7	30885
5	8	916731
6	10	65518401
7	11	2723294673
8	15	1030495099053
9	16	54626982511455
10	18	6417007431590595
11	19	413837985580636167
12	22	116344863173284543665
13	23	8855406003085477228503
14	25	1451836748576538293163705
15	26	1275616828027135000673600 49
16	31	191385927852560927887828084605
17	32	1908061078332069804896422660 1511
18	34	4031194983593309788607032686292335
19	35	4498206045407658361605296974914586 35
20	38	2114910575845607954251483096639143447 15
21	39	2612022238376278114965497075493441703480 5
22	41	6763740228794489740596394161822714805728135
23	42	9161175277997465779846418004843284178874549 35
24	46	1036464696105720129803188216619901855131615837 85
25	47	1527737749484443898348201883095702232439261158768 3
26	49	4686466193428254260496727103508421202836493813611590 3
27	50	7468424771414634117845409123534119931849505977965315619
28	53	4939479898777642568776881638048610516053578367250019222857
29	54	8462847740745651068372378486220532150621531084756546778136 15
30	56	3001218245833010121599708970134288492619739574317885489634676 57
31	57	5501379626043601296191532858138601175291876797659873812681926308 9
32	63	3332400203443697378496750138494651760261429944928066800946897469266 21
33	64	6507605511402207156838189007579350230346280570671701771139054764279954 3
34	66	2619610274003523575267891255313014334003056555277063973821174929763212809 35
35	67	5429503623068290163468199795080936674227387982382152657766482260304499995635
36	70	4631479820569115919052453543664412502364375912603134344155219573167803001916007
37	71	1015438889090892177780991319334821408616101579513325223022119873844127738685843305
38	73	4574341811852406165712981973852745195703097531162983195516691754666602933971260000219
39	74	1057736729256106720773024864888870968827508829383628975105602001157374951934221014134 35
40	78	2007380152042064387738075843917068054599107695331873493476644213557359137472667294778120 03
41	79	4882341371768649695721506193679714428703711845960546204436625547776448084599704119000083710 5

TABLE IV. The coefficients of the strong coupling expansion calculated according to NSPT versus the number of terms N included in the analysis.

m		$N=10$	$N=20$	$N=30$	$N=40$
1	(10^{-1})	1.435 48	1 436 73	1 436 72	1.436 68
2	(-10^{-2})	8.680 0	8.630 9	8.627 4	8.627 5
3	(10^{-4})	8.212	8.180	8.181	8.182
4	(-10^{-5})	8.083	8.227	8.244	8.243
5	(10^{-6})	7.774	8.072	8.075	8.068
6	(-10^{-7})	7.446	7.367	7.276	7.279
7	(10^{-8})	7.076	5.733	5.577	5.622
8	(-10^{-9})	5.64	2.713	2.918	2.964

tum mechanical problems. No one needs such accuracy in quantum field theory; nevertheless, for credible results at least a dozen weak coupling expansion coefficients are needed. It is very unlikely that we will have such a number of coefficients in our possession in the near future. One more drawback of the second method is the lack of a counterpart to the Feynman–Kac formula for the excited states, unnecessary though it might seem for a field theory.

The aforementioned methods could easily be generalized to cases of more complicated anharmonic oscillators, for instance,

$$H=-\frac{1}{2}\frac{d^2}{dx^2}+\frac{1}{2}x^2+\frac{g}{2^p}x^{2p}. \tag{53}$$

In the first approach one should start with

$$y_0=x\left(\sum_{i=0}^{p-1}c_i x^{2i}\right)^{1/2}, \tag{54}$$

where $c_{p-1}=g/2^p$, whereas in the second approach a generalization of (31) and (32) is

$$S_0[\varphi,\rho]=\pi\rho^2+\int\dot{\varphi}^2\,dx+\frac{r}{C}\left\{\pi\rho^2+\int\dot{\varphi}^2\,dx\right\}^{(p-1)/2}\times\left\{\int\varphi^2\,dx\right\}^{(p+1)/2}, \tag{55}$$

$$S_1[\varphi,\rho]=\int\varphi^{2p}\,dx-\frac{r}{C}\left\{\pi\rho^2+\int\dot{\varphi}^2\,dx\right\}^{(p-1)/2}\times\left\{\int\varphi^2\,dx\right\}^{(p+1)/2}, \tag{56}$$

where

$$C=2\frac{(p+1)^{(p+1)/2}}{(p-1)^{(p-1)/2}}\left\{B\left(\frac{p}{p-1},\frac{p}{p-1}\right)\right\}^{p-1}. \tag{57}$$

The relevant generalization of the Sobolev inequality reads

$$\left\{\pi\rho^2+\int\dot{\varphi}^2\,dx\right\}^{(p-1)/2}\left\{\int\varphi^2\,dx\right\}^{(p+1)/2}\geqslant C\int\varphi^{2p}\,dx. \tag{58}$$

It implies that at $r>\frac{1}{2}$ the perturbation is weaker than the unperturbed component, which ensures the convergence property of NSPT [compare with (46)]. We notice that the NSPT terms are finite provided that

$$s>m-(k+2m)/(p+1) \tag{59}$$

[cf. (45)].

We conclude by citing the counterpart of (51) for a relationship between the strong and weak coupling expansion coefficients:

$$b_m=\frac{(-1)^m}{m!}\left(\frac{C}{r}\right)^{(p+2m)/(p+1)}\frac{2}{1-2m}\sum_{n=0}^{\infty}\sum_{l=0}^{n}\frac{\Gamma(n+(p+2m)/(p+1))\Gamma(s-m+(p+2m)/(p+1))}{\Gamma(p/2+(p+1)l/2)\Gamma(s+p/2+(p-1)l/2)}\times\left(\frac{C}{r}\right)^{l}\frac{(l+1)}{(n-l)!}. \tag{60}$$

A few concluding remarks on future prospects: We place our hopes mainly on the second approach. On the one hand, an interesting development would be an analysis of many-dimensional quantum mechanical problems. On the other hand, one could consider many-dimensional field theory models, extending further the results of Refs. 20 and 21. A major problem here is a search for the most suitable decompositions of the action and for novel methods of accelerating the convergence while retaining the linear relationship between the strong and weak coupling expansion coefficients.

APPENDIX A: THE CALCULATION OF WEAK COUPLING EXPANSION COEFFICIENTS

To find the coefficients of the weak coupling expansion we use Dalgarno's F-function method. (See, for example, Ref. 24. In fact, this method was used by Bender and Wu[7] in an analogous calculation.) Let us write the wave function in the form

$$\Psi(x)=e^{-x^2/2}\sum_{n=0}^{\infty}g^n\hat{a}_n(x). \tag{A1}$$

Substituting (A1) and (3) in the Schrödinger equation with Hamiltonian (1) and collecting terms of order g^n, we obtain the equation for the nth correction $\hat{a}_n(x)$. As a consequence of the boundary condition the solution of this equation will be a polynomial in x:

$$\hat{a}_n(x)=(-1)^n\sum_{i=1}^{2n}\hat{a}_{n,i}x^{2i},\quad \hat{a}_0(x)=1. \tag{A2}$$

After the substitution $x^2\to 2y^2$ in (A2) the recursion relations are obtained:

$$2i\hat{a}_{n,i}=(i+1)(2i+1)\hat{a}_{n,i+1}+\hat{a}_{n-1,i-2}+\sum a_{n-j}\hat{a}_{j,i}, \tag{A3}$$

where $a_{n-j} = -\hat{a}_{n-j,1}$.

The coefficients a_n have the form of simple fractions, which were found by means of the symbolic routine REDUCE-2. We have calculated the first 41 coefficients a_n with absolute accuracy. Such an accurate calculation was necessary due to the accumulation of errors in computing the coefficients b_n (see p. 2058).

APPENDIX B: GENERALIZED SOBOLEV INEQUALITIES

The proof of the generalized Sobolev inequalities (47) and (58) is based on the well-known Sobolev embedding theorems.[22] Their essence is as follows.

Let a field φ, defined on the real axis R, belong to the Sobolev space $W_2^1(R)$, i.e., let it be quadratically integrable on R and have a quadratically integrable first derivative. Then φ necessarily belongs to the Lebesgue space L_{2p}, i.e., it is integrable in R with the power exponent $2p$, where p is an arbitrary number exceeding unity. (For a more rigorous definition of the Sobolev and Lebesgue spaces see Ref. 25.) Then for any $\varphi \in W_2^1(R)$, the following inequality (the so-called Sobolev inequality) holds:

$$\left\{\int \dot{\varphi}^2\,dx + \int \varphi^2\,dx\right\}^p \geqslant C^* \int \varphi^{2p}\,dx, \tag{B1}$$

where

$$C^* = 2^{p+1}\frac{p^p}{(p-1)^{p-1}}\left\{B\left(\frac{p}{p-1},\frac{p}{p-1}\right)\right\}^{p-1}. \tag{B2}$$

Now we are in a position to prove (47) and (58). Let W be the set of all nonvanishing fields $\varphi \in W_2^1(R)$. Then the following theorem holds.

Theorem: The functional

$$C[\varphi] = \left(\int \dot{\varphi}^2\,dx\right)^{(p-1)/2}\left(\int \varphi^2\,dx\right)^{(p+1)/2}\left(\int \varphi^{2p}\,dx\right)^{-1} \tag{B3}$$

defined on W reaches on W its exact lower bound, which equals

$$C = \left(\frac{p-1}{2p}\right)^{(p-1)/2}\left(\frac{p+1}{2}\right)^{(p+1)/2} C^*. \tag{B4}$$

Proof: Consider the scaling transformations of fields and coordinates, which leave the fields in the set W:

$$\varphi \to \alpha\varphi, \quad x \to \beta x, \quad \alpha \neq 0, \quad \beta \neq 0. \tag{B5}$$

It is obvious that for any $\varphi(x)$ there do exist such α and β, and that after the transformation (B5) the integrals $\int \dot{\varphi}^2\,dx$ and $\int \varphi^2\,dx$ take any given values C_1 and C_2. On the other hand, the transformations (B5) do not change the value of the functional $C[\varphi]$. Therefore one can write

$$\inf_W C[\varphi] = \inf_{W(C_1,C_2)} C[\varphi] = C_1^{(p-1)/2} C_2^{(p+1)/2} \inf_{W(C_1,C_2)} \left(\int \varphi^{2p}\,dx\right)^{-1}. \tag{B6}$$

Here $W(C_1,C_2)$ is a set of fields $\varphi \in W$, subject to two conditions,

$$\int \dot{\varphi}^2\,dx = C_1, \quad \int \varphi^2\,dx = C_2. \tag{B7}$$

Now, by virtue of the Sobolev inequality (B1), one finds

$$\inf_W C[\varphi] \geqslant C_1^{(p-1)/2} C_2^{(p+1)/2} \frac{C^*}{(C_1+C_2)^p} = \xi^{(p-1)/2}(1-\xi)^{(p+1)/2} C^*. \tag{B8}$$

Here we have introduced

$$\xi = C_1/(C_1+C_2). \tag{B9}$$

Since the inequality (B8) holds for any value of ξ, it is most interesting at the value of ξ at which its rhs takes its maximal value. This extremal value can easily be found:

$$\inf_W C[\varphi] \geqslant \left(\frac{p-1}{2p}\right)^{(p-1)/2}\left(\frac{p+1}{2p}\right)^{(p+1)/2} C^*. \tag{B10}$$

Consider now the specific case of

$$\varphi^* = (1/\cosh x)^{1/(p-1)}. \tag{B11}$$

A straightforward calculation of (B3) for the field (B11) yields

$$C[\varphi^*] = \left(\frac{p-1}{2p}\right)^{(p-1)/2}\left(\frac{p+1}{2p}\right)^{(p+1)/2} C^*. \tag{B12}$$

A comparison of (B10) and (B12) shows that the factor C [see (B4)] is an absolute minimal value of the functional $C[\varphi]$, which, by virtue of the invariance of $C[\varphi]$ under translations and the scaling transformations (B15), is attained for fields of the form

$$\varphi = \alpha/[\cosh(\beta x+\gamma)]^{1/(p-1)}. \tag{B13}$$

The theorem is proved.

The corollary of this theorem is that the inequality of interest to us,

$$\left(\pi\rho^2 + \int \dot{\varphi}^2\,dx\right)^{(p-1)/2}\left(\int \varphi^2\,dx\right)^{(p+1)/2} \geqslant C\int \varphi^{2p}\,dx, \tag{B14}$$

holds for any $p \geqslant 2$, $\rho \in R^{2s}$, and $\varphi \in W_2^1(R)$.

APPENDIX C: CALCULATION OF THE ASYMPTOTIC BEHAVIOR IN NSPT

Let us prove the convergence property of the NSPT series at

$$r > \tfrac{1}{2}. \tag{C1}$$

(Qualitative arguments suggested by the generalized Sobolev inequality were presented in Sec. IV.) For a more rigorous proof of the convergence property let us calculate the large-order terms of the NSPT series by the method formulated in Refs. 19 and 26. This method enables us to express the asymptotics of $G^k_{n,m}$ in terms of the asymptotics of the correlators A^k_m and B^k_m in the limit of $m \to \infty$.

The formula (4) for the asymptotics of the coefficients a_m and the relationship (23) imply that the correlators A^k_m have the asymptotic behavior

$$A^k_m = (-1)^m m!\, a^m n^{\alpha(k)} A(k)(1+O(1/m)), \tag{C2}$$

where

$$\alpha(1)=\alpha+1,\quad \alpha(2)=\alpha+2,\quad A(1)=\tfrac{3}{2}A,$$
$$A(2)=aA.$$

According to Simon's result,[10] the strong coupling expansion (2) has a finite radius of convergence. Besides, Table IV suggests that, starting with the first coefficient, the b_n's have alternating signs. Hence a natural conjecture for the asymptotic behavior of b_n is

$$b_m=(-1)^{m+1}b^m m^\beta B(1+O(1/m)) \tag{C3}$$

[cf. (4)], where b, β, and B are certain constants whose exact values are irrelevant to our purposes. In principle, once a sufficiently large number of the b_m are known, the rest could easily be determined. Assuming them to be known, and making use of (24), one easily finds the asymptotics of the correlators B^k_m:

$$B^k_m=(-1)^m b^m m^{\beta(k)}B(k)(1+O(1/m)), \tag{C4}$$

where

$$\beta(1)=\beta+1,\quad \beta(2)=\beta+1,$$
$$B(1)=b\cdot B,\quad B(2)=\tfrac{2}{3}B.$$

Now we shall proceed further by combining (C2) and (C4) with their representation (36). In the latter we perform integrations in t and u. This leads to

$$G^k_{n,m}=\frac{(-1)^n}{n!}\left(\frac{C}{r}\right)^{(k+2n)/3}\Gamma\left(s-n+\frac{k+2n}{3}\right)\times\frac{2}{3}\frac{\Gamma(m+(k+2n)/3)}{(m+1)}\times\left\{\int_0^r d\nu\left(1-\frac{\nu}{r}\right)^m f(\nu)+(-1)^m\times\int_r^1 d\nu\left(\frac{\nu}{r}-1\right)^m f(\nu)\right\}, \tag{C5}$$

where

$$f(\nu)=\int d^{2s}\rho\int\mathscr{D}\varphi\varphi^{2k}(0)\delta\left(\nu-C\int\varphi^4\,dx\right)\times\delta\left(1-\pi\rho^2-\int\dot\varphi^2\,dx\right)\delta\left(1-\int\varphi^2\,dx\right). \tag{C6}$$

We do not know yet the explicit form of the function $f(\nu)$, but that causes no trouble. Indeed, from (C5) it follows that the asymptotic behavior of terms in $G^k_{n,m}$ in the limit of $m\to\infty$ is controlled solely by the asymptotic behavior of $f(\nu)$ as $\nu\to1$ and $\nu\to0$. An important observation is that the latter is uniquely fixed by the asymptotic behavior (C2) and (C4) of the correlators A^k_m and B^k_m. Indeed, we can write A^k_m and B^k_m in terms of $f(\nu)$. To do so we start with definitions (21) and (22), and representations of A^k_m and B^k_m of the form

$$A^k_m=\frac{(-1)^m}{m!}\int_0^\infty dt\int_0^\infty du\,e^{-t-u}\times\int_0^{t^{1/2}u^{3/2}}dz\left(\frac{z}{C}\right)^m\int d^{2s}\rho\int\mathscr{D}\varphi\varphi^{2k}(0)\times\delta\left(z-C\int\varphi^4\,dx\right)\delta\left(t-\pi\rho^2-\int\dot\varphi^2\,dx\right)\times\delta\left(u-\int\varphi^2\,dx\right), \tag{C7}$$

$$B^k_m=\frac{(-1)^m}{m!}\int_0^\infty dt\int_0^\infty du\,e^{-t}u^m\times\int_0^{t^{1/2}u^{3/2}}dz\,e^{-z/C}\int d^{2s}\rho\int\mathscr{D}\varphi\varphi^{2k}(0)\times\delta\left(z-C\int\varphi^4\,dx\right)\delta\left(t-\pi\rho^2-\int\dot\varphi^2\,dx\right)\times\delta\left(u-\int\varphi^2\,dx\right). \tag{C8}$$

After the scaling transformations (35) and the introduction of a new variable

$$\nu=zt^{-1/2}u^{3/2}, \tag{C9}$$

the formulas (C7) and (C8) take the form

$$A^k_m=\frac{(-1)^m}{m!}\int_0^\infty dt\,e^{-t}t^{s+(k+m)/2-1}\times\int_0^\infty du\,e^{-u}u^{(k+3m)/2-1}C^{-m}\int_0^1 d\nu\,\nu^m f(\nu), \tag{C10}$$

$$B^k_m=\frac{(-1)^m}{m!}\int_0^\infty dt\,e^{-1}t^{s+k/2-1}\times\int_0^\infty du\,e^{-t^{1/2}u^{3/2}}u^{m+k/2-1}\times C^{(k+2m)/3}\int_0^1 d\nu\,\nu^{-(k+2m)/3}f(\nu). \tag{C11}$$

Now one can perform integrations in t and u. The result is

$$A^k_m=\frac{(-1)^m}{m!}\Gamma\left(s+\frac{k+m}{2}\right)\Gamma\left(\frac{k+3m}{2}\right)C^{-m}\times\int_0^1 d\nu\,\nu^m f(\nu), \tag{C12}$$

$$B^k_m=\frac{(-1)^m}{m!}\Gamma\left(\frac{2m+k}{3}\right)\Gamma\left(\frac{k-m}{3}+s\right)C^{(k+2m)/3}\times\frac{2}{3}\int_0^1 d\nu\,\nu^{-(k+2m)/3}f(\nu). \tag{C13}$$

From (C12) and (C13) one observes that the asymptotic behavior of A^k_m and B^k_m is closely related to the asymptotic behavior of the function $f(\nu)$ in the limits $\nu\to1$ and $\nu\to0$, respectively, as the relevant integrands do not vanish only in these limits as $m\to\infty$. Now take the limit $m\to\infty$ in (C12) and (C13), using the Stirling formula for the gamma functions and the asymptotic formulas (C4) and (C2). Then

$$\int_0^1 d\nu\,\nu^m f(\nu)=m^{-s-k+2+\alpha(k)}2^{k-1+s}A(k), \tag{C14}$$

$$\int_0^1 dv\, v^{-(k+2m)/3} f(v)$$

$$= \Gamma\left(\frac{2m}{3}\right)[3\cdot 2^{-4/3} C^{-2/3} b\,]^m m^{-s-2k/3+2+\beta(k)} \times \sin\left(\pi\frac{m-k}{3}\right)\cdot(-1)^{s+1} \times \frac{1}{\pi} 3^{s+2k/3-1/2}(2s)^{-k/3} B(k). \quad \text{(C15)}$$

These formulas are convenient for the determination of the asymptotics $f(v)$ as $v\to 1$ and $v\to 0$. One can easily check explicitly that (C14) and (C15) hold provided that, in the limit $v\to 1$,

$$f(v) \sim \sigma(1-v)^{\tau}, \quad \text{(C16)}$$

and that, in the limit $v\to 0$,

$$f(v) = \mathrm{Re}\{\exp\{i(\omega + i0)/v - i\delta\} v^{\gamma} \varkappa\}. \quad \text{(C17)}$$

Here σ, τ, ω, δ, γ, and $\varkappa$ are the real constants, given by

$$\tau = k + s - \tfrac{1}{2} - \alpha(k), \quad \text{(C18a)}$$

$$\sigma = (\tfrac{2}{3})^{k+s-1} A(k)/\Gamma(k+s-\alpha(k)), \quad \text{(C18b)}$$

$$\omega = \tfrac{4}{3} b^{-3/2}, \quad \text{(C19a)}$$

$$\delta = (\pi/2)\{\beta(k) + s + 1\}, \quad \text{(C19b)}$$

$$\gamma = k + s - \beta(k) + 3, \quad \text{(C19c)}$$

$$\varkappa = \left\{\frac{4}{3} b^{-3/2}\right\}^{\beta(k)+2-s-k} \left(\frac{2}{\sqrt{b}}\right)^k \times \frac{1}{\pi} \frac{3^{s+1/2+k/6}}{2^{2k/3+1}} B(k). \quad \text{(C19d)}$$

Finally, in order to determine the asymptotics of the NSPT series terms in the limit $n\to\infty$, one must determine the corresponding limiting behavior of (C5), with an explicit use of the formulas (C16)–(C19). After simple but tedious algebraic manipulations, one finds

$$G^k_{n,m} = F^k_{n,m} + (-1)^m H^k_{n,m}, \quad m\to\infty, \quad \text{(C20)}$$

where

$$F^k_{n,m} = e^{-f_1\sqrt{m}} (\sqrt{m})^{f_2} \cos(f_3\sqrt{m} + f_4) f_5, \quad \text{(C21)}$$

$$H^k_{n,m} = h_1^m m^{h_2} h_3. \quad \text{(C22)}$$

The constants in (C21) and (C22) are defined as follows:

$$f_1 = \sqrt{2}\,(\omega/r)^{1/2}, \quad \text{(C23a)}$$

$$f_2 = \tfrac{2}{3}(k+2n) - \gamma - \tfrac{7}{2}, \quad \text{(C23b)}$$

$$f_3 = \sqrt{2}(\omega/r)^{1/2}, \quad \text{(C23c)}$$

$$f_4 = -(\pi/4)(\gamma + \tfrac{1}{2}) - \delta - \omega/2r, \quad \text{(C23d)}$$

$$f_5 = \varkappa r^{\gamma+1}\sqrt{\pi}\frac{2}{3}\left(\frac{\omega}{r}\right)^{(\gamma+1/2)/2} \times \frac{(-1)^n}{n!}\left(\frac{C}{r}\right)^{(k+2n)/3} \Gamma\left(s - n + \frac{k+2n}{3}\right), \quad \text{(C23e)}$$

$$h_1 = (1-r)/r, \quad \text{(C24a)}$$

$$h_2 = (k+2n)/3 - \tau - 2, \quad \text{(C24b)}$$

$$h_3 = \sigma(1-r)^{\tau+1}\Gamma(\tau+1)\frac{(-1)^n}{n!}\left(\frac{C}{r}\right)^{(k+2n)/3} \times \Gamma(s - n + (k+2n)/3)\cdot\frac{2}{3}. \quad \text{(C24c)}$$

The set of formulas (C20)–(C22), complemented by the formulas (C23), (C24), (C17), and (C19), gives a complete description of the asymptotics of the large-order terms of the NSPT series. To complete our analysis, consider now how the asymptotic behavior depends on the choice of the parameters r and s. It can easily be seen that at $r<\frac{1}{2}$, the second term dominates in (C20). In this case NSPT series has the alternating signs and diverges. Its radius of convergence is less than 1 and vanishes in the limit $r\to 0$; this is obvious, as in this limit NSPT coincides with the conventional perturbation expansion in the coupling constant. In the opposite case, where $r>\frac{1}{2}$, it is the first term that dominates in (C20). In this case the NSPT series is convergent. The terms of this series oscillate around zero with decreasing amplitude and rising oscillation period. From the formulas we have derived one can see that the larger r is, the lower the rate of convergence is, whereas at larger s the convergence accelerates. However, the larger s is, the later the asymptotic behavior starts. A more complete analysis of the onset of asymptotic behavior will be published elsewhere.

[1] F. Dyson, Phys. Rev. **85**, 631 (1952).
[2] A. V. Turbiner, Usp. Fiz. Nauk **144**, 35 (1984).
[3] E. B. Bogomolny, V. A. Fateev, and L. N. Lipatov, Sov. Sci. Rev. **2**, 247 (1980).
[4] B. S. Shaverdyan and A. G. Ushveridze, Phys. Lett. B **123**, 316 (1983).
[5] F. T. Hioe, D. McMillen, and E. W. Montroll, Phys. Rep. C **43**, 306 (1978).
[6] K. Banerjee, S. P. Bhatnagar, V. Choudhury, and S. S. Konwal, Proc. R. Soc. London Ser. A **360**, 575 (1978).
[7] C. M. Bender and T. T. Wu, Phys. Rev. **184**, 1231 (1969).
[8] A. I. Wainstein, *Divergence Perturbation Theory of the Anharmonic Oscillator* (Institute of Nuclear Physics Siberian Division, USSR Academy of Sciences, Novosibirsk, 1964), in Russian.
[9] S. Graffi, V. Grecchi, and B. Simon, Phys. Lett. B **32**, 631 (1970).
[10] B. Simon, Ann. Phys. **58**, 79 (1970).
[11] W. Y. Crutchfield II, Phys. Lett. B **77**, 109 (1978).
[12] K. Banerjee and J. K. Bhattacharjee, Phys. Rev. D **19**, 1111 (1984).
[13] F. Fernandez, G. A. Arteca, and E. A. Castro, Physica A **122**, 37 (1983).
[14] A. V. Kudinov and M. A. Smondirev, JINR preprint E2-83-412.
[15] P. H. Price, Proc. Phys. Soc. London **67**, 383 (1954).
[16] Ya. B. Zel'dovich, **31**, 1101 (1956).
[17] A. D. Dolgov and V. S. Popov **75**, 2010 (1978).
[18] C. M. Bender, F. Cooper, G. S. Guralnik, and D. H. Sharp, Phys. Rev. D **19**, 1865 (1979).
[19] A. G. Ushveridze and N. I. Shubitidze, Yad. Phiz. **40**, 1065 (1984).
[20] A. G. Ushveridze, Yad. Phiz. **38**, 798 (1983).
[21] A. G. Ushveridze, Phys. Lett. B **142**, 403 (1984).
[22] C. M. Nikol'skii, *Priblizenie funkzii mnogikh peremennikh i teoremi vlozenija* (Nauka, Moscow, 1977).
[23] J. Zinn-Justin, Phys. Rep. C **70**, 2 (1981).
[24] J. O. Hirschfelder, W. B. Brown, and S. T. Epstein, Adv. Quantum Chem. **1**, 225 (1964).
[25] V. M. Trenogin, *Funkzional'nii analiz* (Nauka, Moscow, 1976).
[26] A. G. Ushveridze, Soobshch. Akad. Nauk Gruz. USSR **115**, 261 (1984).

Geometrical and kinematical characterization of parallax-free world models

W. Hasse and V. Perlick
Institut für Theoretische Physik der TU Berlin, Sekr. PN 7-1, Hardenbergstrasse 36, 1000 Berlin 12, Federal Republic of Germany

(Received 21 August 1987; accepted for publication 4 May 1988)

An arbitrary general relativistic world model, i.e., a pseudo-Riemannian manifold along with a timelike vector field V, is considered. Such a kinematical world model is called "parallax-free" iff the angle under which any two observers (i.e., integral curves of V) are seen by any third observer remains constant in the course of time. It is shown that a model is parallax-free iff V is proportional to some conformal Killing field. In this case V, especially, has to be shear-free. Furthermore a relationship between parallaxes and red shift is presented and a reference is made to considerations concerning the visibility of cosmic rotation.

I. INTRODUCTION

In 1985, Treder[1] raised the interesting question of whether cosmic rotation would be observable as an aberrational motion of distant objects. He presented an aberrational formula [formula (9) in Ref. 1], which was claimed to hold approximately in the case of Gödel's rotating cosmos. According to this formula there is an influence of the rotation on aberrational angles, and this influence is proportional to the distance between observer and observed object. Recently Ruben[2] investigated the possible verification of this effect from observational data on quasar aberrations.

Stimulated by these articles we became interested in the following related question: Consider an arbitrary general relativistic world model, i.e., a pseudo-Riemannian manifold (M,g) equipped with a timelike vector field V representing the mean flow of galaxies. What can be learned about the kinematical properties of this model (such as rotation, shear, etc.) from the observation of "parallaxes"? Here we use the word "parallax" in its most general sense characterizing the general relativistic amalgamation of "aberration," "proper motion," and "parallax." We are going to make this concept mathematically precise in Sec. II. In Sec. III, we present a theorem characterizing parallax-free world models by geometrical and kinematical properties. In Sec. IV, we discuss our theorem and its relation to Treder's work. The proof of our theorem is given in the Appendix.

II. DEFINITION OF PARALLAX-FREE WORLD MODELS

First of all we want to remind the reader that in general relativity there is some ambiguity concerning the notions "aberration," "proper motion," and "parallax," since in an arbitrary general relativistic world model these three effects cannot be separated from each other in an invariant manner (see, e.g., Mast and Strathdee,[3] p. 484). Furthermore there is some discrepancy in the use of the word "aberration," which remains even in a special relativistic context: Whereas the common aberrational formulas (given in standard textbooks such as Rindler,[4] p. 57) are completely independent of the motion of the source, Treder[1] uses the word "aberration" for an effect that depends on the motion of the source during the travel time of light. A point of view similar to Treder's can be found in Laue,[5] p. 12.

Now let us make precise what we will mean by "parallax" in this paper. We consider a general relativistic world model, which in our notation is a triple (M,g,V), where (i) M denotes the space-time manifold, which is assumed to be simply connected for simplicity; (ii) g denotes the metric tensor field of signature $(+\,+\,+\,-)$; and (iii) V denotes a timelike vector field normalized to $g(V,V) = -c^2$, where c is the velocity of light. The integral curves of V are interpreted as the world lines of observers representing the mean flow of galaxies (or groups of galaxies). Now let us fix three observers (i.e., three integral curves of V) a_0, a_1, and a_2. Let us assume that a_1 and a_2 are emitting light rays, which are received by a_0. Then a_0 will see a_1 and a_2 as light points at his celestial sphere. In the general case these two light points will move relative to each other, i.e., the angle between a_1 and a_2 will not remain constant over time. This effect we call "parallax." In fact, this is probably the most general meaning that could be given to this word. Now it has already become clear what we will mean by "parallax-free world models."

Definition 1: A world model (M,g,V) is called "parallax-free in the strong sense" iff, for any three observers a_0, a_1, a_2 in V, the angle under which a_1 and a_2 are seen by a_0 remains constant over time.

Figuratively speaking this means the following: If we are traveling on the world line of an observer in V, then all the other observers form a pattern on our celestial sphere that does not change over time. Note that the pattern is allowed to rotate rigidly with respect to an inertial compass.

In the general case, parallaxes depend on distance in the following sense: If at one instant of time a_0 sees a_1 and a_2 in the same spatial direction (one behind the other), then in the general case a_1 and a_2 will not follow the same path at the celestial sphere of a_0 over time. We want to give a special name to the case where parallaxes in the sense of Definition 1 are admitted, but do not depend on distance.

Definition 2: A world model (M,g,V) is called "parallax-free in the weak sense" iff, for any three observers a_0, a_1, a_2 in V, the following holds: If a_0 sees a_1 and a_2 in the same spatial direction at one instant of time, then he will see a_1 and a_2 in the same spatial direction at every instant of time.

In Sec. III, we will show that, in fact, Definitions 1 and 2 are equivalent, which might be somewhat surprising at first sight. Furthermore we want to relate our concept of parallax

 0022-2488/88/092064-05$02.50

to the common decomposition of the velocity gradient into rotation, shear, and volume expansion (see, e.g., Ehlers[6]). Whereas the notion of parallax is concerned with the motion of observers in arbitrary distances relative to each other, the notions of rotation, shear, and volume expansion are concerned with the motion of infinitesimally neighboring observers relative to a Fermi triad (i.e., relative to an inertial compass).

III. CHARACTERIZATION OF PARALLAX-FREE WORLD MODELS

The following theorem characterizes parallax-free world models by geometrical (P3,P4) and kinematical (P4,P5) properties, and presents a (somewhat surprising) relation to red shift. We use the notation introduced in Sec. II.

Theorem: If (M,g,V) is a world model, then the following six properties are equivalent.

P1: (M,g,V) is parallax-free in the strong sense.

P2: (M,g,V) is parallax-free in the weak sense.

P3: V is proportional to some conformal Killing vector field.

P4: There is some scalar function f on M, such that

$$\tfrac{1}{2} L_V g = -df(V)g + g(V,\cdot) \underset{s}{\otimes} df.$$

Here L_V denotes the Lie derivative with respect to V, and $\underset{s}{\otimes}$ denotes the symmetrized tensor product. Note that $\frac{1}{2} L_V g$ equals the symmetrical part of $g(\cdot,\nabla_{\cdot} V)$, where ∇ is the Levi-Civita connection of g.

P5: V is shear-free, and the one-form

$$\rho := c^{-2}[g(\nabla_V V,\cdot) - (\theta/3)g(V,\cdot)]$$

satisfies $d\rho = 0$. Here θ denotes the volume expansion of V.

P6: V admits a "red-shift potential," i.e., there is some scalar function f on M such that the following holds: If a light signal is emitted in the event p_s and received in the event p_0 (by observers in V, respectively), then the red shift $z = \Delta\lambda / \lambda$ is given by

$$\ln(1+z) = f(p_0) - f(p_s).$$

Our assumption that M is simply connected is relevant for P3 and P4, whereas the proper time normalization $g(V,V) = -c^2$ enters into P4 and P5.

This theorem and some of its implications are discussed in Sec. IV. Those readers who are interested in the mathematical details are referred to the Appendix, where the proof of the theorem is given.

IV. CONCLUSIONS

Here we want to discuss our theorem in a few words. The equivalence of P1 and P2 might be somewhat surprising. It means that, if there are parallaxes at all, then they have to be distance dependent (at least for some observers). We would like to mention that the analogous equivalence in Newtonian physics does not hold: Newtonian space-time is $\mathbb{R}^3 \times \mathbb{R}$, light signals are represented by straight lines in $\mathbb{R}^3$, and a Newtonian observer field is represented by a smooth map V: $\mathbb{R}^3 \times \mathbb{R} \to \mathbb{R}^3$. If V is of the special form $V(x,t) = V(0,t) + A(t)x$ with some matrix $A(t)$, then the flow of V maps straight lines in $\mathbb{R}^3$ onto straight lines in $\mathbb{R}^3$, i.e., V is parallax-free in the weak sense. But the angle between straight lines will be preserved by the flow of this V only if the traceless symmetric part of $A(t)$ vanishes.

The equivalence of P1 (and P2) with P3 relates parallaxes to symmetry properties of our space-time. In particular we learn from P3 that on an arbitrary space time (M,g) the existence of parallax-free observer fields V is not guaranteed (since timelike conformal Killing fields need not exist). Stated in different words: (M,g) is conformally stationary iff it admits a parallax-free observer field. Furthermore, we see at once that the property of being parallax-free is invariant under conformal transformations

$$g \mapsto e^f g, \quad V \mapsto e^{-f/2} V,$$

with some scalar function f.

The equivalence of P1 (and P2) with the kinematical characterizations P4 and P5 shows that parallaxes are related to the symmetrical part of the velocity gradient, which includes shear and volume expansion. Especially we learn from P5 that the presence of shear implies the presence of parallaxes. But note that, on the other hand, the absence of shear implies the absence of parallaxes only for observers in infinitesimally small distances; the absence of parallaxes for observers in arbitrary distances requires the further integrability condition $d\rho = 0$, which means that acceleration $\nabla_V V$ and volume expansion θ can be transformed to zero simultaneously by some conformal transformation (cf. Ehlers *et al.*,[7] Lemma 2).

The characterization of parallax-free world models by means of P6 seems to be worthy of further investigation. The existence of a red-shift potential was not expected to be in any relation to parallaxes *a priori*.

From the characterization P4 we learn that the antisymmetrical part of the velocity gradient and hence the rotation of V is not related to parallaxes. Gödel's rotating cosmos, which was considered by Treder,[1] is parallax-free in our sense: Since the standard reference frame in Gödel's cosmos is a Killing field, P3 is obviously satisfied. In Gödel's cosmos the observers do not rotate relative to each others, but all the observers rotate rigidly with respect to an inertial compass. We summarize this special implication of our theorem into the following statement: If we are traveling on some integral curve of V, we cannot learn anything about the rotation of V from the observation of parallaxes.

On this footing we want to discuss the question of whether cosmic rotation would influence the relative motion of distant quasars (as suggested by Treder[1] and Ruben[2]). In this context V represents the mean flow of galaxies (the "Hubble flow"), and our foregoing results are not applicable directly, since the observer living on the Earth does not travel on some integral curve of V. The world line of the Earth spirals around the sun, the world line of the sun spirals around the center of our galaxy, etc. Let us denote by a_0 the integral curve of V that intersects the world line of the Earth "here and now." Let a_1 and a_2 denote the world lines of two distant quasars (assumed to be integral curves of V), and let us assume that the angle under which a_1 and a_2 are seen from

the Earth has been measured during the last few years. Following Treder and Ruben, we have to ask whether the rotation of V (if it is assumed to be different from zero) has any influence of these data increasing with the distance of the quasars. From the foregoing considerations we know that the angle under which a_1 and a_2 are seen from a_0 is not influenced by rotation in any way. The correction term, which comes from the fact that the observer is situated on the Earth and not on a_0, is, in fact, influenced by rotation, but this influence does not increase with the distance of the observed objects. On the contrary, this influence vanishes if the distance between the Earth and the observed objects becomes very large compared to the distance between the Earth and a_0. In this case the world line of the Earth may be replaced by a curve ξ in the tangent bundle TM over the space-time M, such that the projection of ξ to M coincides with a_0. Then the correction term mentioned above is completely determined by the (pseudo-Euclidean) angle between ξ and the tangent to a_0 (cf. Mast and Strathdee,[3] p. 484ff for a similar consideration).

Thus we arrive at a conclusion contrary to Treder's: The apparent motion of very distant objects relative to each other will not give any information about cosmic rotation. Actually, the motion of very distant objects relative to nearby objects (our neighboring galaxies, say) might contain such information, provided the motion of these nearby objects has decoupled from cosmic rotation as it has decoupled from cosmic expansion. But, in any case, the measurement of anisotropies in the cosmic background radiation seems to be the most appropriate tool for obtaining information about cosmic rotation (see Barrow *et al.*[8]).

ACKNOWLEDGMENTS

We would like to thank Professor K.-E. Hellwig and Professor E. Sedlmayr (TU Berlin) for helpful discussions. Furthermore, one of us (V. P.) expresses his gratitude to Professor H.-J. Treder for discussing our results with him at the Einstein house in Caputh (GDR). Special thanks are due to Professor J. Ehlers (MPI Garching) for pointing out an error in an earlier version of this manuscript.

APPENDIX: PROOF OF OUR THEOREM

P1 $\Rightarrow$ P2: This is obvious.

P2 $\Rightarrow$ P5: We want to call a vector field X (defined on some open subset U of M) a "field of view lines" iff (i) X is lightlike, i.e., $g(X,X)=0$; (ii) X is geodesic, i.e., $\nabla_X X$ is parallel to X; and (iii) there is some integral curve a of V, such that every integral curve of X hits a, if it is geodesically extended. Note that the integral curve a cannot be contained in the domain U. Since (M,g,V) is assumed to be parallax-free in the weak sense, V has to form a foliation with every field of view lines X; according to Frobenius' theorem (see, e.g., Westenholz,[9] p. 216) the Lie bracket between V and X may be written on U as

$$[V,X]=h_X V+k_X X, \tag{A1}$$

with some scalar functions h_X and k_X. Now let us consider $[V,\nabla_X X]$ in two different ways: From the definition of the curvature tensor R and the fact that ∇ has vanishing torsion, we find

$$[V,\nabla_X X]=\nabla_X\nabla_V X+R(V,X,X)+\nabla_{[V,X]}X-\nabla_{\nabla_{XX}}V,$$

which, by means of (A1), can be rewritten as

$$\begin{aligned}[V,\nabla_X X]=&(\nabla_X(\nabla_{\cdot}V))(X)+\nabla_{\nabla_X X}V+X(h_X)V\\&+h_X\nabla_X V+X(k_X)X+k_X\nabla_X X+R(V,X,X)\\&+h_X(\nabla_X V+h_X V+k_X X)\\&+k_X\nabla_X X-\nabla_{\nabla_X X}V.\end{aligned}\tag{A2a}$$

Here "$X(h_X)$" should be read as "the derivation X applied to the scalar function h_X." On the other hand, $\nabla_X X$ was assumed to be parallel to X; this implies by means of (A1) that $[V,\nabla_X X]$ has to be some linear combination of V and X. Hence we find from (A2a)

$$(\nabla_X(\nabla_{\cdot}V))(X)+2h_X\nabla_X V+R(V,X,X)\in\operatorname{span}\{X,V\}.\tag{A2b}$$

In order to express the function h_X in a suitable way, we consider the well-known decomposition of the velocity gradient into its irreducible parts (see, e.g., Ehlers,[6] p. 800):

$$\begin{aligned}g(\cdot,\nabla_{\cdot}V)=&(\theta/3)[g+c^{-2}g(V,\cdot)\otimes g(V,\cdot)]\\&+\sigma+\omega-c^{-2}g(\nabla_V V,\cdot)\otimes g(V,\cdot),\end{aligned}\tag{A3}$$

where θ is the volume expansion, σ is the shear, and ω is the rotation. Symmetrization of this equation yields the Lie derivative of the metric with respect to V:

$$\tfrac{1}{2}L_V g=\tilde{\sigma}-c^{-2}g(\nabla_V V,\cdot)\underset{s}{\otimes}g(V,\cdot),\tag{A4}$$

where $\underset{s}{\otimes}$ denotes the symmetrized tensor product and

$$\tilde{\sigma}:=(\theta/3)[g+c^{-2}g(V,\cdot)\otimes g(V,\cdot)]+\sigma.\tag{A5}$$

For later convenience we rewrite Eq. (A4) by introducing the one-form ρ, which was defined in P5:

$$\tfrac{1}{2}L_V g=\rho(V)g-\rho\underset{s}{\otimes}g(V,\cdot)+\sigma.\tag{A4'}$$

Now we are ready to determine the function h_X. According to (A1) this can be written as

$$h_X=g([V,X],X)/g(V,X)=-\tfrac{1}{2}(L_V g)(X,X)/g(V,X).$$

This becomes, with (A4) and (A4'), respectively,

$$h_X=-\tilde{\sigma}(X,X)/g(V,X)+c^{-2}g(\nabla_V V,X),\tag{A6a}$$

$$h_X=\rho(X)-\sigma(X,X)/g(V,X).\tag{A6b}$$

Inserting (A6a) into (A2b) results in

$$\begin{aligned}&(\nabla_X(\nabla_{\cdot}V))(X)+2(c^{-2}g(\nabla_V V,X)\\&\quad-\tilde{\sigma}(X,X)/g(V,X))\nabla_X V+R(V,X,X)\in\operatorname{span}\{X,V\}.\end{aligned}\tag{A7}$$

Up to here X denoted a field of view lines. But note that (A7) does not contain any derivative of X. Since every lightlike vector (given at an arbitrary point p of M) can be extended to some field of view lines, (A7) has to hold for all lightlike X. We reinterpret (A7) in such a way that X should denote some lightlike vector (not a vector field) given at an arbitrary point p of M. From this purely algebraical condition we will show the vanishing of the shear σ. Let $E\neq 0$ be some vector (at the arbitrary point p) with $g(E,V)=0$. Then

$X_\pm := E \pm \sqrt{g(E,E)}\,V$ defines two lightlike vectors, which both may be inserted into (A7). Taking the difference of both these expressions and projecting onto the orthocomplement of V results in

$$[\tilde{\sigma}(E,E)/g(E,E)]\nabla_E V + A(E) = \varphi(E)E. \tag{A8}$$

Here A denotes some (1,1) tensor, and $\varphi(E)$ is a scalar factor. Whereas the special form of A will not be of any interest later on, we want to determine $\varphi(E)$ by taking the scalar product of (A8) with E. Thus we arrive at

$$g(E,E)\tilde{\sigma}(E,E)\nabla_E V + (g(E,E))^2 A(E) = [(\tilde{\sigma}(E,E))^2 + g(E,E)g(E,A(E))]E, \tag{A9}$$

which has to hold for every E with $g(E,V) = 0$. Note that $g(E,\nabla_E V) = \tilde{\sigma}(E,E)$ from (A3) and (A5). Now let us choose an orthonormal basis $\{E_1,E_2,E_3\}$ in the orthocomplement of V (at the arbitrary point p), which diagonalizes the shear σ, i.e., the E_i determine principal shear directions. Inserting some linear combination $E = \alpha E_i + \beta E_j$ of two basis vectors E_i and E_j with $i \neq j$ into (A9) and taking the scalar product with E_j results in a polynomial equation for the real coefficients α and β. Comparing the coefficients in front of $\alpha^4\beta$, we find

$$\tilde{\sigma}(E_i,E_i)\tilde{\sigma}(E_j,E_j) + g(E_j,A(E_j)) = (\tilde{\sigma}(E_i,E_i))^2 + g(E_i,A(E_i)),$$

where $\tilde{\sigma}(E_i,E_i) = g(\nabla_{E_i}V,E_i)$ has been used. Symmetrization of this equation with respect to i and j results in $(\tilde{\sigma}(E_i,E_i) - \tilde{\sigma}(E_j,E_j))^2 = 0$. Together with (A5) this implies $\sigma(E_1,E_1) = \sigma(E_2,E_2) = \sigma(E_3,E_3)$. Since σ is traceless, this can hold if and only if the shear vanishes at the arbitrary point p:

$$\sigma = 0. \tag{A10}$$

In order to complete the proof, we have to show that $d\rho = 0$. For that purpose fix an arbitrary point p of M, and consider four linearly independent fields of view lines X_1, X_2, X_3, X_4 on some neighborhood of p. Without loss of generality we may assume that

$$[V,X_i] = \rho(X_i)V, \quad \text{for } i = 1,2,3,4. \tag{A11}$$

[Remember that every field of view lines has to satisfy (A1); h_{X_i} is determined by (A6b) and (A10), whereas k_{X_i} may be assumed to vanish. If we multiply a field of view lines X by an appropriate scalar function, we get a new field of view lines $\tilde{X}$ with $k_{\tilde{X}} = 0$.] Using (A4'), (A10), (A11), the Ricci identity $\nabla g = 0$, and the fact that ∇ has vanishing torsion, we find, after some lengthy but straightforward calculation,

$$g([V,\nabla_{X_i}X_i],X_j) = 2g(X_i,X_j)[\rho(V)\rho(X_i) + X_i(\rho(V))] - 2(d\rho)(X_i,X_j)g(V,X_i) + \rho(\nabla_{X_i}X_i)g(V,X_j). \tag{A12}$$

Here "$X_i(\rho(V))$" should be read as "the derivation X_i applied to the scalar function $\rho(V)$." On the other hand, we know that $\nabla_{X_i}X_i = w_i X_i$ with some scalar function w_i. Inserting this into (A12) we find

$$(d\rho)(X_i,X_j) = u_i g(X_i,X_j), \tag{A13}$$

with some scalar function u_i (depending on X_i but not on X_j). Since $d\rho$ is antisymmetric, whereas g is symmetric, (A13) implies

$$u_i = -u_j \quad \text{if } i \neq j. \tag{A14}$$

[Note that $g(X_i,X_j) \neq 0$ if $i \neq j$.] However, (A14) can hold for $i,j = 1,2,3,4$ if and only if all the u_i vanish. This implies, by means of (A13), the desired equation $d\rho = 0$ on the neighborhood U of the arbitrarily chosen point p.

P5⇔P3: This is well known, see Ehlers,[6] p. 803 or Ehlers *et al.*,[7] p. 1346.

P3⇒P1: Assume V is proportional to some conformal Killing vector field, i.e.,

$$V = e^f \tilde{V}, \tag{A15}$$

where f is some scalar function and the Lie derivative of g with respect to $\tilde{V}$ is proportional to g:

$$L_{\tilde{V}}g = kg. \tag{A16}$$

The factor k is determined by the normalization condition $g(V,V) = -c^2$ to be

$$k = -2df(\tilde{V}), \tag{A17}$$

such that (A16) becomes

$$L_{\tilde{V}}g = -2df(\tilde{V})g. \tag{A18}$$

Now we have to show from (A15) and (A18) that (M,g,V) is parallax-free in the strong sense. Let us assume three integral curves a_0, a_1, and a_2 of V. Choose a vector field X_1 on M such that every light ray passing from a_1 to a_0 is an integral curve of X_1; similarly choose a vector field X_2 such that every light ray passing from a_2 to a_0 is an integral curve of X_2. Now X_1 and X_2 may be assumed to satisfy the following conditions everywhere:

$$g(X_i,X_i) = 0, \quad \text{for } i = 1,2, \tag{A19}$$

$$[\tilde{V},X_i] = v_i\tilde{V}, \quad \text{for } i = 1,2, \tag{A20}$$

with some scalar functions v_i. Applying $\tilde{V}$ as a derivation to (A19), we find, by means of (A18) and (A20), that the functions v_i have to vanish; hence (A20) becomes

$$[\tilde{V},X_i] = 0. \tag{A21}$$

The cosine of the angle under which a_1 and a_2 are seen from a_0 is given by

$$h_{12} = g(Y_1,Y_2)/[g(Y_1,Y_1)g(Y_2,Y_2)]^{1/2} \tag{A22}$$

along a_0, where

$$Y_i = X_i + c^{-2}g(X_i,V)V, \quad \text{for } i = 1,2, \tag{A23}$$

is that part of X_i orthogonal to V. Applying the derivation V to the function h_{12} given in (A22), we find, by means of (A15), (A18), (A21), and (A23),

$$V(h_{12}) = 0, \tag{A24}$$

i.e., h_{12} is constant along every integral curve of V (especially along a_0), which means that (M,g,V) has to be parallax-free in the strong sense.

P3⇔P4: If P3 is satisfied, we have (A15) and (A18) at our disposal as in the preceding proof. It is an easy exercise (using the usual derivation rules) to show the equivalence with P4.

P5⇔P6: According to Schrödinger,[10] p. 49, the red shift $z = \Delta\lambda/\lambda$ is given quite generally by

$$1 + z = g(V,\dot{b}(0))/g(V,\dot{b}(1)), \tag{A25}$$

where b is the affinely parametrized lightlike geodesic connecting the event of emission $p_s = b(0)$ with the event of reception $p_0 = b(1)$. Introducing the Lie derivative $L_V g$, (A25) may be easily rewritten as an integral, where $\nabla_{\dot{b}}\dot{b} = 0$ and $g(\dot{b},\dot{b}) = 0$ have to be used:

$$\ln(1+z) = -\frac{1}{2}\int_0^1 \frac{(L_V g)(\dot{b}(s),\dot{b}(s))}{g(V,\dot{b}(s))}\,ds. \tag{A26}$$

From this representation of the red shift, which is much less common than the Schrödinger formula (A25), we see at once the invariance of red shift under transformations $g \mapsto e^f g$, leaving V unchanged, and we can implement the representation (A4′) of the Lie derivative $L_V g$:

$$\ln(1+z) = \int_b \rho - \int_0^1 \frac{\sigma(\dot{b}(s),\dot{b}(s))}{g(V,\dot{b}(s))}\,ds. \tag{A27}$$

Equation (A27) proves the equivalence of P5 and P6. Assume P5 is satisfied. Then the second integral in (A27) vanishes and the first one may be rewritten in the following way: Since $d\rho = 0$, by means of Poincaré's lemma there is some scalar function f such that $\rho = df$. Hence (A27) reads

$$\ln(1+z) = \int_b df = f(p_0) - f(p_s), \tag{A28}$$

which proves P6. Assume, on the other hand, P6 is satisfied. Then (A27) becomes

$$\int_b (\rho - df) = \int_0^1 \frac{\sigma(\dot{b}(s),\dot{b}(s))}{g(V,\dot{b}(s))}\,ds, \tag{A29}$$

for all lightlike geodesics b. This implies

$$\sigma(X,X) = (\rho - df)(X)\,g(V,X) \tag{A30}$$

for all lightlike X, which is equivalent to

$$\sigma - (\rho - df) \underset{s}{\otimes} g(V,\cdot) = -\tfrac{1}{4}(\rho - df)(V)\,g. \tag{A31}$$

Here the coefficient on the right-hand side in front of the metric g has been determined from the fact that σ has vanishing trace. Using $\sigma(V,\cdot) = 0$, we find that (A31) can hold if and only if $\sigma = 0$. Inserting this into (A30) we arrive at $\rho = df$, which implies $d\rho = 0$, thereby proving P5. As a by-product this proof yields some further information about the physical meaning of the one form ρ. According to (A27) the integral over ρ equals that part of the red shift that is related to some one-form. Therefore ρ may be called the "red-shift one-form." In addition, we refer the reader to an article by Dautcourt[11] concerning space-time models admitting a red-shift potential.

[1] H.-J. Treder, Ann. Phys. (Leipzig) **42**, 71 (1985).
[2] G. Ruben, Ann. Phys. (Leipzig) **44**, 150 (1987).
[3] C. B. Mast and J. Strathdee, Proc. R. Soc. London Ser. A **252**, 476 (1959).
[4] W. Rindler, *Essential Relativity* (Springer, New York, 1977).
[5] M. v. Laue, *Das Relativitätsprinzip* (Vieweg, Braunschweig, 1911).
[6] J. Ehlers, Akad. Wiss. Lit. (Mainz), Abh. Math. Nat. Kl. **1961** (11), 791.
[7] J. Ehlers, P. Geren, and R. K. Sachs, J. Math. Phys. **9**, 1344 (1968).
[8] J. D. Barrow, R. Juskiewicz, and D. H. Sonoda, Mon. Not. R. Astron. Soc. **213**, 917 (1985).
[9] C. v. Westenholz, *Differential Forms in Mathematical Physics* (North-Holland, Amsterdam, 1981).
[10] E. Schrödinger, *Expanding Universes* (Cambridge U.P., Cambridge, 1956).
[11] G. Dautcourt, Astron. Nachr. **308**, 293 (1987).

Gravitational instability in a multicomponent cosmological medium

A. M. Mathai
Department of Mathematics and Statistics, McGill University, Burnside Hall, 805 Sherbrooke Street West, Montreal, Quebec H3A 2K6, Canada

H. J. Haubold, J. P. Mücket, S. Gottlöber, and V. Müller
Central Institute for Astrophysics, Academy of Sciences of the GDR, DDR-1591 Potsdam-Babelsberg, German Democratic Republic

(Received 23 February 1988; accepted for publication 20 April 1988)

A systematic approach to exact solutions for the evolution of small density inhomogeneities of a two-component and in special cases for a three-component cosmological medium is presented. In general, there exists just one increasing perturbation mode. Expressing the solutions of the respective differential equations in terms of Meijer's G function [Proc. Nederl. Akad. Wetensch. A **49**, 344 (1946)] various exact and asymptotic formulas for the perturbation modes are derived.

I. INTRODUCTION

Surely gravitational condensation must be considered to be the reason for the formation of the basic building blocks of the Universe, i.e., the stars and galaxies and systems of them at various scales. Originally the effect of gravitational clumping acting in an initially uniform diffuse matter was suggested by Newton. The first theory of gravitational condensation of perturbations in a homogeneous medium was given in the classic work of Jeans.[1] The main result of Jeans is his criterion that in a homogeneous nonrelativistic medium oscillations with a wavelength λ greater than some critical wavelength λ_J cannot exist, and all perturbations with $\lambda > \lambda_J$ condense gravitationally. However, Jeans assumed a static homogeneous medium, an assumption not in accordance with the real situation in the Universe. In 1946 Lifshits[2] gave a general-relativistic theory of the evolution of instabilities in the expanding universe and in 1957 Bonnor[3] obtained results in the Newtonian limit (the theory is developed in detail in some textbooks, see, e.g., Weinberg,[4] Zeldovich and Novikov,[5] and Peebles[6]). These basic papers were concerned with single-component baryonic matter.

The inflationary cosmological models developed during the last years suggest that the spatial curvature in the homogeneous and isotropic world models is zero with high accuracy and the cosmological constant can be assumed to be zero, too (some arguments for this are presented, for example, by Gottlöber *et al.*[7]). Therefore, the total density parameter must be equal to 1. It is a matter of fact that galaxies and clusters of galaxies contain matter that is hidden from direct observation, hence called nonluminous or dark matter. From cosmological arguments it appears that this material cannot be composed only of baryons and of particles having a zero rest mass (Shandarin *et al.*[8]). The best candidate for the dark matter component in the Universe is still the neutrino of finite rest mass which could dominate the dynamics of the Universe through its gravitational interaction with other forms of matter.

The linear theory of density perturbations in a universe containing a neutrino component and baryonic matter has been analyzed by Wasserman.[9] The influence of the component's relative motion upon the gravitational instability was investigated in the Newtonian limit by Polyachenko and Fridman[10] and Grishchuk and Zeldovich.[11] Fargion[12] and Soloveva and Nurgaliev[13] have treated the corresponding problems for a multicomponent medium in the expanding universe. In Newtonian approximation they obtained asymptotic expressions for the solutions as $k \to 0$ and $k \to \infty$, where $k = 2\pi a(t)/\lambda$ is the perturbation wave number and $a(t)$ is the universal expansion scale of the Friedmann model. Assuming an equation of state $P_i \propto \rho_i^{\gamma_i}$ and a sound velocity $\beta_i^2 = dP_i/d\rho_i \propto \rho_i^{\gamma_i - 1}$, Fargion[12] and Soloveva and Starobinski[14] established exact solutions for the evolution of small density perturbations in a two-component medium against the background of an expanding Friedmann model.

In this paper we shall present a systematic approach to exact solutions for the evolution of small density perturbations of a two-component medium against the background of an expanding Friedmann universe in the Newtonian approximation. In particular, we give closed-form expressions for the following cases of the physical parameters:

$$\begin{aligned}
&\beta_1 = \beta_2 = 0, \\
&\gamma_1 = \gamma_2 = \tfrac{4}{3}, && \beta_1, \beta_2 \text{ arbitrary}, \\
&\gamma_2 \neq \tfrac{4}{3}, && \beta_1 = 0, \\
&\gamma_2 = \tfrac{4}{3}, && \beta_1 \neq 0, \\
&\gamma_1 = \gamma_2 \neq \tfrac{4}{3}, && \beta_1 = \beta_2 \neq 0.
\end{aligned}$$

Further, the behavior of perturbations in those three-component models are discussed, and they can be explicitly solved. In particular, a special solution is given for a case of special cosmological interest that leads to qualitatively new physical effects. In this case we have two types of dark matter, one dustlike component with $\beta_3 = 0$ and one relativistic component with $\gamma_2 = \frac{4}{3}$. The third component, favorable to the baryonic component, has arbitrary γ_1 and β_1.

Special attention will be devoted to the mathematical method of obtaining closed-form solutions of the coupled systems of fundamental second-order differential equations that govern the growth or decay of gravitational condensations in an expanding Friedmann universe. This is the appli-

 0022-2488/88/092069-09$02.50

cation of the theory of Meijer's G function that received some interest in the astrophysical literature recently (cf. Soloveva and Starobinski[14] and Mathai and Haubold[15]). The formalism is of interest in an other context, too. Recent investigations (Haubold *et al.*[16] and Mücket *et al.*[17]) have shown that the coupling of perturbation modes on different length scales can lead to a strong condensation effect. Exact solutions for that case may be obtained by the methods developed in this paper.

II. DIFFERENTIAL EQUATIONS FOR DENSITY PERTURBATIONS

Considering a two-component medium with densities ρ_i ($i = 1,2$) in the Newtonian limit the equations of hydrodynamics and gravitation are given by

$$\frac{\partial \rho_i}{\partial t} + \nabla(\rho_i \mathbf{U}_i) = 0, \tag{1}$$

$$\frac{\partial \mathbf{U}_i}{\partial t} + (\mathbf{U}_i \nabla)\mathbf{U}_i + \frac{1}{\rho_i}\nabla P_i + \nabla\varphi = 0, \tag{2}$$

$$\Delta\varphi = \nabla\nabla\varphi = 4\pi G \sum_i \rho_i, \tag{3}$$

where ρ_i, $\mathbf{U}_i$, P_i, and φ denote the density, velocity, pressure, and gravitational fields, respectively. The expansion will be described by the Hubble flow $\mathbf{U}_i = H\mathbf{x}_i$, where H is the Hubble parameter. We may employ the hydrodynamic approximation for the description of cosmological perturbations given by plane waves with wavelength $\lambda = 2\pi|\mathbf{k}|^{-1}$, where $\mathbf{k}(t)$ is the wave vector. Thus it holds for the density perturbations $\delta\rho_i/\rho_i = \delta_i(t)\exp\{i|\mathbf{k}|\cdot|\mathbf{x}|\}$. The evolution of the amplitudes $\delta_i(t)$ for one particular mode is described in the linearized approach by the following system of second-order differential equations (Weinberg,[4] Zeldovich and Novikov,[5] and Peebles[6]):

$$\frac{d^2\delta_1}{dt^2} + 2\left(\frac{1}{a}\frac{da}{dt}\right)\frac{d\delta_1}{dt} + k^2\beta_1^2\delta_1 = 4\pi G(\rho_1\delta_1 + \rho_2\delta_2), \tag{4}$$

$$\frac{d^2\delta_2}{dt^2} + 2\left(\frac{1}{a}\frac{da}{dt}\right)\frac{d\delta_2}{dt} + k^2\beta_2^2\delta_2 = 4\pi G(\rho_1\delta_1 + \rho_2\delta_2), \tag{5}$$

where for both perturbations the same wave vector $\mathbf{k}(t)$ is used. During the dustlike stage of the cosmological evolution, i.e., when the matter dominates over radiation (the pressure $P_i \ll \rho_i c^2$), one has

$$a(t) \propto t^{2/3}, \quad \rho_i(t) = \Omega_i/6\pi G t^2,$$
$$\Omega_i = \text{const}, \quad \Omega_1 + \Omega_2 = 1, \tag{6}$$

where G is the gravitational constant, and Ω_1 and Ω_2 denote the density parameters of the two-fluid medium. Specifying the equation of state for the fluids,

$$P_i \propto \rho_i^{\gamma_i}, \tag{7}$$

the sound velocities may be written as

$$\beta_i^2 = \frac{dP_i}{d\rho_i} \propto \rho_i^{\gamma_i - 1} \propto t^{2(1-\gamma_i)}. \tag{8}$$

During the expansion the Hubble parameter changes as $H = \eta t^{-1}$ with $\eta = \frac{2}{3}$. Note that the following equations are also valid, if the evolution of two perturbations not coupled to the radiation is considered in the radiation dominated universe. Then one has to put $\eta = \frac{1}{2}$. The wave vector is proportional to a^{-1} so that $k^2\beta_i^2 = k_i^2 t^{2(1-\eta-\gamma_i)}$, where the constants k_i come from both the wave vector and the velocity of sound. If $k_i = 0$ the sound velocity β_i equals zero, in which case the adiabatic index γ_i loses its sense. Multiplying Eqs. (4) and (5) by t^2 and setting $\alpha_i = 2(2 - \eta - \gamma_i)$, one obtains

$$t^2\frac{d^2\delta_1}{dt^2} + 2\eta t\frac{d\delta_1}{dt} + k_1^2 t^{\alpha_1}\delta_1 = \frac{2}{3}(\Omega_1\delta_1 + \Omega_2\delta_2), \tag{9}$$

$$t^2\frac{d^2\delta_2}{dt^2} + 2\eta t\frac{d\delta_2}{dt} + k_2^2 t^{\alpha_2}\delta_2 = \frac{2}{3}(\Omega_1\delta_1 + \Omega_2\delta_2). \tag{10}$$

With the operator

$$\Delta = t\frac{d}{dt} \Rightarrow t^2\frac{d^2}{dt^2} = \Delta^2 - \Delta, \tag{11}$$

Eqs. (9) and (10) read

$$\Delta^2\delta_1 + (2\eta - 1)\Delta\delta_1 + k_1^2 t^{\alpha_1}\delta_1 = \tfrac{2}{3}(\Omega_1\delta_1 + \Omega_2\delta_2), \tag{12}$$

$$\Delta^2\delta_2 + (2\eta - 1)\Delta\delta_2 + k_2^2 t^{\alpha_2}\delta_2 = \tfrac{2}{3}(\Omega_1\delta_1 + \Omega_2\delta_2). \tag{13}$$

Taking δ_2 from (12) one has

$$\delta_2 = \frac{3}{2\Omega_2}\Delta^2\delta_1 + (2\eta - 1)\frac{3}{2\Omega_2}\Delta\delta_1 + \left(\frac{3k_1^2 t^{\alpha_1}}{2\Omega_2} - \frac{\Omega_1}{\Omega_2}\right)\delta_1,$$

$$\Delta\delta_2 = \frac{3}{2\Omega_2}\Delta^3\delta_1 + (2\eta - 1)\frac{3}{2\Omega_2}\Delta^2\delta_1 + \left(\frac{3k_1^2 t^{\alpha_1}}{2\Omega_2} - \frac{\Omega_1}{\Omega_2}\right)\Delta\delta_1 + \frac{3k_1^2\alpha_1 t^{\alpha_1}}{2\Omega_2}\delta_1,$$

$$\Delta^2\delta_2 = \frac{3}{2\Omega_2}\Delta^4\delta_1 + (2\eta - 1)\frac{3}{2\Omega_2}\Delta^3\delta_1 + \left(\frac{3k_1^2 t^{\alpha_1}}{2\Omega_2} - \frac{\Omega_1}{\Omega_2}\right)\Delta^2\delta_1 + \frac{6k_1^2\alpha_1 t^{\alpha_1}}{2\Omega_2}\Delta\delta_1 + \frac{3k_1^2\alpha_1^2 t^{\alpha_1}}{2\Omega_2}\delta_1.$$

Substituting δ_2, $\Delta\delta_2$, and $\Delta^2\delta_2$ in (13) we have the following result:

$$\begin{aligned}
&\Delta^4\delta_1 + 2(2\eta - 1)\Delta^3\delta_1 \\
&\quad + \left[k_1^2 t^{\alpha_1} + k_2^2 t^{\alpha_2} - \tfrac{2}{3} + (2\eta - 1)^2\right]\Delta^2\delta_1 \\
&\quad + \left[(2\eta - 1)k_1^2 t^{\alpha_1} + (2\eta - 1)k_2^2 t^{\alpha_2}\right. \\
&\quad \left. + 2k_1^2\alpha_1 t^{\alpha_1} - (2\eta - 1)\tfrac{2}{3}\right]\Delta\delta_1 \\
&\quad + \left[k_1^2\alpha_1^2 t^{\alpha_1} + (2\eta - 1)k_1^2\alpha_1 t^{\alpha_1} - \tfrac{2}{3}\Omega_2 k_1^2 t^{\alpha_1}\right. \\
&\quad \left. - \tfrac{2}{3}\Omega_1 k_2^2 t^{\alpha_2} + k_1^2 k_2^2 t^{\alpha_1 + \alpha_2}\right]\delta_1 = 0.
\end{aligned} \tag{14}$$

Equation (14) is the fundamental fourth-order differential equation that governs the growth and decay of gravitational condensations in the expanding two-fluid universe.

III. CLASSIFICATION OF THE TWO-COMPONENT MODELS

We shall now classify the system of differential equations (14) dependent of the values of the formal parameters α_i and k_i, i.e., independent of the various physical models for the parameter γ_i in the equation of state and the velocity of sound β_i. Note that Eq. (14) is an equation for one solution δ_1 only, and we obtain the second solution δ_2 by writing the corresponding symmetric case. In the manner, the cases (3) and (4) and the cases (5) and (6) of the following classification describe the complementary solutions in the same physical situation.

(1) $k_1 = k_2 = 0$ ($\beta_1 = \beta_2 = 0$). In this case the differential equation (14) reduces to the following form:

$$\{\Delta^4 + 2(2\eta-1)\Delta^3 + [(2\eta-1)^2 - \tfrac{2}{3}]\Delta^2 - \tfrac{2}{3}(2\eta-1)\Delta\}\delta_1 = 0. \tag{15}$$

Then one has

$$\{\Delta[\Delta + (2\eta-1)][\Delta^2 + (2\eta-1)\Delta - \tfrac{2}{3}]\}\delta_1 = 0, \tag{16}$$

which implies that

$$\{(\Delta - a_1)(\Delta - a_2)(\Delta - a_3)(\Delta - a_4)\}\delta_1 = 0, \tag{17}$$

where

$$\begin{aligned} a_1 &= 0, \quad a_2 = -(2\eta-1), \\ a_3 &= (\tfrac{1}{2} - \eta) - [(\tfrac{1}{2} - \eta)^2 + \tfrac{2}{3}]^{1/2}, \\ a_4 &= (\tfrac{1}{2} - \eta) + [(\tfrac{1}{2} - \eta)^2 + \tfrac{2}{3}]^{1/2}. \end{aligned} \tag{18}$$

Hence we have the solution

$$\delta_1 = C_1 + C_2 t^{a_2} + C_3 t^{a_3} + C_4 t^{a_4}, \tag{19}$$

where C_1, C_2, C_3, and C_4 are arbitrary constants. Note that from Eq. (14) we can write down an equation similar to (17) for δ_2 that leads to a similar solution for δ_2 as given in (19). Then by using the initial conditions one can get relationships between the arbitrary constants C_1, C_2, C_3, and C_4 for δ_1 and δ_2 (see Sec. VII).

(2) $\alpha_1 = \alpha_2 = 0$, k_1, k_2 arbitrary ($\gamma_1 = \gamma_2 = \frac{4}{3}$, β_1, β_2 arbitrary). In this case we have for the differential equation (14)

$$\begin{aligned} \{\Delta^4 &+ 2(2\eta-1)\Delta^3 + [(2\eta-1)^2 - \tfrac{2}{3} + k_1^2 + k_2^2]\Delta^2 \\ &+ [-\tfrac{2}{3}(2\eta-1) + (2\eta-1)(k_1^2 + k_2^2)]\Delta \\ &+ [-\tfrac{2}{3}\Omega_2 k_1^2 - \tfrac{2}{3}\Omega_1 k_2^2 + k_1^2 k_2^2]\}\delta_1 = 0. \end{aligned} \tag{20}$$

Then the general solution is of the form

$$\delta_1 = C_1 t^{a_1} + C_2 t^{a_2} + C_3 t^{a_3} + C_4 t^{a_4}, \tag{21}$$

where C_1, C_2, C_3, and C_4 are arbitrary constants and a_1, a_2, a_3, and a_4 are the roots of the equation

$$\begin{aligned} a^4 &+ 2(2\eta-1)a^3 + [(2\eta-1)^2 - \tfrac{2}{3} + k_1^2 + k_2^2]a^2 \\ &+ [-\tfrac{2}{3}(2\eta-1) + (2\eta-1)(k_1^2 + k_2^2)]a \\ &+ [-\tfrac{2}{3}\Omega_2 k_1^2 - \tfrac{2}{3}\Omega_1 k_2^2 + k_1^2 k_2^2] = 0. \end{aligned} \tag{22}$$

(3) $\alpha_2 \neq 0$, $k_1 = 0$ ($\beta_1 = 0$). The case $\alpha_2 = 0$ is excluded since it is already covered as a special case of (2). Here it is useful to transform the time variable t into x by

$$x = k_2^2 t^{\alpha_2}/\alpha_2^2,$$

so that

$$\Delta^l = \alpha_2^l \underset{\sim}{\Delta}^l, \quad l = 1,2,$$

where we have denoted

$$\underset{\sim}{\Delta} = x\frac{d}{dx}.$$

With this transformation and dividing Eq. (14) by α_2^4 we have

$$\begin{aligned} \Big\{\underset{\sim}{\Delta}^4 &+ \frac{2(2\eta-1)}{\alpha_2}\underset{\sim}{\Delta}^3 \\ &+ \frac{(2\eta-1)^2 - \frac{2}{3}}{\alpha_2^2}\underset{\sim}{\Delta}^2 - \frac{2(2\eta-1)}{3\alpha_2^3}\underset{\sim}{\Delta}\Big\}\delta_1 \\ &+ x\Big\{\underset{\sim}{\Delta}^2 + \frac{(2\eta-1)}{\alpha_2}\underset{\sim}{\Delta} - \frac{2\Omega_1}{3\alpha_2^2}\Big\}\delta_1 = 0. \end{aligned} \tag{23}$$

Equation (23) can be written in the form

$$\begin{aligned} &\{(\underset{\sim}{\Delta} - b_1)(\underset{\sim}{\Delta} - b_2)(\underset{\sim}{\Delta} - b_3)(\underset{\sim}{\Delta} - b_4)\}\delta_1 \\ &\quad + x\{(\underset{\sim}{\Delta} - a_1)(\underset{\sim}{\Delta} - a_2)\}\delta_1 = 0, \end{aligned} \tag{24}$$

where the constants are

$$\begin{aligned} b_1 &= 0, \quad b_2 = -(2\eta-1)/\alpha_2, \\ b_3 &= \frac{(\frac{1}{2} - \eta)}{\alpha_2} - \left[\frac{(\eta - \frac{1}{2})^2}{\alpha_2^2} + \frac{2}{3\alpha_2^2}\right]^{1/2}, \\ b_4 &= \frac{(\frac{1}{2} - \eta)}{\alpha_2} + \left[\frac{(\eta - \frac{1}{2})^2}{\alpha_2^2} + \frac{2}{3\alpha_2^2}\right]^{1/2}, \\ a_1 &= \frac{(\frac{1}{2} - \eta)}{\alpha_2} - \left[\frac{(\frac{1}{2} - \eta)^2}{\alpha_2^2} + \frac{2\Omega_1}{3\alpha_2^2}\right]^{1/2}, \\ a_2 &= \frac{(\frac{1}{2} - \eta)}{\alpha_2} + \left[\frac{(\frac{1}{2} - \eta)^2}{\alpha_2^2} + \frac{2\Omega_1}{3\alpha_2^2}\right]^{1/2}. \end{aligned} \tag{25}$$

Equation (24) is a special case of the differential equation satisfied by Meijer's G function (see, for example, Meijer,[18] Luke,[19] and Mathai and Saxena[20]). In the next three cases we transform the original Eq. (14) into an equation with the same structure as Eq. (24), but with other constants than those given in (25). Hence we will reduce all the cases to a differential equation of the type (24), and we can solve them together.

(4) $\alpha_1 \neq 0$, $k_2 = 0$ ($\beta_2 = 0$). Proceeding as in the preceding case we transform Eq. (14) into

$$\begin{aligned} &\{(\underset{\sim}{\Delta} - b_1')(\underset{\sim}{\Delta} - b_2')(\underset{\sim}{\Delta} - b_3')(\underset{\sim}{\Delta} - b_4')\}\delta_1 \\ &\quad + x\{(\underset{\sim}{\Delta} - a_1')(\underset{\sim}{\Delta} - a_2')\}\delta_1 = 0, \end{aligned} \tag{26}$$

where we have denoted

$$\underset{\sim}{\Delta} = \underset{\sim}{x}\frac{d}{d\underset{\sim}{x}}, \quad \underset{\sim}{x} = \frac{k_1^2 t^{\alpha_1}}{\alpha_1^2},$$

and the constants $b_i' = b_i$ are given by Eq. (25) and the constants a_i' are

$$\begin{aligned} a_1' &= \left[-1 + \frac{\frac{1}{2} - \eta}{\alpha_1}\right] - \left[\frac{(\eta - \frac{1}{2})^2}{\alpha_1^2} + \frac{2\Omega_2}{3\alpha_1^2}\right]^{1/2}, \\ a_2' &= \left[-1 + \frac{\frac{1}{2} - \eta}{\alpha_1}\right] + \left[\frac{(\eta - \frac{1}{2})^2}{\alpha_1^2} + \frac{2\Omega_2}{3\alpha_1^2}\right]^{1/2}. \end{aligned} \tag{27}$$

(5) $\alpha_1 = 0$, $\alpha_2 \neq 0$ ($\gamma_1 = \frac{4}{3}$). With the same transformations as in case (3), we obtain again Eq. (24) where now the

constants a_1 and a_2 are given by

$$a_1 = \frac{\frac{1}{2}-\eta}{\alpha_2^2} - \left[\frac{(\frac{1}{2}-\eta)^2}{\alpha_2^2} + \frac{2\Omega_1}{3\alpha_2^2} - \frac{k_1^2}{\alpha_2^2}\right]^{1/2},$$
$$a_2 = \frac{\frac{1}{2}-\eta}{\alpha_2^2} + \left[\frac{(\frac{1}{2}-\eta)^2}{\alpha_2^2} + \frac{2\Omega_1}{3\alpha_2^2} - \frac{k_1^2}{\alpha_2^2}\right]^{1/2}, \quad (28)$$

and the constants b_i are the roots of the equation

$$\alpha_2^4 b^4 + 2(2\eta-1)\alpha_2^3 b^3 + [(2\eta-1)^2 - \tfrac{2}{3} + k_1^2]\alpha_2^2 b^2 + [-\tfrac{2}{3}(2\eta-1) + (2\eta-1)k_1^2]\alpha_2 b - \tfrac{2}{3}\Omega_2 k_1^2 = 0. \quad (29)$$

(6) $\alpha_1 \neq 0, \alpha_2 = 0$ ($\gamma_2 = \frac{4}{3}$). With the same transformations as in case (4), we obtain again Eq. (26) where now the constants a_1' and a_2' are given by

$$a_1' = \left[1 + \frac{\eta - \frac{1}{2}}{\alpha_1}\right] - \left[\frac{(\eta-\frac{1}{2})^2}{\alpha_1^2} + \frac{2\Omega_2}{3\alpha_1^2} - \frac{k_2^2}{\alpha_1^2}\right]^{1/2},$$
$$a_2' = \left[1 + \frac{\eta - \frac{1}{2}}{\alpha_1}\right] + \left[\frac{(\eta-\frac{1}{2})^2}{\alpha_1^2} + \frac{2\Omega_2}{3\alpha_1^2} - \frac{k_2^2}{\alpha_1^2}\right]^{1/2}, \quad (30)$$

and the constants b_i' are roots of Eq. (29) after substituting α_2 by α_1.

(7) $\alpha_1 = \alpha_2 = \alpha \neq 0$, $k_1 = k_2 = k \neq 0$ ($\gamma_1 = \gamma_2 \neq \frac{4}{3}$). In this case we make the following substitutions:

$$\delta_1 = f_1 + f_2, \quad \delta_2 = f_1 - (\Omega_1/\Omega_2)f_2.$$

Equations (12) and (13) are reduced to

$$\Delta^2(f_1+f_2) + (2\eta-1)\Delta(f_1+f_2) + k^2t^2(f_1+f_2) = \tfrac{2}{3}f_1, \quad (31)$$

$$\Delta^2(f_1 - (\Omega_1/\Omega_2)f_2) + (2\eta-1)\Delta(f_1 - (\Omega_1/\Omega_2)f_2) + k^2t^\alpha(f_1 - (\Omega_1/\Omega_2)f_2) = \tfrac{2}{3}f_1. \quad (32)$$

Multiply Eq. (31) by Ω_1 and Eq. (32) by Ω_2 and add to get

$$[\Delta^2 + (2\eta-1)\Delta - \tfrac{2}{3}]f_1 + k^2t^2 f_1 = 0. \quad (33)$$

Change the time variable t to

$$u = \frac{k^2 t^\alpha}{\alpha^2}, \quad \underset{\sim}{\Delta} = u\frac{d}{du},$$

and proceed as in case (3) to get

$$[(\underset{\sim}{\Delta} - b_1)(\underset{\sim}{\Delta} - b_2)]f_1 + uf_1 = 0, \quad (34)$$

where the constants b_1 and b_2 are given by

$$b_1 = \frac{\frac{1}{2}-\eta}{\alpha} - \left[\frac{(\eta-\frac{1}{2})^2}{\alpha^2} + \frac{2}{3\alpha^2}\right]^{1/2},$$
$$b_2 = \frac{\frac{1}{2}-\eta}{\alpha} + \left[\frac{(\eta-\frac{1}{2})^2}{\alpha^2} + \frac{2}{3\alpha^2}\right]^{1/2}. \quad (35)$$

From Eqs. (31) and (32) the equation for f_2 is

$$[\underset{\sim}{\Delta}^2 + (2\eta-1)\underset{\sim}{\Delta}]f_2 + k^2t^\alpha f_2 = 0, \quad (36)$$

from which after transforming t into u Eq. (34) again follows, but now with the constants $b_1 = 0$ and $b_2 = -(2\eta-1)/\alpha$. Note that all the above procedures cannot be applied for the only remaining cases $\alpha_1 = \alpha_2 = \alpha \neq 0$, $k_1 \neq k_2 \neq 0$, and $\alpha_1 \neq \alpha_2 \neq 0$.

IV. GENERAL SOLUTIONS FOR THE CASES (3)–(6)

In all the considered cases the differential equation for δ_1 reduces to the following form:

$$\{(\Delta - b_1)(\Delta - b_2)(\Delta - b_3)(\Delta - b_4)\}\delta_1 + x\{(\Delta - a_1)(\Delta - a_2)\}\delta_1 = 0, \quad (37)$$

where the quantities $a_1, a_2, b_1, \ldots, b_4$ and x change from case to case as given in Sec. III. Equation (37) is a particular case of the differential equation satisfied by Meijer's G function of the form $G^{m,n}_{p,q}(x|^{a_1,\ldots,a_p}_{b_1,\ldots,b_q})$, see, for example, Meijer,[18] Luke,[19] and Mathai and Saxena.[20] In Eq. (37) we have the G-function parameters $q = 4$, $p = 2$, $(-1)^{p-m-n} = -1$, a_1, a_2, and $b_1, \ldots, b_4$. From Luke[19] or from Meijer[18] we have the explicit forms of the solutions as given in the following.

A. The solution near $x = 0$

In this case we have

$$\delta_1 = C_1G_1 + C_2G_2 + C_3G_3 + C_4G_4,$$

where the C_1, C_2, C_3, and C_4 are arbitrary constants and

$$G_j = \frac{[\Pi_{k=1}^2 \Gamma(-a_k + b_j)]}{[\Pi_{k=1}^4 \Gamma(1 - b_k + b_j)]} x^{b_j} {}_2F_3(-a_1 + b_j, -a_2 + b_j; 1 - b_1 + b_j, .., *, 1 - b_4 + b_j; -x). \quad (38)$$

An $*$ indicates that the parameter of the type $1 - b_j + b_j$ and the corresponding gamma are absent and it is assumed that $b_i - b_j \neq 0, \pm 1, \pm 2, \ldots$ for all $i, j = 1, \ldots, 4$, $i \neq j$. In Eq. (38) ${}_2F_3$ denotes a hypergeometric function. Note also that for small values of x the function ${}_2F_3$ behaves like 1 and hence for small x one can approximate G_j by

$$G_j \approx \frac{[\Pi_{k=1}^2 \Gamma(-a_k + b_j)]}{[\Pi_{k=1}^4 \Gamma(1 - b_k + b_j)]} x^{b_j}. \quad (39)$$

Equation (38) is also suitable for computing explicit values of G_j for finite values of x.

B. The solution for $x \to \infty$

Since the G-function parameters are $q = 4$ and $p = 2$ in our case, we have $q > p$. Then the four fundamental solutions of (37) are the following (see also Luke[19]):

$$\delta_1 = C_1G_1 + C_2G_2 + C_3G_3 + C_4G_4,$$

where $C_1, \ldots, C_4$ are constants and

$$G_1 = G^{4,1}_{2,4}\left(x \middle| \begin{matrix} 1 + a_1, 1 + a_2 \\ b_1, \ldots, b_4 \end{matrix}\right),$$
$$G_2 = G^{4,1}_{2,4}\left(x \middle| \begin{matrix} 1 + a_2, 1 + a_1 \\ b_1, \ldots, b_4 \end{matrix}\right),$$
$$G_3 = G^{4,0}_{2,4}\left(xe^{i\pi} \middle| \begin{matrix} 1 + a_1, 1 + a_2 \\ b_1, \ldots, b_4 \end{matrix}\right),$$
$$G_4 = G^{4,0}_{2,4}\left(xe^{-i\pi} \middle| \begin{matrix} 1 + a_1, 1 + a_2 \\ b_1, \ldots, b_4 \end{matrix}\right), \quad (40)$$

where $i^2 = -1$.

C. Evaluation of G_1 and G_2

For example, writing the G function as an inverse Mellin transformation, we have

$$G_1=\frac{1}{2\pi i}\int_L ds\,\frac{[\Pi_{j=1}^4\Gamma(b_j+s)\Gamma(-a_1-s)]}{\Gamma(1+a_2+s)}x^{-s}, \tag{41}$$

where L is a contour enclosing the poles of $\Gamma(b_j+s)$, $j=1,...,4$. The poles of $\Gamma(b_j+s)$ are $s=-b_j-\nu$, $\nu=0,1,...$. If $b_j-b_k\neq 0,\pm 1,\pm 2,...$ and $j\neq k=1,...,4$, then the residue at $s=-b_j-\nu$ is given by

$$\begin{aligned}R_{\nu j}&=\lim_{s\to -b_j-\nu}\Gamma(s+b_j+\nu)\\&\quad\times\frac{[\Pi_{j=1}^4\Gamma(b_j+s)\Gamma(-a_1-s)]}{\Gamma(a_2+s)}x^{-s}\\&=x^{b_j}[\Pi'^4_{k=1}\Gamma(b_k-b_j-\nu)]\\&\quad\times\frac{\Gamma(-a_1+b_j+\nu)}{\Gamma(a_2+s)}x^{-s},\end{aligned}\tag{42}$$

where Π' denotes that the gamma of the form $\Gamma(b_j-b_j-\nu)$ is absent. Now using the formula

$$\Gamma(\alpha-\nu)=\frac{\Gamma(\alpha)(-1)^\nu}{(-\alpha+1)_\nu},$$
$$(a)_\nu=a(a+1)\cdots(a+\nu-1),\quad (a)_0=1,$$
$$\Gamma(\alpha+\nu)=\Gamma(\alpha)(\alpha)_\nu,$$

we have

$$\begin{aligned}R_{\nu j}&=x^{b_j}\left[\prod_{k=1}^4\Gamma(b_k-b_j)\right]\frac{\Gamma(-a_1+b_j)}{\Gamma(1+a_2-b_j)}\\&\quad\times\frac{[\Pi'^2_{k=1}(-a_k+b_j)_\nu]}{[\Pi'^4_{k=1}(1-b_k+b_j)_\nu]}\frac{(-1)_\nu}{\nu!}x^\nu.\end{aligned}\tag{43}$$

Thus the sum of the residues, denoted by R_j, is therefore

$$\begin{aligned}R_j&=\sum_{\nu=0}^\infty R_{\nu j}\\&=x^{b_j}\left[\prod_{k=1}^4{}'\Gamma(b_k-b_j)\right]\frac{\Gamma(-a_1+b_j)}{\Gamma(1+a_2-b_j)}\\&\quad\times{}_2F_3(-a_1+b_j,-a_2+b_j;1-b_1+b_j,..*,\\&\quad 1-b_4+b_j;-x),\end{aligned}\tag{44}$$

where an $*$ indicates that the parameter of the type $1-b_j+b_j$ is absent. Thus

$$G_1=\sum_{j=1}^4 R_j,\tag{45}$$

where R_j is given in (44). Note that G_2 is obtained by interchanging a_1 and a_2 in G_1 as can be seen in the G functions given in (40).

In the case $x\to\infty$ we can use the standard formula (Mathai and Saxena[20])

$$G_{p,q}^{m,n}\left(z\middle|\begin{matrix}a_1,...,a_p\\b_1,...,b_q\end{matrix}\right)=G_{q,p}^{n,m}\left(\frac{1}{z}\middle|\begin{matrix}1-b_1,...,1-b_q\\1-a_1,...,1-a_p\end{matrix}\right).$$

Hence

$$\begin{aligned}G_1&=G_{4,2}^{1,4}\left(\frac{1}{x}\middle|\begin{matrix}1-b_1,...,1-b_4\\-a_1,-a_2\end{matrix}\right)\\&=\frac{1}{2\pi i}\int_L ds\,\frac{\Gamma(-a_1+s)[\Pi_{j=1}^4\Gamma(b_j-s)]}{\Gamma(1+a_2-s)}\left(\frac{1}{x}\right)^{-s}.\end{aligned}$$

Evaluating, as before, as the sum of the residues of $\Gamma(-a_1+s)$, we have

$$G_1=\sum_{\nu=0}^\infty\frac{(-1)^\nu}{\nu!}\left(\frac{1}{x}\right)^{-a_1+\nu}\frac{[\Pi_{k=1}^4\Gamma(b_k-a_1+\nu)]}{\Gamma(1+a_2-a_1+\nu)}.$$

Thus G_1 is approximated to the leading terms. That is,

$$G_1\approx\frac{[\Pi_{k=1}^4\Gamma(b_k-a_1)]}{\Gamma(1+a_2-a_1)}\left(\frac{1}{x}\right)^{-a_1}.\tag{46}$$

Interchange a_1 and a_2 to get an approximation for G_2 when $x\to\infty$.

D. Evaluation of G_3 and G_4

Similarly as in the case of G_1 in (41) we write for G_3 given in (40),

$$G_3=\frac{1}{2\pi i}\int_L ds\,\frac{[\Pi_{j=1}^4\Gamma(b_j+s)]}{[\Pi_{j=1}^2\Gamma(1+a_j+s)]}(xe^{i\pi})^{-s}.\tag{47}$$

Let $b_j-b_k\neq 0,\pm 1,..., j\neq k=1,...,4$. Then the poles are simple and evaluating as before we have the following:

$$\begin{aligned}G_3&=\sum_{j=1}^4 x^{b_j}e^{i\pi b_j}\left[\frac{[\Pi'^4_{k=1}\Gamma(b_k-b_j)]}{\Pi^2_{k=1}\Gamma(1+a_k-b_j)}\right]\\&\quad\times{}_2F_3(b_j-a_1,b_j-a_2;\\&\quad 1+b_j-b_1,..*,1+b_j-b_4;-x),\end{aligned}\tag{48}$$

where Π' indicates that $\Gamma(b_j-b_j)$ is absent and an $*$ indicates that the parameter of the type $1+b_j-b_j$ is absent. The function G_4 is obtained by replacing $\exp(i\pi b_j)$ by $\exp(-i\pi b_j)$ in the function G_3.

In the case $x\to\infty$ we can approximate G_3 (cf. Luke[19]):

$$G_3\approx\pi^{1/2}e^{-2(xe^{i\pi})^{1/2}}(xe^{i\pi})^\Theta,\tag{49}$$

where $i^2=-1$, and

$$\Theta=-\tfrac{1}{4}+\tfrac{1}{2}(b_1+b_2+b_3+b_4-a_1-a_2-2).$$

That is,

$$G_3\approx\pi^{1/2}(xe^{i\pi})^\Theta e^{-2ix^{1/2}}.\tag{50}$$

Hence G_3 and G_4, when $x\to\infty$, are approximated by

$$G_3\approx\pi^{1/2}(xe^{+i\pi})^\Theta e^{-2(+i)x^{1/2}},\tag{51}$$

$$G_4\approx\pi^{1/2}(xe^{-i\pi})^\Theta e^{-2(-i)x^{1/2}},\tag{52}$$

where Θ is given in (49).

Equations (38), (39), (44), (46), (48), (51), and (52) complete the explicit representations as well as approximations of all solutions for the cases (3)–(6).

V. EVALUATION OF THE G FUNCTIONS IN CASE 7

Comparing Eqs. (34) and (36) with a G-function differential equation, we find that the parameters of the G function, $G_{p,q}^{m,n}(u|_{b_1,...,b_q}^{a_1,...,a_p})$, are $q=2, p=0, m+n=$ odd integer. Here the independent variable is u and the dependent variables are f_1 and f_2 in Eqs. (34) and (36), respectively.

A. The solution near $u=0$

For $b_1-b_2\neq 0,\pm 1,\pm 2,...$ the two solutions are available from Luke[19] as follows:

$$G_1 = [\Gamma(1+b_1-b_2)]^{-1} u^{b_1}\, {}_0F_1(-;1+b_1-b_2;-u),$$
$$G_2 = [\Gamma(1+b_2-b_1)]^{-1} u^{b_2}\, {}_0F_1(-;1+b_2-b_1;-u),$$
$$f_1 = C_1G_1 + C_2G_2. \tag{53}$$

Thus

$$\begin{aligned}\delta_1 = {} & C_1[u^{b_1}/\Gamma(1+b_1-b_2)] \\ & \times {}_0F_1(-;1+b_1-b_2;-u) \\ & + C_2[u^{b_2}/\Gamma(1+b_2-b_1)] \\ & \times {}_0F_1(-;1+b_2-b_1;-u) \\ & + C_3[u^{b_3}/\Gamma(1+b_3-b_4)] \\ & \times {}_0F_1(-;1+b_3-b_4;-u) \\ & + C_4[u^{b_4}/\Gamma(1+b_4-b_3)] \\ & \times {}_0F_1(-;1+b_4-b_3;-u)\,,\end{aligned} \tag{54}$$

where the C_i's ($i=1,\ldots,4$) are constants, and

$$\begin{aligned} u &= k^2t^\alpha/\alpha^2, \\ b_1 &= \frac{\frac{1}{2}-\eta}{\alpha} - \left[\frac{(\eta-\frac{1}{2})^2}{\alpha^2} + \frac{2}{3\alpha^2}\right]^{1/2}, \\ b_1 &= \frac{\frac{1}{2}-\eta}{\alpha} + \left[\frac{(\eta-\frac{1}{2})^2}{\alpha^2} + \frac{2}{3\alpha^2}\right]^{1/2}, \\ b_3 &= 0, \quad b_4 = -(2\eta-1)/\alpha. \end{aligned} \tag{55}$$

The leading terms in Eq. (54) can be taken as approximations for $u \to 0$.

B. The solution for $u \to \infty$

As before, in this case we can show that the two independent solutions for $b_1 - b_2 \neq 0,1,\ldots$ are G_1 and G_2 and

$$\begin{aligned} f_1 &= C_1G_1 + C_2G_2 \\ &= C_1G^{2,0}_{0,2}(ue^{i\pi}|_{b_1,b_2}) + C_2G^{2,0}_{0,2}(ue^{-i\pi}|_{b_1,b_2}), \end{aligned} \tag{56}$$

where C_1 and C_2 are constants. For example,

$$\begin{aligned} & G^{2,0}_{0,2}(ue^{i\pi}|_{b_1,b_2}) \\ &= \frac{1}{2\pi i}\int_L ds\, \Gamma(b_1+s)\Gamma(b_2+s)(ue^{i\pi})^{-s} \\ &= u^{b_1}e^{i\pi b_1}\Gamma(b_2-b_1)\, {}_0F_1(-;1+b_1-b_2;-u) \\ &\quad + u^{b_2}e^{i\pi b_2}\Gamma(b_1-b_2)\, {}_0F_1(-;1+b_2-b_1;-u). \end{aligned} \tag{57}$$

In a similar manner we get

$$\begin{aligned} & G^{2,0}_{0,2}(ue^{-i\pi}|_{b_1,b_2}) \\ &= u^{b_1}e^{-i\pi b_1}\Gamma(b_2-b_1)\, {}_0F_1(-;1+b_1-b_2;-u) \\ &\quad + u^{b_2}e^{-i\pi b_2}\Gamma(b_1-b_2)\, {}_0F_1(-;1+b_2-b_1;-u). \end{aligned} \tag{58}$$

Thus we have the solution

$$\begin{aligned} \delta_1 = f_1 + f_2 = {} & C_1G^{2,0}_{0,2}(ue^{i\pi}|_{b_1,b_2}) + C_2G^{2,0}_{0,2}(ue^{-i\pi}|_{b_1,b_2}) \\ & + C_3G^{2,0}_{0,2}(ue^{i\pi}|_{b_3,b_4}) \\ & + C_4G^{2,0}_{0,2}(ue^{-i\pi}|_{b_3,b_4}), \end{aligned} \tag{59}$$

where the C_i ($i=1,\ldots,4$) are arbitrary constants and the b_i are those defined in (55).

Using the results for asymptotic expansions for the G function (see Luke[19]) we get the following:

$$G^{2,0}_{0,2}(ue^{i\pi}|_{b_1,b_2}) \approx \pi^{1/2}(ue^{i\pi})^{\Theta}e^{-2(+i)u^{1/2}}, \tag{60}$$
$$G^{2,0}_{0,2}(ue^{-i\pi}|_{b_1,b_2}) \approx \pi^{1/2}(ue^{-i\pi})^{\Theta}e^{-2(-i)u^{1/2}}, \tag{61}$$

where $\Theta = -\frac{1}{4} + \frac{1}{2}(b_1+b_2)$. By using (60) and (61) in Eq. (59), we can get an approximation for δ_1 when $u \to \infty$.

VI. PERTURBATIONS IN A THREE-COMPONENT MEDIUM

For perturbations in a three-component fluid we have the following set of differential equations:

$$\begin{aligned} & \Delta^2\delta_i + (2\eta-1)\Delta\delta_i + k_i^2t^{\alpha_i}\delta_i \\ & \quad = \frac{2}{3}\left(\sum_{i=1}^{3}\Omega_i\delta_i\right), \quad i=1,2,3, \end{aligned} \tag{62}$$

where $\Sigma_i\Omega_i = 1$. The parameters k_i, Ω_i, and α_i are quantities defined analogously as given in Sec. II. Proceeding as in Sec. II, we obtain after straightforward calculation

$$\begin{aligned} & \Delta^4\delta_1 + 2(2\eta-1)\Delta^3\delta_1 \\ & \quad + [k_1^2t^{\alpha_1} + k_3^2t^{\alpha_3} + (2\eta-1)^2 - \tfrac{2}{3}]\Delta^2\delta_1 \\ & \quad + [2k_1^2\alpha_1t^{\alpha_1} + (2\eta-1)(k_1^2t^{\alpha_1} + k_3^2t^{\alpha_3}) \\ & \quad - \tfrac{2}{3}(2\eta-1)]\Delta\delta_1 + [k_1^2\alpha_1^2t^{\alpha_1} + (2\eta-1)k_1^2\alpha_1t^{\alpha_1} \\ & \quad + k_3^2t^{\alpha_3}k_1^2t^{\alpha_1} - \tfrac{2}{3}\Omega_1k_3^2t^{\alpha_3} - \tfrac{2}{3}k_1^2t^{\alpha_1}(\Omega_2+\Omega_3)]\delta_1 \\ & \quad = \tfrac{2}{3}\Omega_2[k_3^2t^{\alpha_3} - k_2^2t^{\alpha_2}]\delta_2 = \tfrac{2}{3}\Omega_2\delta_2 f(k_2,k_3,\alpha_2,\alpha_3). \end{aligned} \tag{63}$$

For the density contrast of the other components δ_2 and δ_3, analogous equations may be obtained by means of cyclic interchanging the indices 1,2,3.

For certain parameter combinations the right-hand side of (63) becomes equal to zero, and hence δ_1 is determined by the fourth-order differential equation. This concerns the following combination of parameters α_i and k_i:

(1) $k_2 = k_3 = 0$, α_i arbitrary,

(2) $\alpha_2 = \alpha_3 = 0$, $k_2 = k_3 \neq 0$,

(3) $\alpha_2 = \alpha_3 \neq 0$, $k_2 = k_3 = 0$.

Imposing the following stronger conditions on the parameters, we can obtain analogous fourth-order equations for δ_2 and δ_3:

(4) $k_i = 0$, α_i arbitrary,

(5) $\alpha_i = 0$, k_i arbitrary,

(6) $\alpha_i = \alpha \neq 0$, $k_i = k \neq 0$.

Then the general solutions follow from the cases considered in Sec. III for a two-fluid medium. An essentially new situation arises in the more complicated case $f(k_2,k_3,\alpha_2,\alpha_3) = \text{const}$. This holds if

(7) $\alpha_2 = 0$, $k_3 = 0$,

(8) $\alpha_3 = 0$, $k_2 = 0$,

(9) $\alpha_2 = \alpha_3 = 0$, $k_2 = k_3 \neq 0$.

Clearly (7) and (8) describe the same physical situation. After straightforward calculations, a sixth-order differential equation for δ_1 can be obtained. In case (7) we introduce the

variable $x = k_1^2 t^\alpha$ as in case (3) of Sec. III, which leads to the following equation:

$$\prod_{i=1}^{6} (\Delta - b_i) + x \prod_{j=1}^{4} (\Delta - a_j) = 0, \tag{64}$$

where $\Delta = \alpha_1 x\, d/dx$. It is of the type of equation satisfied by Meijer's G function $G^{m,n}_{p,q}(x|^{a_1,\ldots,a_p}_{b_1,\ldots,b_q})$ with $p=4$, $q=6$, $(-1)^{p-m-n} = -1$, $a_1,\ldots,a_4$, and $b_1,\ldots,b_6$. The parameters a_i and b_j are the roots of the two algebraic equations,

$$\begin{aligned} &a^4 + a^3[4\alpha_1 + 2\nu] + a^2\left[6\alpha_1^2 + 3\nu\alpha_1 - \tfrac{2}{3}(\Omega_1+\Omega_2)\right. \\ &\quad \left. + 3\alpha_1\nu + \nu^2 + B\right] + a\left[4\alpha_1^3 + 6\nu\alpha_1^2 - \tfrac{4}{3}\alpha_1(\Omega_2+\Omega_3)\right. \\ &\quad \left. + 2\nu^2\alpha_1 - \tfrac{2}{3}\nu(\Omega_2+\Omega_3) + B(2\alpha_1+\nu)\right] + \alpha_1^4 + 2\nu\alpha_1^3 \\ &\quad + \nu^2\alpha_1^2 - \tfrac{2}{3}\nu\alpha_1(\Omega_2+\Omega_3) + B(\alpha_1^2 + \nu\alpha_1 - \tfrac{2}{3}\Omega_3) = 0, \end{aligned} \tag{65}$$

and

$$\begin{aligned} &b^6 + 3\nu b^5 + (3\nu^2 - \tfrac{2}{3} + B)b^4 + (\nu^3 - \tfrac{4}{3}\nu + 2B\nu)b^3 \\ &\quad + \left[\nu^2(B - \tfrac{2}{3}) - \tfrac{2}{3}B(1-\Omega_2)\right]b^2 \\ &\quad - \tfrac{2}{3}\nu B(1-\Omega_2)b = 0, \end{aligned} \tag{66}$$

where $B = k_2^2$ and $\nu = 2\eta - 1$. These equations can be rewritten in a more convenient form:

$$\begin{aligned} &[a^2 + (2\alpha_1+\nu)a + \alpha_1(\alpha_1+\nu) + B][a^2 + (2\alpha_1+\nu)\alpha \\ &\quad + \alpha_1(\alpha_1+\nu) - \tfrac{2}{3}(\Omega_2+\Omega_3)] + \tfrac{2}{3}\Omega_2 B = 0, \end{aligned} \tag{67}$$

and

$$\begin{aligned} &b(b^2 + \nu b - \tfrac{2}{3})(b+\nu)(b^2 + \nu b + B) \\ &\quad + \tfrac{2}{3}\Omega_2 B b(b+\nu) = 0. \end{aligned} \tag{68}$$

The explicit solutions read

$$\begin{aligned} &b_{1,2,3,4} = -\nu/2 \pm [\nu^2/4 + Q_{1,2}]^{1/2}, \\ &b_5 = -\nu, \quad b_6 = 0, \\ &Q_{1,2} = [-\tfrac{1}{2}(B - \tfrac{2}{3})] \pm [\tfrac{1}{4}(B-\tfrac{2}{3})^2 + \tfrac{2}{3}B(1-\Omega_2)]^{1/2}, \end{aligned} \tag{69}$$

and

$$\begin{aligned} &a_{1,2,3,4} = [-(\alpha_1 + \nu/2)] \pm [\nu/4 + P_{1,2}]^{1/2}, \\ &P_{1,2} = [\tfrac{1}{3}(\Omega_2+\Omega_3) - B/2] \\ &\qquad \pm [(B/2 - \tfrac{1}{3}(\Omega_2+\Omega_3))^2 + \tfrac{2}{3}B\Omega_2]^{1/2}. \end{aligned} \tag{70}$$

For case (9) we also obtain a differential equation of the form (64), where the a_i and b_j are roots of the algebraic equations

$$\begin{aligned} &[a^2 + (2\alpha_1+\nu)\alpha + \alpha_1(\alpha_1+\nu) + B + C] \\ &\quad \times [a^2 + (2\alpha_1+\nu)\alpha + \alpha_1(\alpha_1+\nu) - \tfrac{2}{3}(\Omega_2+\Omega_3)] \\ &\quad + \tfrac{2}{3}\Omega_2(B-C) = 0, \end{aligned} \tag{71}$$

and

$$\begin{aligned} &b^6 + 3\nu b^5 + (3\nu^2 - \tfrac{2}{3} + B + C)b^4 \\ &\quad + [\nu^2 - 2(\tfrac{2}{3} - B - C)]b^3\nu \\ &\quad + [(B + C - \tfrac{2}{3})\nu^2 + BC - \tfrac{2}{3}B(\Omega_1+\Omega_3) \\ &\quad - \tfrac{2}{3}C(\Omega_1+\Omega_2)]b^2 \\ &\quad + [BC - \tfrac{2}{3}B(\Omega_1+\Omega_3) - \tfrac{2}{3}C(\Omega_1+\Omega_2)]b\nu \\ &\quad - \tfrac{2}{3}\Omega_1 BC = 0, \end{aligned}$$

where $C = k_3^2$, and B and ν are defined in Eq. (66). The last equation can be written in a more convenient form,

$$\begin{aligned} &(b^2 + b\nu)(b^2 + b\nu - \tfrac{2}{3})(b^2 + b\nu + B + C) - (b^2 + b\nu) \\ &\quad \times [BC - \tfrac{2}{3}B(\Omega_1+\Omega_3) - \tfrac{2}{3}C(\Omega_1+\Omega_2)] \\ &\quad - \tfrac{2}{3}\Omega_1 BC = 0. \end{aligned} \tag{72}$$

From Eq. (71) we get explicit expressions for the solutions,

$$\begin{aligned} &a_{1,2,3,4} = [-(\alpha_1 + \nu/2)] \pm [\nu^2/4 + P_{1,2}]^{1/2}, \\ &P_{1,2} = [\tfrac{1}{3}(\Omega_2+\Omega_3) - B/2] \pm [(B/2 - \tfrac{1}{3}(\Omega_2+\Omega_3))^2 \\ &\quad + \tfrac{2}{3}\Omega_2(B-C)]^{1/2}. \end{aligned} \tag{73}$$

Substituting $Q = b^2 + b\nu$ in Eq. (72), we get instead of the sixth-order equation a cubic one, that is,

$$\begin{aligned} &Q(Q - \tfrac{2}{3})(Q + B + C) + Q[BC - \tfrac{2}{3}B(\Omega_1+\Omega_3) \\ &\quad - \tfrac{2}{3}C(\Omega_1+\Omega_2)] - \tfrac{2}{3}\Omega_1 BC = 0. \end{aligned} \tag{74}$$

One gets the solutions for the b_i's in terms of the solutions for $Q = b^2 + b\nu$,

$$b_{1,2,3,4,5,6} = -\nu/2 \pm [\nu^2/4 + Q_{1,2,3}]^{1/2}. \tag{75}$$

The asymptotic behavior can be obtained by using the known properties of the corresponding G functions as it was demonstrated in Secs. IV and V, respectively.

VII. DISCUSSION

The exact solutions obtained in a systematic way using Meijer's G function will be discussed now more explicitly. Attention will be drawn to the asymptotic behavior for small and large values of the time t and to the decision in which cases growing more solutions exist. To this end we consider the seven cases given in Sec. III for the two-fluid problem systematically. After that we discuss shortly the new solutions for the three-fluid problem.

Note that all considerations provided above are given in terms of the arbitrary expansion parameter η, describing powerlike behavior of the scale factor $a(t)$ given in Eq. (6). Besides the case of exact solution $a(t) \propto t^{2/3}$ this can be used to get information for epochs approximately described by a power law $a(t) \propto t^\eta$. In this case the numerical factor equal to $\frac{2}{3}$ appearing in many of the formulas has to be replaced by $\frac{2}{3}\eta^2$.

A. The two-component medium solutions

(1) $k_1 = k_2 = 0$. This corresponds to the gravitational interaction of two dustlike components. The solution was given by Wasserman.[9] From (18) and (19) we obtain immediately

$$\begin{aligned} &\delta_1 = C_{11} + C_{12}t^{-1/3} + C_{13}t^{-1} + C_{14}t^{2/3}, \\ &\delta_2 = C_{21} + C_{22}t^{-1/3} + C_{23}t^{-1} + C_{14}t^{2/3}. \end{aligned} \tag{76}$$

The constants of integration C_{ij} are restricted by the conditions

$$\begin{aligned} &\Omega_1 C_{11} + \Omega_2 C_{21} = \Omega_1 C_{12} + \Omega_2 C_{22} = 0, \\ &C_{13} = C_{23}, \quad C_{14} = C_{24}. \end{aligned}$$

There exists only one growing solution proportional to $t^{2/3}$.

(2) $\alpha_1 = \alpha_2 = 0$, k_1, k_2 arbitrary. This case describes the gravitational interaction of two fluids; the equation of state is given by (7) with $\gamma_1 = \gamma_2 = \frac{4}{3}$. This case is also described by

Fargion.[12] The solution is given by (21) and (22). The constants occurring in the general solution (21) must fulfill some relations as in case (1). A consideration whether an unstable solution (i.e., an $a_i > 0$) exists can be provided straightforwardly using Eq. (22) and the rule of Descartes. For $\eta = \frac{2}{3}$ Eq. (22) reads

$$a^4 + \tfrac{2}{3}a^3 + (k_1^2 - k_2^2 - \tfrac{8}{9})a^2 + \tfrac{1}{3}(k_1^2 + k_2^2 - \tfrac{2}{3})\alpha + (k_1^2 k_2^2 - \tfrac{2}{3}\Omega_2 k_1^2 - \tfrac{2}{3}\Omega_1 k_2^2) = 0. \quad (77)$$

It is shown that for

$$k_1^2 k_2^2 - \tfrac{2}{3}\Omega_2 k_1^2 - \tfrac{2}{3}k_2^2\Omega_1 < 0 \quad (78)$$

exactly one change of the signs of the coefficients before the powers of α takes place. Therefore, if (78) is fulfilled, only one growing mode exists.

For $\gamma = \frac{4}{3}$ the parameters $k_{1,2}^2 = \beta_{1,2}^2 k^2 t_{1,2}^2$ does not depend on the time t, and hence also condition (78) is time independent. The stability criterion (78) generalizes the corresponding one-fluid conditions

$$k_2^2 < \tfrac{2}{3}\Omega_2 \quad \text{and} \quad k_1^2 < \tfrac{2}{3}\Omega_1.$$

If one of them is fulfilled, the inequality (78) is fulfilled automatically. For larger k_1 and k_2 the solution behaves like an acoustic wave. The exact solution of (77) can be obtained writing it as

$$(Q - \tfrac{2}{3})Q + (k_1^2 + k_2^2)Q + (k_1^2 k_2^2 - \tfrac{2}{3}k_1^2\Omega_2 - \tfrac{2}{3}k_2^2\Omega_1) = 0,$$

where $Q = a^2 + a/3$. Then we get immediately

$$Q_{1,2} = \left[\tfrac{1}{2}(k_1^2 + k_2^2 - \tfrac{2}{3})\right] \pm \left[\tfrac{1}{4}(k_1^2 + k_2^2 - \tfrac{2}{3})^2 - k_1^2 k_2^2 + \tfrac{2}{3}k_1^2\Omega_2 + \tfrac{2}{3}k_2^2\Omega_1\right]^{1/2},$$

and

$$a_{1,2,3,4} = -\nu/2 \pm [\nu^2/4 + Q_{1,2}]^{1/2}.$$

(3) $\alpha_2 \neq 0$, $k_1 = 0$.

(4) $\alpha_1 \neq 0$, $k_2 = 0$. Case (4) describes the evolution of the density contrast for two gravitationally coupled fluids: one of them consisted of dust ($k_2 = 0$) and the other one was characterized by the equation of state (7) with arbitrary γ_1. The parameters for the corresponding solution δ_1 are given by (25). Case (3) can be used to obtain the solution for δ_2 in case (4) and vice versa.

The parameters for the solutions δ_1 and δ_2 are, respectively,

$$\delta_{1,2}: b_1 = 0, \quad b_2 = 1/3\alpha_1, \quad b_3 = 2/3\alpha_1, \quad b_4 = -1/\alpha_1,$$
$$\delta_1: a_{1,2}^{(1)} = 1 - (1/6\alpha_1)(1 \mp \omega), \quad \omega = (1 + 24\Omega_2)^{1/2},$$
$$\delta_2: a_{1,2}^{(2)} = 1 + a_{1,2}^{(1)}. \quad (79)$$

The general solution for δ_1 (and analogously for δ_2) reads

$$\delta_i = C_{i1}G_1 + C_{i2}G_2 + C_{i3}G_3 + C_{i4}G_4, \quad i = 1,2, \quad (80)$$

where the G_j are the G functions of Eqs. (38) and (40), respectively. To determine the behavior of the asymptotic case $x \approx 0$ we use (38), then $\delta_{1,2} \propto x^{b_j} \propto t^{\alpha_1 b_j}$, and finally we get the asymptotic solutions ($x \approx 0$),

$$G_1 \propto t^0, \quad G_2 \propto t^{-1/3}, \quad G_3 \propto t^{2/3}, \quad G_4 \propto t^{-1}. \quad (81)$$

The coefficients in Eq. (80) depend on α_1 and Ω_2; they can easily be calculated from (39), e.g., for δ_1, we have

$$G_1 = \frac{\Gamma(1 + (1/6\alpha_1)(1 - \omega))\Gamma(1 + (1/6\alpha_1)(1 + \omega))}{\Gamma(1)\Gamma(1 + 1/3\alpha_1)\Gamma(1 - 1/3\alpha_1)\Gamma(-1/\alpha_1)}. \quad (82)$$

For $\alpha_1 < 0$ (81) corresponds to the asymptote $t \to \infty$ and for $\alpha_1 > 0$ to the asymptote $t \approx 0$, respectively. The considered asymptotic case reproduces the solution for dustlike fluids, and there exists only one growing mode.

The asymptotic behavior for $x \to \infty$ (which corresponds for $\alpha_1 > 0$ to $t \to \infty$ and for $\alpha_1 < 0$ to $t \approx 0$) is given by (46) and (51) and (52):

$$\delta_1: \begin{matrix} G_1 \propto x^{-1-(1/6\alpha_1)(1-\omega)} \propto t^{-\alpha_1 - (1/6)(1-\omega)}, \\ G_2 \propto x^{-1-(1/6\alpha_1)(1+\omega)} \propto t^{-\alpha_1 - (1/6)(1+\omega)}, \end{matrix} \quad (83)$$

$$\delta_2: \begin{matrix} G_1 \propto x^{-(1/6\alpha_1)(1-\omega)} \propto t^{-(1/6)(1-\omega)}, \\ G_2 \propto x^{-(1/6\alpha_1)(1+\omega)} \propto t^{-(1/6)(1+\omega)}, \end{matrix} \quad (84)$$

i.e., it is $\delta_1 \propto (t/t_0)^{-\alpha_1}\delta_2$.

For $\Omega_2 > 0$ a growing mode solution (84) exists. A growing solution (83) for δ_1 exists also, independent on α_1; namely, if $\alpha_1 < (\omega - 1)/6$, i.e., $\gamma_1 > \frac{4}{3} - \frac{1}{12}(\omega - 1)$. Using (51) and (52) we obtain, with

$$\Theta = -\frac{1}{4} + \frac{1}{2}\left(\sum_i b_i - \sum_i a_i - 2\right),$$

and

$$\sum_i b_i = -\frac{2}{3\alpha_1}, \quad \sum_i a_i^{(1)} = -2 - \frac{1}{3\alpha_1}, \quad \sum_i a_i^{(2)} = -\frac{1}{3\alpha_1},$$

the expressions

$$\Theta = (\gamma_1 - \tfrac{5}{3})/2\alpha_1, \quad \Theta_2 = \Theta_1 - 1 \quad (85)$$

(see also Soloveva and Starobinski[14]), and

$$\delta_1 \propto G_{3,4} \approx \pi^{1/2}\left(\frac{k_1^2}{\alpha_1^2}\right)^{\Theta_1}\left(\frac{t}{t_0}\right)^{(1/2)(\gamma_1 - 5/3)} \times \exp\left\{\pm i\pi\left[\Theta_1 + \frac{2}{\pi}\frac{k_1}{\alpha_1}\left(\frac{t}{t_0}\right)^{\alpha_1/2}\right]\right\}, \quad (86)$$

$$\delta_2 \propto G_{3,4} \propto \delta_1\left(\frac{t}{t_0}\right)^{-\alpha_1}. \quad (87)$$

As mentioned by Soloveva and Starobinski,[14] for $\gamma_1 = \frac{5}{3}$ the amplitude of the acoustic waves remains constant, whereas for $\gamma_1 > \frac{5}{3}$ the amplitudes grow proportionally to $t^{(\gamma_1 - 5/3)/2}$. Although a growing mode for the first fluid in most all cases exists [see (83)], the comparison with (81) shows that only the mode proportional to G_3 can survive during the further evolution. The moment at which this mode goes over from the oscillatory regime to monotonic growth is determined by

$$|(2k_1/\alpha_1)(t/t_0)^{\alpha_1/2}| < 2. \quad (88)$$

For $\gamma_1 = \frac{5}{3}$ this happens at $t_1 = (3k_1)^3 t_0$. The quantity t_1 is the Jeans time for a single component fluid with $\gamma = \frac{5}{3}$.

(5) $\alpha_1 = 0$, $\alpha_2 \neq 0$.

(6) $\alpha_2 = 0$, $\alpha_1 \neq 0$. The solutions for δ_1 are given by (29) and (30). Solving case (6) for δ_1 we can obtain the corresponding solution for δ_2 using case (5), and vice versa.

Cases (5) and (6) have been considered by Soloveva and Starobinski.[14] For discussion of the asymptotic behavior we can avoid the explicit solution of (29) and use the rule of Descartes to determine under which conditions unstable solutions occur, i.e., if $b_j > 0$. Indeed, for $x \approx 0$ the asymptotic solution has the form $\delta_1 \propto x^{b_j} \propto t^{\alpha_2 b_j}$ because of the used variable $x(t)$. Equation (29) can be written in the form

$$\alpha_2^4 b^4 + 2\nu\alpha_2^3 b^3 + (k_1^2 + \nu^2 - \tfrac{2}{3})\alpha_2^2 b^2 + (-\tfrac{2}{3} + k_1^2)\nu\alpha_2 b - \tfrac{2}{3}\Omega_2 k_1^2 = 0 . \quad (89)$$

It is shown that for arbitrary k_1^2 the sign of the coefficients changes only once. Therefore, exactly one solution with positive exponent $\alpha_2 b_j > 0$ exists. The exact solution for $\alpha_2 b$ follows as in case (2): With

$$Q_{1,2} = \left[-\tfrac{1}{2}(k_1^2 - \tfrac{2}{3}) \right] \pm \left[\tfrac{1}{4}(k_1^2 - \tfrac{2}{3})^2 + \tfrac{2}{3}k_1^2\Omega_2 \right]^{1/2},$$

we obtain

$$(\alpha_2 b)_{1,2,3,4} = -\nu/2 \pm \left[\nu^2/4 + Q_{1,2} \right]^{1/2}.$$

In the region $x \to \infty$ the $a_1^{(1)}$ and $a_2^{(1)}$ become important. With

$$\tilde{\omega} = (1 + 24\Omega_1 - 36k_1^2)^{1/2},$$

we obtain

$$\delta_1 \propto t^{-(1/6)(1 \mp \tilde{\omega})} + \text{acoustic waves.} \quad (90)$$

For

$$k_1^2 < \tfrac{2}{3}\Omega_1 , \quad (91)$$

a growing mode exists [i.e., Eq. (78) represents the Jeans criterion for this case]. A similar condition holds in case (2).

For δ_2 in the case $x \approx 0$ the same growing mode as for δ_1 exists, and for the case $x \to \infty$ we obtain

$$\delta_2 \propto (t/t_0)^{-\alpha_2}\delta_1 + \text{acoustic waves.} \quad (92)$$

A growing mode exists if $\gamma > \frac{4}{3} + \frac{1}{12}(1 - \tilde{\omega})$.

(7) $\alpha_2 = \alpha_2 = \alpha \neq 0$, $k_1 = k_2 = k \neq 0$. This case was considered in more detail by Fargion.[12] For δ_1 we obtain in the asymptotic case $u \approx 0$ the dust asymptote (62),

$$\delta_1 \propto u^{b_j} \propto t^{ab_j}, \quad (93)$$

where the b_j are the same ones as for cases (3) and (4), respectively. For $u \to \infty$ (i.e., for $\alpha < 0$ this corresponds to $t \approx 0$) we get for the parameter

$$\Theta(b_1, b_2) = (\tfrac{1}{2} - \eta)/\alpha = -\tfrac{1}{4} - 1/6\alpha .$$

This corresponds to the case of Eqs. (85) and (86), respectively. For $\Theta(b_3, b_4)$ we get with $\Theta = -\frac{1}{4} + (\frac{1}{2} - \eta)/\alpha$ the same result. Thus only acoustic wave solutions exist, which go over into a monotonically growing solution if (88) is satisfied, i.e., the usual Jeans criterion.

B. The three-component medium solutions

For the three-component medium, solutions are given for special cases and in most cases only for the δ_1 perturbation. For cases (4)–(6) in Sec. VI a general solution for all three components has been obtained. However, in this case the equations degenerate to fourth-order differential equations as for a two-fluid medium with the solutions described above. Essentially new solutions are obtained under the assumptions in cases (7)–(9) in Sec. VI. In these cases the behavior of the perturbation δ_1 will also be described by G functions.

Cases (7) and (8) in Sec. VI correspond to two different kinds of dark matter (a dustlike and a relativistic one) and a component with arbitrary γ. Using (69) we can immediately see that in the asymptotic region $x \approx 0$ (for $\alpha_2 < 0$ this corresponds to $t \to \infty$) exactly one growing mode exists independent of the other parameters.

The existence of a growing mode for the opposite asymptotic case $x \to \infty$ is connected with the condition

$$\nu^2/4 + P_{1,2} > (\alpha_1 + \nu/2)^2 , \quad (94)$$

where $P_{1,2}$ is given in (70).

For $\nu = 2\eta - 1 = \frac{1}{3}$ and $\alpha_1 = -\frac{2}{3}$ (i.e., $\gamma = \frac{5}{3}$) this condition reads

$$P_{1,2} > \tfrac{2}{9}, \quad B(1 - 3\Omega_2) + \tfrac{2}{9} < \tfrac{2}{3}(\Omega_2 + \Omega_3) , \quad (95)$$

where B is given in (66). Case (9) in Sec. VI describes two relativistic fluids coupled to a component with arbitrary γ_1 and β_1. A growing mode exists for $x \approx 0$ if $Q > 0$, where Q is explained in (74). In the opposite asymptotic case $x \to \infty$ a growing mode exists if

$$P_{1,2} > \alpha_1^2 + \alpha_1\nu .$$

For $\nu = \frac{1}{3}$, $\alpha_1 = -\frac{2}{3}$ we obtain the condition

$$B(1 - 3\Omega_2) + 3\Omega_2 C + \tfrac{2}{9} < \tfrac{2}{3}(\Omega_2 + \Omega_3) , \quad (96)$$

where C is given in (72). These conditions generalize the Jeans criterion to the three-component cosmological medium.

[1] J. H. Jeans, *Astronomy and Cosmology* (Cambridge U.P., Cambridge, 1929).
[2] E. M. Lifshitz, J. Phys. (Moscow) **10**, 116 (1946).
[3] W. B. Bonnor, Mon. Not. R. Astron. Soc. **117**, 104 (1957).
[4] S. Weinberg, *Gravitation and Cosmology* (Wiley, New York, 1972).
[5] Ya. B. Zeldovich and I. D. Novikov, *The Structure and Evolution of the Universe, Relativistic Astrophysics* (Nauka, Moscow, 1975; University of Chicago Press, Chicago, 1983), Vol. 2.
[6] P. J. E. Peebles, *The Large-scale Structure of the Universe* (Princeton U.P., Princeton, NJ, 1980).
[7] S. Gottlöber, H. J. Haubold, J. P. Mücket, and V. Müller, *Modern Problems in Cosmology*, edited by A. M. Mathai (Centre for Mathematical Sciences, Trivandrum, India, 1987).
[8] S. F. Shandarin, A. G. Doroshkewich, and Ya. B. Zeldovich, Usp. Fiz. Nauk **139**, 83 (1983) [Sov. Phys. Usp. **26**, 46 (1983)].
[9] I. Wasserman, Astrophys. J. **248**, 1 (1981).
[10] V. L. Polyachenko and A. M. Fridman, Zh. Eksp. Teor. Fiz. **81**, 13 (1981) [Sov. Phys. JETP **54**, 7 (1981)].
[11] L. P. Grishchuk and Ya. B. Zeldovich, Astron. Zh. **58**, 472 (1981) [Sov. Astron. **25**, 267 (1981)].
[12] D. Fargion, Nuovo Cimento B **77**, 111 (1983).
[13] L. V. Soloveva and I. S. Nurgaliev, Astron. Zh. **62**, 459 (1985) [Sov. Astron. **29**, 267 (1985)].
[14] L. V. Soloveva and A. A. Starobinski, Astron. Zh. **62**, 625 (1985) [Sov. Astron. **29**, 367 (1985)].
[15] A. M. Mathai and H. J. Haubold, *Modern Problems in Nuclear and Neutrino Astrophysics* (Akademie, Berlin, 1988).
[16] H. J. Haubold, J. P. Mücket, and A. Tornambè, Ann. Phys. (Leipzig) **44**, 519 (1987).
[17] J. P. Mücket, H. J. Haubold, V. Müller, S. Gottlöber, and A. M. Mathai, "Coupling of modes in Newtonian cosmology and the formation of primordial objects," McGill University, Montreal, Department of Mathematics and Statistics, Preprint No. 88-1, 1988.
[18] C. S. Meijer, Proc. Nederl. Akad. Wetensch. A **49**, 344 (1946).
[19] Y. L. Luke, *The Special Functions and Their Approximations* (Academic, New York, 1969), Vol. I.
[20] A. M. Mathai and R. K. Saxena, *Generalized Hypergeometric Functions with Applications in Statistics and Physical Sciences, Lecture Notes in Mathematics*, Vol. 348 (Springer, Berlin, 1973).

The Teukolsky–Starobinsky identities in type D vacuum backgrounds with cosmological constant

G. F. Torres del Castillo
Departamento de Física Matemática, Instituto de Ciencias de la Universidad Autónoma de Puebla, 72000 Puebla, Mexico and Departamento de Física, Centro de Investigación y de Estudios Avanzados del Instituto Politécnico Nacional, Apartado Postal 14-740, 07000 México, D. F., Mexico

(Received 9 December 1987; accepted for publication 4 May 1988)

It is shown that in all the type D solutions to the Einstein vacuum field equations with cosmological constant, the one-variable functions appearing in the separable maximal spin-weight components of the neutrino, electromagnetic, and gravitational perturbations are related by certain differential operators and the corresponding proportionality constants are obtained. It is also shown that analogous relations hold in the case of perturbations by a Rarita–Schwinger field if the cosmological constant vanishes.

I. INTRODUCTION

Several interesting and unexpected problems have emerged from investigations involving type D solutions of the Einstein field equations. In particular, the remarkable properties found in the studies associated with the Kerr metric have led to the discovery of similar regularities for the whole class of the type D vacuum space-times and, in some cases, the underlying structure responsible for these properties has also been found.

It has been shown that, just like for the Kerr metric,[1,2] in all the type D vacuum space-times with cosmological constant the maximal spin-weight components of a Weyl neutrino field, an electromagnetic field, or a gravitational perturbation satisfy decoupled (second-order) equations that can be solved by separation of variables.[3,4] Similarly, Güven's analysis of the perturbations of the Kerr metric by a spin-$\frac{3}{2}$ field in the framework of linearized supergravity[5] has been extended by Kamran[6] to the whole class of space-times mentioned above; the spin-$\frac{3}{2}$ perturbations are determined (modulo supersymmetry transformations) by two (maximal spin-weight) components that obey decoupled equations also solvable by separation of variables.

In the case of the electromagnetic and the gravitational perturbations of the Kerr metric, it was discovered that there exist differential relations between the one-variable functions contained in the opposite maximal spin-weight components of the perturbations.[7,8] The existence of such relations, called the Teukolsky–Starobinsky identities,[2] turns out to be related to that of a two-index Killing spinor.[9,10]

The aim of this paper is to show explicitly that in all the type D vacuum space-times with cosmological constant, relations similar to the Teukolsky–Starobinsky identities hold. Furthermore, we show that analogous relations apply to the case of the spin-$\frac{3}{2}$ perturbations if the cosmological constant is set equal to zero. In the case of the Weyl neutrino perturbations the analog of the Teukolsky–Starobinsky identities follows directly from the Weyl neutrino equation.

In Sec. II some basic relations and the decoupled equations for the perturbations are given. In Secs. III and IV we make use of the explicit expressions for the type D vacuum metrics given in Ref. 11 (see also Ref. 12) obtaining the ordinary differential equations satisfied by the separated functions, from which the Teukolsky–Starobinsky identities are derived. In Sec. V the proportionality constants involved in the Teukolsky–Starobinsky identities are evaluated and from the equality of some of these constants certain identities relating the functions contained in the metric are obtained. We employ the Newman–Penrose notation and the spinor formalism following the conventions of Ref. 13.

II. PERTURBATIONS IN TYPE D VACUUM BACKGROUNDS

In a type D space-time one can choose a spin frame $\{o_A, \iota_A\}$ (with $o_A \iota^A = 1$) such that o_A and ι_A are (double) principal spinors of the Weyl spinor. Then Ψ_2 is the only nonvanishing component of Ψ_{ABCD} and, assuming that the Einstein vacuum field equations with cosmological constant are satisfied, it follows from the Bianchi identities that

$$\Psi_2 = -M\phi^3 , \tag{1}$$

where M is a constant and ϕ is a function such that

$$\begin{aligned} &\rho = D\ln\phi, \quad \tau = \delta\ln\phi, \\ &\pi = -\bar{\delta}\ln\phi, \quad \mu = -\Delta\ln\phi , \end{aligned} \tag{2}$$

and

$$\kappa = \sigma = \lambda = \nu = 0 . \tag{3}$$

Owing to the assumptions stated above, the remaining spin coefficients can be expressed (locally) as

$$\begin{aligned} &\epsilon = D\ln\zeta, \quad \beta = \delta\ln\zeta, \\ &\alpha = -\bar{\delta}\ln\xi, \quad \gamma = -\Delta\ln\xi , \end{aligned} \tag{4}$$

where ζ and ξ are some functions. The transformation properties of the spin coefficients under the spin transformations that preserve the direction of o_A and ι_A can be accounted for by assigning to ϕ, ζ, and ξ the weights $\{0,0\}$, $\{1,0\}$, and $\{-1,0\}$, respectively, in the sense of Geroch *et al.*[14] [The functions ϕ, ζ, and ξ are not defined uniquely by (2) and (4), however; ϕ is defined modulo a constant factor and ζ and ξ are defined up to multiplicative functions of two variables.]

The massless field equations and the equations for the gravitational perturbations consist of sets of differential

 0022-2488/88/092078-06$02.50

equations in which the components of the fields are coupled with each other and with the background geometry. Nevertheless, in a type D vacuum space-time the components obtained by fully contracting these fields with each of the principal spinors of the conformal curvature satisfy decoupled equations that can be solved by separation of variables. If we denote by $\chi_{1/2}$ and $\chi_{-1/2}$ the components η_0 and η_1 of a Weyl neutrino field, by χ_1 and χ_{-1} the components φ_0 and φ_2 of an electromagnetic field, and by χ_2 and χ_{-2} the first-order perturbations of the components Ψ_0 and Ψ_4 of the Weyl spinor, then from the corresponding field equations it follows that[1]

$$\begin{aligned}&[(D-(2s-1)\epsilon+\bar{\epsilon}-2s\rho-\bar{\rho})(\Delta-2s\gamma+\mu)\\&\quad-(\delta-(2s-1)\beta-\bar{\alpha}-2s\tau+\bar{\pi})(\bar{\delta}-2s\alpha+\pi)\\&\quad-(s-1)(2s-1)\Psi_2]\chi_s=0\end{aligned}\tag{5}$$

and

$$\begin{aligned}&[(\Delta+(2s-1)\gamma-\bar{\gamma}+2s\mu+\bar{\mu})(D+2s\epsilon-\rho)\\&\quad-(\bar{\delta}+(2s-1)\alpha+\bar{\beta}+2s\pi-\bar{\tau})(\delta+2s\beta-\tau)\\&\quad-(s-1)(2s-1)\Psi_2]\chi_{-s}=0\,.\end{aligned}\tag{6}$$

Equations (5) and (6) are also valid when $s=0$; they coincide with the (conformally invariant) spin-0 massless field equation

$$(\nabla^{\mu}\nabla_{\mu}+R/6)\chi=0\,,\tag{7}$$

with $\chi=\chi_0=\chi_{-0}$. [The equivalence of (7) with (5) and (6) depends on the fact that we are assuming $\kappa\nu-\sigma\lambda=0$.] Furthermore, when $s=\frac{3}{2}$, Eqs. (5) and (6) are precisely the equations satisfied by $\nabla_0{}^{C'}\psi_{00C'}$ and $\nabla_1{}^{C'}\psi_{11C'}$, where $\psi_{ABC'}$ are the components of a Rarita–Schwinger field, provided that the Ricci tensor vanishes. These derivatives of $\psi_{ABC'}$ are invariant under supersymmetry transformations and correspond to the quantities $\Psi_2 b_2$ and $\Psi_2 a_1$ introduced by Güven,[5] which determine the solutions of the Rarita–Schwinger equation that cannot be generated from a vacuum by supersymmetry transformations (see also Ref. 15).

All the type D solutions to the Einstein vacuum field equations with cosmological constant have been given in Refs. 11 and 12. Each of these metrics possesses (at least) two commuting Killing vectors and therefore, in all cases, there exist local coordinates $\{x,y,u,v\}$ such that ∂_u and ∂_v are Killing vectors. The orbits under the isometry group generated by ∂_u and ∂_v are called null or non-null according to whether $g_{uu}g_{vv}-(g_{uv})^2$ vanishes or not. In Secs. III and IV these two cases are treated separately, summarizing the relevant information about the metrics. In each case, a general form of the null tetrad is given and Eqs. (5) and (6) are separated. The resulting ordinary differential equations are then used to find the differential relations between the separated functions corresponding to opposite helicities.

III. NON-NULL ORBIT METRICS

As mentioned above, in all cases of the type D vacuum metrics there exist coordinates $\{x,y,u,v\}$ such that ∂_u and ∂_v are Killing vectors, therefore Eqs. (5) and (6) admit separable solutions with a dependence in the ignorable coordinates u and v of the form

$$e^{i(ku+lv)}\,,\tag{8}$$

where k and l are constants. We shall choose the null tetrad in such a way that, in all non-null orbit type D vacuum metrics, on quantities with a dependence of the form (8) the tetrad vectors can be replaced according to

$$\begin{aligned}&D\to\mathscr{D}_0,\quad \Delta\to-\tfrac{1}{2}\phi\bar{\phi}Q\mathscr{D}_0^{\dagger},\\&\delta\to(1/\sqrt{2})\bar{\phi}\mathscr{L}_0^{\dagger},\quad \bar{\delta}\to(1/\sqrt{2})\phi\mathscr{L}_0,\end{aligned}\tag{9}$$

where

$$\begin{aligned}&\mathscr{D}_n\equiv\partial_y+i(q/Q)+nQ^{(1)}/Q=Q^{-n}\mathscr{D}_0Q^n\,,\\&\mathscr{D}_0^{\dagger}\equiv\partial_y-i(q/Q)+nQ^{(1)}/Q=Q^{-n}\mathscr{D}_0^{\dagger}Q^n\,,\\&\mathscr{L}_n\equiv\sqrt{P}\left(\partial_x+\frac{p}{P}+\frac{n}{2}\frac{P^{(1)}}{P}\right)=P^{-n/2}\mathscr{L}_0P^{n/2}\,,\\&\mathscr{L}_n^{\dagger}\equiv\sqrt{P}\left(\partial_x-\frac{p}{P}+\frac{n}{2}\frac{P^{(1)}}{P}\right)=P^{-n/2}\mathscr{L}_0^{\dagger}P^{n/2}\,,\end{aligned}\tag{10}$$

$q(y)$ and $p(x)$ are polynomials of degree not greater than 2 that contain the separation constants k and l, and $Q(y)$ and $P(x)$ are polynomials of degree not greater than 4 that contain the cosmological constant λ_0 and other arbitrary parameters present in the metric. The polynomials p, q, P, and Q are listed in Table I, following Ref. 11 with some minor changes in notation. The spin coefficients and the conformal curvature for each metric given in Table I can be obtained by means of Eqs. (1)–(4) with the aid of Table II. The asymmetry between D and Δ, given in (9), introduces certain asymmetry in the expressions for χ_s and χ_{-s} in terms of the separated functions. As remarked in Refs. 4, 6, and 10, any other null tetrad satisfying Eq. (3) would lead to results equivalent to those obtained by means of (9).

Using Eqs. (1), (2), (4), (9), and (10) and the expressions given in Tables I and II one finds that the solutions to Eq. (5), with a dependence of the form (8), are given in terms of the solutions of the ordinary differential equations

$$\begin{aligned}&[Q\mathscr{D}_{1-s}\mathscr{D}_0^{\dagger}-(2s-1)iq^{(1)}\\&\quad+(s-1)(2s-1)Q^{(2)}/6]Q^sR_{+s}=A_sQ^sR_{+s}\end{aligned}\tag{11}$$

and

$$\begin{aligned}&[\mathscr{L}_{1-s}^{\dagger}\mathscr{L}_s+(2s-1)p^{(1)}+(s-1)(2s-1)P^{(2)}/6]S_{+s}\\&\quad=-A_sS_{+s}\,,\end{aligned}\tag{12}$$

where A_s is a separation constant and s can take the values 0, $\frac{1}{2}$, 1, $\frac{3}{2}$, and 2. Similarly, the separable solutions of Eq. (6) are given in terms of the functions $R_{-s}(y)$ and $S_{-s}(x)$, which satisfy

$$\begin{aligned}&[Q\mathscr{D}_{1-s}^{\dagger}\mathscr{D}_0+(2s-1)iq^{(1)}\\&\quad+(s-1)(2s-1)Q^{(2)}/6]R_{-s}=A_{-s}R_{-s}\end{aligned}\tag{13}$$

and

$$\begin{aligned}&[\mathscr{L}_{1-s}\mathscr{L}_s^{\dagger}-(2s-1)p^{(1)}+(s-1)(2s-1)P^{(2)}/6]S_{-s}\\&\quad=-A_{-s}S_{-s}\end{aligned}\tag{14}$$

[cf. Refs. 2 and 10]. The expressions for χ_s and χ_{-s} in terms of $R_{\pm s}(y)$ and $S_{\pm s}(x)$ are given in Table III. It may be noticed that the equations for R_{-s} and S_{-s} can be ob-

TABLE I. Explicit form of the functions that determine the null tetrad. The cases called *pGS* and *pDM* are the null orbit solutions. Here a, b, m, n, γ_0, ϵ_0, and λ_0 are arbitrary constants.

Metric	$p(x)$	$q(y)$	$P(x)$	$Q(y)$
B	$k+l$	k	$1-\lambda_0 x^2$	$1-\lambda_0 y^2$
gS	$k+l$	k	$1-\epsilon_0 x^2$	$-\lambda_0/3+\epsilon_0 y^2-2my^3$
g^*S	$k+l$	k	$-\lambda_0/3-\epsilon_0 x^2+2nx^3$	$1+\epsilon_0 y^2$
gC	$k+l$	k	$-\lambda_0/6+b-\epsilon_0 x^2-2mx^3$	$-\lambda_0/6-b+\epsilon_0 y^2-2my^3$
$pC\tilde{B}(+)$	$l+2akx$	$-k(a^2+y^2)$	$1-\epsilon_0 x^2$	$\lambda_0(a^4-2a^2y^2-\frac{1}{3}y^4)$ $-2my+\epsilon_0(y^2-a^2)$
$pC\tilde{B}(-)$	$k(a^2+x^2)$	$-l-2aky$	$\lambda_0(a^4-2a^2x^2-\frac{1}{3}x^4)$ $+2nx-\epsilon_0(x^2-a^2)$	$1+\epsilon_0 y^2$
pGS	$k(a^2+x^2)$	$-l-2aky$	$\lambda_0(a^4-2a^2x^2-\frac{1}{3}x^4)$ $+2nx$	⋯
pCA	$l+kx^2$	$l-ky^2$	$b+2nx-\epsilon_0 x^2-(\lambda_0/3)x^4$	$b-2my+\epsilon_0 y^2-(\lambda_0/3)y^4$
pDM	$l+kx^2$	$l-ky^2$	$2nx$	⋯
pPD	$l+kx^2$	$l-ky^2$	$-\lambda_0/6+\gamma_0+2nx-\epsilon_0 x^2$ $+2mx^3-(\lambda_0/6+\gamma_0)x^4$	$-\lambda_0/6+\gamma_0-2my+\epsilon_0 y^2$ $-2ny^3-(\lambda_0/6+\gamma_0)y^4$

tained from those for $Q^s R_{+s}$ and S_{+s} by replacing (k,l) by $(-k,-l)$; which follows from the fact that, under the replacement $(k,l)\to(-k,-l)$, $p\to -p$, $q\to -q$ and therefore $\mathscr{D}_n \leftrightarrow \mathscr{D}_n^\dagger$, $\mathscr{L}_n \leftrightarrow \mathscr{L}_n^\dagger$. On the other hand, the equation for R_{-s} is the complex conjugate of that for $Q^s R_{+s}$.

A straightforward computation, using Eq. (10) and the fact that $q^{(3)} = Q^{(5)} = 0$, shows that, for $s = \frac{1}{2}, 1, \frac{3}{2}, 2$,

$$\begin{aligned} &Q^s(\mathscr{D}_0^\dagger)^{2s}[Q\mathscr{D}_{1-s}\mathscr{D}_0^\dagger - (2s-1)iq^{(1)} \\ &\quad + (s-1)(2s-1)Q^{(2)}/6] \\ &= [Q\mathscr{D}_{1-s}^\dagger\mathscr{D}_0 + (2s-1)iq^{(1)} \\ &\quad + (s-1)(2s-1)Q^{(2)}/6]Q^s(\mathscr{D}_0^\dagger)^{2s}. \end{aligned} \tag{15}$$

Therefore $Q^s(\mathscr{D}_0^\dagger)^{2s}Q^s R_{+s}$ must be a multiple of R_{-s} and, conversely, $Q^s(\mathscr{D}_0)^{2s}R_{-s}$ must be proportional to $Q^s R_{+s}$ [which can be seen by changing (k,l) to $(-k,-l)$ in Eq. (15)], furthermore, the set of values of the separation constants in Eqs. (11) and (13), A_s and A_{-s}, must be the same.

In a similar way one can show that

$$\begin{aligned} &\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_s[\mathscr{L}_{1-s}^\dagger\mathscr{L}_s + (2s-1)p^{(1)} \\ &\quad + (s-1)(2s-1)P^{(2)}/6] \\ &= [\mathscr{L}_{1-s}\mathscr{L}_s^\dagger - (2s-1)p^{(1)} \\ &\quad + (s-1)(2s-1)P^{(2)}/6]\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_s. \end{aligned} \tag{16}$$

In fact, Eq. (16) can be written in the same form as Eq. (15) since, for example, using Eq. (10),

$$\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_s = P^{s/2}(P^{-1/2}\mathscr{L}_0)^{2s}P^{s/2}$$

and the operator $P^{-1/2}\mathscr{L}_0$ has a form analogous to that of $\mathscr{D}_0$. [Actually, by defining operators equivalent to (10) in a more symmetrical way, all separated equations in the variables x and y can be written in a more symmetrical form. However, we make use of the definitions (10) in order to get expressions similar to those obtained in previous works.] From Eqs. (12), (14), and (16) it follows that $\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_s S_{+s}$ is proportional to S_{-s} and, by changing (k,l) to $(-k,-l)$ in Eq. (16), that $\mathscr{L}_{1-s}^\dagger\mathscr{L}_{2-s}^\dagger\cdots\mathscr{L}_s^\dagger S_{-s}$ is proportional to S_{+s}.

These differential relations also follow directly from the field equations: using Eqs. (2)–(4), (9), and (10) and the

TABLE II. Expressions for the functions that determine the conformal curvature and the spin coefficients. With the exception of only the null orbit metric *pDM*, where the scalar curvature R vanishes, in all other cases $R = 4\lambda_0$.

Metric	M	ϕ	ζ	ξ
B	$\lambda_0/3$	1	$P^{1/4}$	$P^{1/4}Q^{1/2}$
gS	m	y	$P^{1/4}y^{-1}$	$P^{1/4}Q^{1/2}$
g^*S	in	$-ix$	$P^{1/4}x^{-1}$	$P^{1/4}Q^{1/2}$
gC	m	$x+y$	$P^{1/4}(x+y)^{-1}$	$P^{1/4}Q^{1/2}$
$pC\tilde{B}(+)$	$m+i(\epsilon_0 a-\frac{4}{3}\lambda_0 a^3)$	$(y+ia)^{-1}$	$P^{1/4}$	$P^{1/4}Q^{1/2}(y+ia)^{-1}$
$pC\tilde{B}(-)$	$in+\epsilon_0 a+\frac{4}{3}\lambda_0 a^3$	$(a+ix)^{-1}$	$P^{1/4}$	$P^{1/4}Q^{1/2}(a+ix)^{-1}$
pGS	$in+\frac{4}{3}\lambda_0 a^3$	$(a+ix)^{-1}$	$P^{1/4}$	$P^{1/4}(a+ix)^{-1}$
pCA	$m+in$	$(y+ix)^{-1}$	$P^{1/4}$	$P^{1/4}Q^{1/2}(y+ix)^{-1}$
pDM	in	$(y+ix)^{-1}$	$P^{1/4}$	$P^{1/4}(y+ix)^{-1}$
pPD	$m+in$	$(1-xy)/(y+ix)$	$P^{1/4}(1-xy)^{-1}$	$P^{1/4}Q^{1/2}(y+ix)^{-1}$

TABLE III. Decoupled components in terms of the separated functions.

Metric	χ_s	χ_{-s}
$B,pC\bar{B}(+),pC\bar{B}(-)$	$e^{i(ku+lv)}R_{+s}S_{+s}$	$[1/(-\sqrt{2})^{2s}]\phi^{2s}e^{i(ku+lv)}R_{-s}S_{-s}$
pGS,pCA,pDM		
gS,g^*S,gC	$\phi e^{i(ku+lv)}R_{+s}S_{+s}$	$[1/(-\sqrt{2})^{2s}]\phi^{2s+1}e^{i(ku+lv)}R_{-s}S_{-s}$
pPD	$(1-xy)e^{i(ku+lv)}R_{+s}S_{+s}$	$[(1-xy)/(-\sqrt{2})^{2s}]\phi^{2s}e^{i(ku+lv)}R_{-s}S_{-s}$

expressions given in Tables I–III, the neutrino Weyl equation takes the form

$$\begin{aligned}&\mathscr{D}_0(R_{-1/2}S_{-1/2}) = -\mathscr{L}_{1/2}(R_{+1/2}S_{+1/2}),\\ &\mathscr{L}^\dagger_{1/2}(R_{-1/2}S_{-1/2}) = Q^{1/2}\mathscr{D}^\dagger_0(Q^{1/2}R_{+1/2}S_{+1/2});\end{aligned} \tag{17}$$

similarly, using Maxwell's equations, it follows that[7,10]

$$\begin{aligned}&\mathscr{D}_0\mathscr{D}_0(R_{-1}S_{-1}) = \mathscr{L}_0\mathscr{L}_1(R_{+1}S_{+1}),\\ &\mathscr{L}^\dagger_0\mathscr{L}^\dagger_1(R_{-1}S_{-1}) = Q\mathscr{D}^\dagger_0\mathscr{D}^\dagger_0(QR_{+1}S_{+1}),\end{aligned} \tag{18}$$

and, assuming that the Ricci tensor vanishes, from the Rarita–Schwinger equation one obtains[16]

$$\begin{aligned}&\mathscr{D}_0\mathscr{D}_0\mathscr{D}_0(R_{-3/2}S_{-3/2})\\ &\quad = -\mathscr{L}_{-1/2}\mathscr{L}_{1/2}\mathscr{L}_{3/2}(R_{+3/2}S_{+3/2}),\\ &\mathscr{L}^\dagger_{-1/2}\mathscr{L}^\dagger_{1/2}\mathscr{L}^\dagger_{3/2}(R_{-3/2}S_{-3/2})\\ &\quad = Q^{3/2}\mathscr{D}^\dagger_0\mathscr{D}^\dagger_0\mathscr{D}^\dagger_0(Q^{3/2}R_{+3/2}S_{+3/2}).\end{aligned} \tag{19}$$

In the case of the gravitational perturbations there is a pair of equations similar to Eqs. (17)–(19) when Ψ_2 is also zero; however, the presence of Ψ_2 introduces additional terms and complicates the derivation of an analog of Eqs. (17)–(19).

From Eqs. (15)–(19) it follows that, for each value of s, the functions $R_{\pm s}$ and $S_{\pm s}$, present in the components χ_s and χ_{-s} of a given perturbation with a dependence of the form (8), correspond to the same value of the separation constant and, since Q^sR_{+s} and R_{-s} satisfy complex-conjugate equations, these functions can be normalized in such a way that

$$\begin{aligned}&Q^s(\mathscr{D}_0)^{2s}R_{-s}(y) = B_sQ^sR_{+s}(y),\\ &Q^s(\mathscr{D}^\dagger_0)^{2s}Q^sR_{+s}(y) = B_sR_{-s}(y),\end{aligned} \tag{20}$$

with B_s being a real constant. Therefore, in the cases $s=\frac{1}{2}, 1, \frac{3}{2}$, the functions $S_{\pm s}$ satisfy

$$\begin{aligned}&\mathscr{L}_{1-s}\mathscr{L}_{2-s}\cdots\mathscr{L}_sS_{+s}(x) = (-1)^{2s}B_sS_{-s}(x),\\ &\mathscr{L}^\dagger_{1-s}\mathscr{L}^\dagger_{2-s}\cdots\mathscr{L}^\dagger_sS_{-s}(x) = B_sS_{+s}(x).\end{aligned} \tag{21}$$

In the case $s=2$, Q^2R_{+2} and R_{-2} can be normalized according to (20), with B_2 being a real constant; however, if Ψ_2 is different from zero, S_{+2} and S_{-2} would be connected by relations of the form (21) with proportionality constants, different from B_2, complex in general. We can absorb in S_{+2} and S_{-2} the phase factor of those complex constants, thus making Eq. (21) valid also in this case, but then Eq. (20) has to be replaced by

$$\begin{aligned}&Q^2(\mathscr{D}_0)^4R_{-2}(y) = C_2Q^2R_{+2}(y),\\ &Q^2(\mathscr{D}^\dagger_0)^4Q^2R_{+2}(y) = \overline{C_2}R_{-2}(y),\end{aligned} \tag{22}$$

where C_2 is a complex constant. Only the value of $|C_2|^2$ is uniquely fixed and is independent of the normalization of $R_{\pm 2}$. In the cases of the Schwarzschild and the Kerr metrics, Eqs. (21) and (22) lead to simpler expressions than the choice based on (20) (see Ref. 2).

IV. NULL ORBIT METRICS

In the case of the null orbit type D vacuum metrics, the null tetrad can be chosen in such a way that, on quantities with a dependence of the form (8), the tetrad vectors can be replaced according to

$$\begin{aligned}&D\to 2iq, \quad \Delta\to -\tfrac{1}{2}\phi\bar{\phi}\,\partial_y,\\ &\delta\to(1/\sqrt{2})\bar{\phi}\mathscr{L}^\dagger_0, \quad \bar{\delta}\to(1/\sqrt{2})\phi\mathscr{L}_0,\end{aligned} \tag{23}$$

with $\mathscr{L}_0$ and $\mathscr{L}^\dagger_0$ given by Eq. (10). As in the preceding section, $p(x)$ and $q(y)$ are polynomials of degree not greater than 2 and $P(x)$ is a polynomial of degree not greater than 4; but, in this case, the polynomial $Q(y)$ is absent and, moreover, the two principal null directions are not equivalent. The polynomials $p(x)$, $q(y)$, and $P(x)$ corresponding to the null orbit metrics are also given in Table I. The spin coefficients and the conformal curvature can be easily obtained by using Eqs. (1)–(4) and (23) and the expressions listed in Table II.

Following the steps indicated in the preceding section, one finds that the separable solutions to Eqs. (5) and (6) are of the form given in Table III, where the separated functions $R_{\pm s}$ satisfy the equations

$$[2iq\,\partial_y - (2s-1)iq^{(1)}]R_{+s} = A_sR_{+s}, \tag{24}$$

$$[2iq\,\partial_y + (2s+1)iq^{(1)}]R_{-s} = A_sR_{-s}, \tag{25}$$

while S_{+s} and S_{-s} continue satisfying Eqs. (12) and (14), respectively. It may be noticed[4,6] that Eqs. (24) and (25) are first-order equations in contrast to Eqs. (11)–(14).

Using the fact that $q^{(3)}=0$, it is easy to see that

$$\begin{aligned}&(\partial_y)^{2s}[2iq\,\partial_y - (2s-1)iq^{(1)}]\\ &\quad = [2iq\,\partial_y + (2s+1)iq^{(1)}](\partial_y)^{2s}\end{aligned} \tag{26}$$

and

$$\begin{aligned}&(2iq)^{2s}[2iq\,\partial_y + (2s+1)iq^{(1)}]\\ &\quad = [2iq\,\partial_y - (2s-1)iq^{(1)}](2iq)^{2s},\end{aligned} \tag{27}$$

therefore $(\partial_y)^{2s}$ maps solutions of Eq. (24) into solutions of Eq. (25) and a solution of Eq. (25) multiplied by $(2iq)^{2s}$ is a solution of Eq. (24).

Since the functions $S_{\pm s}$ in this case also satisfy Eqs. (12) and (14), they can be normalized to satisfy Eq. (21).

Then, by virtue of the field equations as in Eqs. (17)–(19), it follows that, for $s = \frac{1}{2}, 1$ and, provided the cosmological constant vanishes, for $s = \frac{3}{2}$,

$$(2iq)^{2s}R_{-s}(y) = B_s R_{+s}(y)\,,$$
$$(\partial_y)^{2s}R_{+s}(y) = B_s R_{-s}(y)\,. \tag{28}$$

When $s = 2$ we can only assert that

$$(2iq)^4 R_{-2}(y) = ER_{+2}(y)\,,$$
$$(\partial_y)^4 R_{+2}(y) = FR_{-2}(y)\,, \tag{29}$$

where E and F are complex constants.

V. PROPORTIONALITY CONSTANTS

The proportionality factor B_s, contained in the differential relations (20), can be determined by substituting the second of Eqs. (20) into the first one and using Eqs. (10) and (11). This gives, for example,

$$\begin{aligned}(B_{1/2})^2 Q^{1/2}R_{+1/2} &= Q^{1/2}\mathscr{D}_0 Q^{1/2}\mathscr{D}_0^\dagger Q^{1/2}R_{+1/2}\\ &= Q\mathscr{D}_{1/2}\mathscr{D}_0^\dagger Q^{1/2}R_{+1/2}\\ &= A_{1/2}Q^{1/2}R_{+1/2}\,,\end{aligned}$$

i.e.,

$$(B_{1/2})^2 = A_{1/2}\,. \tag{30}$$

Similarly, a somewhat lengthy computation yields

$$(B_1)^2 = (A_1)^2 + (q^{(1)})^2 - 2qq^{(2)} \tag{31}$$

and

$$\begin{aligned}(B_{3/2})^2 = (A_{3/2})^3 &+ \{-\tfrac{1}{6}[QQ^{(4)} - Q^{(1)}Q^{(3)} + \tfrac{1}{2}(Q^{(2)})^2] + 4(q^{(1)})^2 - 8qq^{(2)}\}A_{3/2}\\ &+ \tfrac{4}{3}\{\tfrac{1}{2}q^2Q^{(4)} - qq^{(1)}Q^{(3)} + (qq^{(2)} + (q^{(1)})^2)Q^{(2)} - 3q^{(1)}q^{(2)}Q^{(1)} + 3(q^{(2)})^2Q\}\\ &+ \tfrac{1}{36}\{-2QQ^{(2)}Q^{(4)} + \tfrac{3}{2}(Q^{(1)})^2Q^{(4)} + Q(Q^{(3)})^2 - Q^{(1)}Q^{(2)}Q^{(3)} + \tfrac{1}{3}(Q^{(2)})^3\}\,.\end{aligned} \tag{32}$$

It can be verified, using the fact that $q^{(3)} = Q^{(5)} = 0$, that the expressions (31) and (32) are actually constant; moreover, each term between braces in (32) is constant.

On the other hand, using Eqs. (10), (12), and (21), one obtains Eq. (30) and the alternative expressions

$$(B_1)^2 = (A_1)^2 - (p^{(1)})^2 + 2pp^{(2)}\,, \tag{33}$$

$$\begin{aligned}(B_{3/2})^2 = (A_{3/2})^3 &+ \{-\tfrac{1}{6}[PP^{(4)} - P^{(1)}P^{(3)} + \tfrac{1}{2}(P^{(2)})^2] - 4(p^{(1)})^2 + 8pp^{(2)}\}A_{3/2}\\ &+ \tfrac{4}{3}\{\tfrac{1}{2}p^2P^{(4)} - pp^{(1)}P^{(3)} + (pp^{(2)} + (p^{(1)})^2)P^{(2)} - 3p^{(1)}p^{(2)}P^{(1)} + 3(p^{(2)})^2P\}\\ &+ \tfrac{1}{36}\{2PP^{(2)}P^{(4)} - \tfrac{3}{2}(P^{(1)})^2P^{(4)} - P(P^{(3)})^2 + P^{(1)}P^{(2)}P^{(3)} - \tfrac{1}{3}(P^{(2)})^3\}\,,\end{aligned} \tag{34}$$

and

$$\begin{aligned}(B_2)^2 = \{(A_2)^2 &- \tfrac{1}{2}[PP^{(4)} - P^{(1)}P^{(3)} + \tfrac{1}{2}(P^{(2)})^2]\}^2 - 10\{(p^{(1)})^2 - 2pp^{(2)}\}(A_2)^2\\ &+ 8\{\tfrac{1}{2}p^2P^{(4)} - pp^{(1)}P^{(3)} + (pp^{(2)} + (p^{(1)})^2)P^{(2)} - 3p^{(1)}p^{(2)}P^{(1)} + 3(p^{(2)})^2P\}A_2 + 9\{(p^{(1)})^2 - 2pp^{(2)}\}^2\\ &+ \{p^2(2P^{(2)}P^{(4)} - (P^{(3)})^2) + pp^{(2)}(6PP^{(4)} - (P^{(2)})^2) + 3(p^{(2)})^2(4PP^{(2)} - 3(P^{(1)})^2) + 2pp^{(1)}(P^{(2)}P^{(3)} - 3P^{(1)}P^{(4)})\\ &+ 3(p^{(1)})^2(PP^{(4)} + P^{(1)}P^{(3)} - \tfrac{5}{6}(P^{(2)})^2) - 6p^{(1)}p^{(2)}(2PP^{(3)} - P^{(1)}P^{(2)})\}\,.\end{aligned} \tag{35}$$

In a similar manner, from Eqs. (10), (11), and (22) it follows that

$$\begin{aligned}|C_2|^2 = \{(A_2)^2 &- \tfrac{1}{2}[QQ^{(4)} - Q^{(1)}Q^{(3)} + \tfrac{1}{2}(Q^{(2)})^2]\}^2 + 10\{(q^{(1)})^2 - 2qq^{(2)}\}(A_2)^2\\ &+ 8\{\tfrac{1}{2}q^2Q^{(4)} - qq^{(1)}Q^{(3)} + (qq^{(2)} + (q^{(1)})^2)Q^{(2)} - 3q^{(1)}q^{(2)}Q^{(1)} + 3(q^{(2)})^2Q\}A_2 + 9\{(q^{(1)})^2 - 2qq^{(2)}\}^2\\ &- \{q^2(2Q^{(2)}Q^{(4)} - (Q^{(3)})^2) + qq^{(2)}(6QQ^{(4)} - (Q^{(2)})^2) + 3(q^{(2)})^2(4QQ^{(2)} - 3(Q^{(1)})^2) + 2qq^{(1)}(Q^{(2)}Q^{(3)} - 3Q^{(1)}Q^{(4)})\\ &+ 3(q^{(1)})^2(QQ^{(4)} + Q^{(1)}Q^{(3)} - \tfrac{5}{6}(Q^{(2)})^2) - 6q^{(1)}q^{(2)}(2QQ^{(3)} - Q^{(1)}Q^{(2)})\}\,.\end{aligned} \tag{36}$$

(As indicated in Sec. III, the results concerning $S_{\pm s}$ can be obtained from those corresponding to $R_{\pm s}$, and vice versa, by means of appropriate substitutions.) It can be readily seen that each term between braces in (34)–(36) is constant.

In the case of the null orbit metrics the computations are much simpler and one finds that the proportionality constants appearing in Eq. (28) are given by Eqs. (30)–(32), with Q replaced by zero. Similarly, the product, EF, of the constants introduced in Eq. (29) is given by the right-hand side of (36) if Q is set equal to zero.

From the equivalence of Eqs. (31) and (33) one finds that

$$(q^{(1)})^2 - 2qq^{(2)} = -(p^{(1)})^2 + 2pp^{(2)}\,, \tag{37}$$

which provides a relationship between the (coefficients of the) polynomials p and q corresponding to each metric. Similarly, the equality of (32) and (34) gives another relationship valid when the cosmological constant vanishes. Nevertheless, by explicit computations using the expressions given

in Table I we find that the constants appearing in Eqs. (32) and (34)–(36) are related by

$$QQ^{(4)} - Q^{(1)}Q^{(3)} + \tfrac{1}{2}(Q^{(2)})^2 = PP^{(4)} - P^{(1)}P^{(3)} + \tfrac{1}{2}(P^{(2)})^2, \tag{38}$$

$$\begin{aligned}&\tfrac{1}{2}q^2Q^{(4)} - qq^{(1)}Q^{(3)} \\ &\quad + (qq^{(2)} + (q^{(1)})^2)Q^{(2)} - 3q^{(1)}q^{(2)}Q^{(1)} + 3(q^{(2)})^2Q \\ &\quad = \tfrac{1}{2}p^2P^{(4)} - pp^{(1)}P^{(3)} + (pp^{(2)} + (p^{(1)})^2)P^{(2)} \\ &\qquad - 3p^{(1)}p^{(2)}P^{(1)} + 3(p^{(2)})^2P,\end{aligned} \tag{39}$$

$$\begin{aligned}&-2QQ^{(2)}Q^{(4)} + \tfrac{3}{2}(Q^{(1)})^2Q^{(4)} \\ &\quad + Q(Q^{(3)})^2 - Q^{(1)}Q^{(2)}Q^{(3)} + \tfrac{1}{3}(Q^{(2)})^3 \\ &\quad = 2PP^{(2)}P^{(4)} - \tfrac{3}{2}(P^{(1)})^2P^{(4)} - P(P^{(3)})^2 + P^{(1)}P^{(2)}P^{(3)} \\ &\qquad - \tfrac{1}{3}(P^{(2)})^3 - \tfrac{1}{3}\lambda_0|12M|^2,\end{aligned} \tag{40}$$

even if the cosmological constant λ_0 is different from zero. Furthermore, we find that

$$|C_2|^2 = (B_2)^2 + |12Mh|^2, \tag{41}$$

where h is defined by

$$\phi^{-1}(\rho\Delta + \mu D - \tau\bar{\delta} - \pi\delta)e^{i(ku+lv)} = he^{i(ku+lv)}. \tag{42}$$

The vector field $\phi^{-1}(\rho\Delta + \mu D - \tau\bar{\delta} - \pi\delta)$ is a Killing vector that turns out to be a linear combination of ∂_u and ∂_v. Hence h is a combination of the separation constants k and l.

Equations (38)–(42) also apply for the case of the null orbit metrics if Q is set equal to zero, with $|C_2|^2$ replaced by EF in Eq. (41).

VI. CONCLUDING REMARKS

As shown in Ref. 10, in the case of the electromagnetic perturbations the Teukolsky–Starobinsky identities are intimately related to the separability of Maxwell's equations and in the case of the Weyl neutrino equation something similar occurs. In view of the results presented here, one may conjecture that an analogous relationship exists also in the cases of the spin-$\frac{3}{2}$ and of the gravitational perturbations.

It is somewhat surprising the fact that certain relations between the functions contained in the metric arise from the equations for perturbations [Eqs. (37)–(41)]. Probably, these relations can also be derived directly from Einstein's equations. The relation (41), obtained here by explicit computation, must be obtainable directly from the linearized Bianchi identities.

ACKNOWLEDGMENTS

The author would like to thank Professor R. Penrose for the hospitality provided at the Mathematical Institute, University of Oxford, where this work was begun.

The author also acknowledges the support received from the Sistema Nacional de Investigadores (Mexico).

[1] S. A. Teukolsky, Astrophys. J. **185**, 635 (1973).
[2] S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Clarendon, Oxford, 1983).
[3] A. L. Dudley and J. D. Finley, III, J. Math. Phys. **20**, 311 (1979).
[4] N. Kamran and R. G. McLenaghan, in *Gravitation and Geometry: a Volume in Honor of I. Robinson,* edited by W. Rindler and A. Trautman (Bibliopolis, Naples, 1987).
[5] R. Güven, Phys. Rev. D **22**, 2327 (1980).
[6] N. Kamran, J. Math. Phys. **26**, 1740 (1985).
[7] S. A. Teukolsky and W. H. Press, Astrophys. J. **193**, 443 (1974).
[8] A. A. Starobinsky and S. M. Churilov, Zh. Eksp. Teor. Fiz. **65**, 3 (1973) [Sov. Phys. JETP **38**, 1 (1974)].
[9] B. P. Jeffryes, Proc. R. Soc. London Ser. A **392**, 323 (1984).
[10] G. F. Torres del Castillo, J. Math. Phys. **29**, 971 (1988).
[11] A. García D. and H. Salazar I., Gen. Relativ. Gravit. **16**, 417 (1984).
[12] R. Debever, N. Kamran, and R. G. McLenaghan, Phys. Lett. A **93**, 399 (1983); Bull. Cl. Sci. Acad. R. Belg. **68**, 592 (1982); J. Math. Phys. **25**, 1955 (1984).
[13] R. Penrose and W. Rindler, *Spinors and Space-time* (Cambridge U.P., Cambridge, 1984), Vol. 1.
[14] R. Geroch, A. Held, and R. Penrose, J. Math. Phys. **14**, 874 (1973).
[15] P. C. Aichelburg and H. K. Urbantke, Gen. Relativ. Gravit. **13**, 817 (1981).
[16] G. F. Torres del Castillo, "Rarita–Schwinger fields in algebraically special vacuum space-times," submitted to J. Math. Phys.

A 29-root soluble model for the nonlinear calculation of a four-body propagator

B. G. Giraud, S. Kessal,[a] and A. Weiguny[b]
Service de Physique Théorique, DPH-G, CEN Saclay, F91191, Gif-sur-Yvette, France

(Received 2 October 1987; accepted for publication 18 May 1988)

The time-independent mean-field theory of collisions is applied to the collision of four one-dimensional particles. Their two-body interactions are taken as separable, with Lorentzian form factors. This allows completely analytical solutions. A unique and satisfactory physical branch emerges out of 29 candidate solutions. It is very stable when the strength of the interaction is modified.

I. INTRODUCTION

The N-body problem can often be formalized, in quantum mechanics, by the calculation of a propagator $G=(E-H)^{-1}$, where the Hamiltonian $H=\mathbb{T}+\mathbb{V}$ reads, in obvious notation,

$$H=\sum_{i=1}^{N} t_i+\sum_{i>j}^{N} v_{ij}, \tag{1.1}$$

and E is a complex parametric energy.

This is formally, a trivial linear problem, with a unique solution when $\mathrm{Im}\,E>0$. In practice, as soon as there are more than three particles, and Faddeev equations are no longer available, this would-be linear problem amounts to the formidable inversion of a huge matrix in a large-dimensional space. The numerical accuracy demanded by the intricacies of the on-shell limit, $\mathrm{Im}\,E=0$, is out of reach.

This is why we have proposed a nonlinear approximation[1] to the calculation of a matrix element such as

$$D\equiv\langle\chi|(E-H)^{-1}|\chi\rangle\,, \tag{1.2}$$

where χ is a generic product of single-particle orbitals

$$\chi(\mathbf{r}_1,\ldots,\mathbf{r}_N)=\chi_1(\mathbf{r}_1)\cdots\chi_N(\mathbf{r}_N)\,. \tag{1.3}$$

For that purpose, it is trivial to show that D is the stationary value of the functional F of ϕ,

$$F\equiv(\phi|\chi)^2/(\phi|(E-H)|\phi)\,, \tag{1.4}$$

where round brackets $(\ |\ )$ denote the absence of complex conjugation for bras. An ansatz for ϕ similar to Eq. (1.3) for χ,

$$\phi=\varphi_1\cdots\varphi_N, \tag{1.5}$$

provides the variational, nonlinear equations

$$(\eta_i-h_i)\varphi_i=\chi_i\,, \tag{1.6}$$

with the complex, single-particle self-energies

$$\eta_i\equiv E-\sum_{j\neq i}\frac{\theta_j}{n_j}-\sum_{\substack{j>k\\ j\neq i\\ k\neq i}}\frac{w_{jk}}{n_j n_k}, \tag{1.7}$$

where

$$n_i=(\varphi_i|\varphi_i)\,, \tag{1.8}$$

$$\theta_i=(\varphi_i|t|\varphi_i)\,, \tag{1.9}$$

$$w_{jk}=(\varphi_j\varphi_k|v|\varphi_j\varphi_k)\,, \tag{1.10}$$

$$h_i=t_i+U_i\,, \tag{1.11}$$

and finally

$$\langle\mathbf{p}'|U_i|\mathbf{p}\rangle=\int d\mathbf{p}''\,d\mathbf{p}'''\,\langle\mathbf{p}'\mathbf{p}''|v|\mathbf{p}\mathbf{p}'''\rangle \times\sum_{j\neq i} n_j^{-1}\varphi_j(\mathbf{p}'')\varphi_j(\mathbf{p}''')\,. \tag{1.12}$$

This set of equations, Eqs. (1.6)–(1.12), is plain Hartree theory, except for the right-hand side of Eq. (1.6), the source term χ_i. Antisymmetrization is trivial and has been published elsewhere.[2]

This problem has been investigated in the framework of a soluble model[3,4] where only the kinetic energy is present. It may be pointed out here that, although T is a trivial one-body operator, the bare propagator $(E-\mathrm{T})^{-1}$ is a nontrivial, many-body operator. It was thus a significant result that the approximation generated by Eq. (1.6) turned out to be excellent[3] and that one out of its five competing solutions could easily be selected as the unique, physical approximation[4] for outgoing waves.

The purpose of the present paper is to generalize the soluble model[3] into a model including interactions. As will be shown, an analytical control of the whole algebra can be obtained all the way. It will turn out, once again, that only one of the many solutions (29 in that case) generated by nonlinearity is physically acceptable.

In Sec. II we describe the model and the basic equations that make it soluble. The technical, analytical solution of the model is explained in Sec. III. Then we study the main properties of the solution in Sec. IV, where numerical results are displayed. Our conclusion is proposed in Sec. V.

II. THE MODEL

Consider the symmetry group S of a polyhedron in a d-dimensional space, and a rotation invariant wave packet $\chi(\mathbf{p})$ centered at the origin in that space. A vector $\mathbf{K}$ will be associated to the first particle, via the wave packet

$$\chi_1(\mathbf{p}_1)\equiv\chi(\mathbf{p}_1-\mathbf{K})\,. \tag{2.1}$$

[a] Permanent address: Institut de Physique, Université des Sciences et de la Technologie Houari Boumediène, PB 32 El Alia, Alger, Algeria.
[b] Permanent address: Institut für Theoretische Physik, Universität Münster, Domagkstrasse, Münster, Federal Republic of Germany.

 0022-2488/88/092084-06$02.50

Then we let the group S act upon $\mathbf{K}$ and generate $\mathbf{K}_2,\ldots,\mathbf{K}_N$. We define the wave packets

$$\chi_i(\mathbf{p}_i)\equiv\chi(\mathbf{p}_i-\mathbf{K}_i),\quad i=2,\ldots,N, \tag{2.2}$$

and consider the N-body amplitude $D=\langle\chi|G|\chi\rangle$, where $\chi=\Pi_{i=1}^{N}\chi_i$.

It is clear that S is a symmetry group for an exact solution of this problem and that the mean-field approximation can take great advantage of this symmetry. As already discussed in Ref. 3, the mean-field orbitals φ_i span a representation of S, provided of course the two-body interaction commutes with S.

In the following we take dimension $d=1$, and our potential is defined by

$$\langle\mathbf{p}'\mathbf{p}''|v|\mathbf{p}\mathbf{p}'''\rangle=-\lambda_0 f(\mathbf{p}')f(\mathbf{p}'')f(\mathbf{p})f(\mathbf{p}'''), \tag{2.3}$$

where λ_0 is a strength constant and the form factor f is given as a Lorentzian with width γ,

$$f(\mathbf{p})=(2/\pi)^{1/2}\gamma(p^2+\gamma^2)^{-1}, \tag{2.4}$$

which allows easy analytical calculations. Many other choices and generalizations are obviously available. For the sake of definitiveness, the width γ is taken as $\gamma=1$ fm^{-1}, a typical value for nuclear physics as an example.

Finally we set

$$\chi(p-K)=(2/\pi)^{1/2}\gamma[(p-K)^2+\gamma^2]^{-1}\quad(\text{fm}), \tag{2.5}$$

a boosted Lorentzian with the same width γ. In the following we list dimensions after each new symbol. We obtain dimension fm^{-4} for λ_0, so that the matrix element $\langle p'p''|v|pp'''\rangle$ becomes dimensionless.

The symmetry group S and the separability of v induce a degeneracy of all the quantities n, θ, and w listed in Eqs. (1.8)–(1.10). Hence it is trivial to omit the subscript i and write Eq. (1.6) as

$$(\eta-p^2)\varphi(p)=\chi(p-K)+\int dp'\langle p|U|p'\rangle\varphi(p'), \tag{2.6}$$

where $\hbar^2/2m=1$ (hence the dimension of E is fm^{-2}), and

$$\eta=E-(N-1)\theta/n+\Lambda\alpha^4/n^2\quad(\text{fm}^{-2}), \tag{2.7}$$

$$\Lambda=\tfrac{1}{2}(N-1)(N-2)\lambda_0\quad(\text{fm}^{-4}), \tag{2.8}$$

$$\alpha=(\varphi|f)\quad(\text{fm}^3), \tag{2.9}$$

$$\langle p|U|p'\rangle=-\lambda(\alpha^2/n)f(p)f(p')\quad(\text{fm}^{-1}), \tag{2.10}$$

$$\lambda=(N-1)\lambda_0\quad(\text{fm}^{-4}). \tag{2.11}$$

Hence the self-consistency demanded by the variational principle reduces to the determination of the three parameters θ (or η), α, and n of the mean-field orbital

$$\varphi(p)=\frac{\chi(p-K)}{\eta-p^2}-\lambda\frac{\alpha^3}{n}\frac{f(p)}{\eta-p^2}\quad(\text{fm}^3). \tag{2.12}$$

As will be shown below, this self-consistency condition will reduce to only the search for the roots of a polynomial.

For that purpose we define the three basic integrals

$$I_1(\eta)\equiv\int dp\,\frac{\chi^2(p-K)}{\eta-p^2}\quad(\text{fm}^3), \tag{2.13a}$$

$$I_2(\eta)\equiv\int dp\,\frac{\chi(p-K)f(p)}{\eta-p^2}\quad(\text{fm}^3), \tag{2.13b}$$

$$I_3(\eta)\equiv\int dp\,\frac{f^2(p)}{\eta-p^2}\quad(\text{fm}^3), \tag{2.13c}$$

and their derivatives, $I'\equiv dI/d\eta$. It is easy to insert Eq. (2.12) into Eqs. (1.8), (1.9), and (2.9), respectively, and obtain

$$n=-I_1'+2\lambda\alpha^3I_2'n^{-1}-\lambda^2\alpha^6I_3'n^{-2}\quad(\text{fm}^5), \tag{2.14}$$

$$\theta=\eta n-I_1+2\lambda\alpha^3I_2n^{-1}-\lambda^2\alpha^6I_3n^{-2}\quad(\text{fm}^3), \tag{2.15}$$

$$\alpha=I_2-\lambda\alpha^3I_3n^{-1}\quad(\text{fm}^3). \tag{2.16}$$

Through the integrals I,I', these three equations, Eqs. (2.14)–(2.16), are parametrized by η, Eq. (2.7). This closes the self-consistency algebra.

A numerical iterative procedure thus seems to be in order, and easy. A fully analytical solution is, however, available. We first define the auxiliary quantity

$$A\equiv\alpha^2/n, \tag{2.17}$$

and obtain from Eq. (2.16)

$$\alpha=I_2/(1+\lambda AI_3). \tag{2.18}$$

Hence Eqs. (2.14) and (2.15) can be written in terms of A only,

$$n=-I_1'+2\lambda A\frac{I_2}{1+\lambda AI_3}I_2'-\lambda^2A^2\left(\frac{I_2}{1+\lambda AI_3}\right)^2I_3', \tag{2.19}$$

$$\theta=\eta n-I_1+2\lambda A\frac{I_2}{1+\lambda I_3}I_2-\lambda^2A^2\left(\frac{I_2}{1+\lambda AI_3}\right)^2I_3. \tag{2.20}$$

Since by definition $A=\alpha^2/n$, we find from Eqs. (2.18) and (2.19)

$$A=I_2^2/[-(1+\lambda AI_3)^2I_1'+2\lambda A(1+\lambda AI_3)I_2I_2'-\lambda^2A^2I_2^2I_3'], \tag{2.21}$$

where A and η are present. Returning to the definition of η, Eq. (2.7), we find

$$\eta=E-(N-1)\eta-(N-1)\frac{(1+\lambda AI_3)^2I_1-2\lambda A(1+\lambda AI_3)I_2^2+\lambda^2A^2I_2^2I_3}{(1+\lambda AI_3)^2I_1'-2\lambda A(1+\lambda AI_3)I_2I_2'+\lambda^2A^2I_2^2I_3'}+\Lambda A^2, \tag{2.22}$$

which simplifies, through Eq. (2.21), into

$$0=E-N\eta+(N-1)A[(1+\lambda AI_3)^2I_1-2\lambda A(1+\lambda AI_3)I_2^2+\lambda^2A^2I_2^2I_3]/I_2^2+\Lambda A^2, \tag{2.23}$$

another condition between only A and η.

It is remarkable that, as far as A is concerned, Eqs. (2.21) and (2.23) are only cubic and read

$$aA^3+bA^2+cA+d=0, \tag{2.21'}$$

$$a'A^3 + b'A^2 + c'A + d' = 0\,, \tag{2.23'}$$

with

$$a = \lambda^2(I_2^2 I_3' - 2I_2 I_2' I_3 + I_1' I_3^2)\,, \tag{2.24a}$$

$$b = 2\lambda(I_1' I_3 - I_2 I_2')\,, \tag{2.24b}$$

$$c = I_1'\,, \tag{2.24c}$$

$$d = I_2^2\,, \tag{2.24d}$$

$$a' = \lambda^2(N-1)I_3(I_1 I_3 - I_2^2)\,, \tag{2.24e}$$

$$b' = \lambda\{2(N-1)I_1 I_3 - [(3N-2)/2]I_2^2\}, \tag{2.24f}$$

$$c' = (N-1)I_1\,, \tag{2.24g}$$

$$d' = (E - N\eta)I_2^2\,. \tag{2.24h}$$

The traditional compatibility condition between two such cubic equations reads

$$\det \equiv \begin{bmatrix} a & a' & 0 & 0 & 0 & 0 \\ b & b' & a & a' & 0 & 0 \\ c & c' & b & b' & a & a' \\ d & d' & c & c' & b & b' \\ 0 & 0 & d & d' & c & c' \\ 0 & 0 & 0 & 0 & d & d' \end{bmatrix} = 0 \tag{2.25}$$

which is an equation where only η is left.

An elementary manipulation, after inspection of Eqs. (2.24), shows that det is a sum of three terms, of order λ^6, λ^5, and λ^4, respectively,

$$\det_6 = (ad' - da')^3\,, \tag{2.26a}$$

$$\det_5 = (ab' - ba')[(bd' - db')^2 - 2\det_0(ad' - da')] - (ac' - ca')(ad' - da')(bd' - db')\,, \tag{2.26b}$$

$$\det_4 = \det_0 \det_2\,, \tag{2.26c}$$

with

$$\det_0 = cd' - dc'\,, \tag{2.26d}$$

$$\det_2 = (ac' - ca')^2 - (ab' - ba')(bc' - cb')\,. \tag{2.26e}$$

As is now obvious from the definitions of the integrals I, Eqs. (2.13), and the explicit forms of f and χ, Eqs. (2.4) and (2.5), the coefficients $a, b, \ldots, d'$, and the determinant det are rational fractions with respect to the auxiliary variable $\omega \equiv \eta^{1/2}$. Hence the solution of Eq. (2.25) boils down to the search for the roots of a polynomial. Our model is thus analytical. Its properties are the subject of Sec. III.

III. ANALYTICAL PROPERTIES OF THE SOLUTION

A straightforward contour integration in the upper half of the complex p plane provides

$$I_1 = -\frac{1}{z^2 + K^2}\left[1 + \frac{2z^2}{(z-1)(z^2+K^2)}\right], \tag{3.1a}$$

$$I_2 = -\frac{2}{K^2+4}\,\frac{2z(z+1) + K^2}{z(z-1)(z^2+K^2)}\,, \tag{3.1b}$$

$$I_3 = -\frac{z+1}{z^2(z-1)}\,, \tag{3.1c}$$

hence,

$$I_1' = -\frac{z[z^4 + z^3 + (K^2-1)z^2 - 3K^2 z + 3K^2]}{[(z-1)(z^2+K^2)]^3}\,, \tag{3.1d}$$

$$I_2' = -\frac{1}{K^2+4} \times \frac{4z^5 + 4z^4 + 4(K^2-1)z^3 + K^2 z^2 + 2K^4 z - K^4}{z^2(z-1)^3(z^2+K^2)^2}\,, \tag{3.1e}$$

$$I_3' = -\frac{z^2 + z - 1}{[z(z-1)]^3}\,, \tag{3.1f}$$

where we have defined z from

$$\omega = i(z - \gamma)\,, \quad \gamma = 1\,, \tag{3.2}$$

and ω from

$$\eta = \omega^2, \quad \operatorname{Im}\omega > 0\,. \tag{3.3}$$

This condition, $\operatorname{Im}\omega > 0$, is linked to the calculation of our contour integrals, see Eqs. (2.13), in the upper half of the complex p plane.

Insertion of Eqs. (3.1) into Eqs. (2.24) gives $a, b, \ldots, d'$ as rational fractions of z (or ω). We notice that all of these coefficients $a, b, \ldots, d'$ depend on K. Also most of them depend on N, if only because λ depends on N. The only coefficient that depends on E, however, is d'. For the sake of definiteness, we set $\operatorname{Re} E = NK^2$, and $N = 4$. The symmetry group S is realized if, for instance, $K_1 = K_2 = K$ and $K_3 = K_4 = -K$. These specifications leave λ, $\operatorname{Im} E$, and K as the only parameters for our analysis of the model.

Before going into a detailed discussion of Eq. (2.25), a remark on the behavior of the solution as a function of λ is in order. On one hand, when $\lambda = 0$, corresponding to a pure kinetic model similar to that studied in Ref. 3, the coefficients a, b, a', b' vanish. There only $\det_0$, Eq. (2.26d), makes sense. As a matter of fact, $\det_4$ is the product of $\det_0$ and $\det_2$. When λ takes on nonvanishing values, it is clear that those roots of det that are continuations of roots of $\det_0$ are likely to be more physical than the roots of det that are continuations of the roots of $\det_2$.

On the other hand, the model should also accommodate a "potential" limit, corresponding to the situation where $|\lambda|$ is very large. There, the dominant term of det is $\det_6$, which leads to threefold degenerate roots. We must thus search, and keep track of, "physical" roots ω of det, which are continuous functions of λ, and which should interpolate between roots of $\det_0$ and roots of $\det_6$.

The straightforward, but tedious calculation of det, or rather of its polynomial numerator, demanded our use of symbolic language AMP, since an explicit listing of this numerator demands ten pages of computer output. This polynomial cannot be published here, naturally, but is available on request. It turns out to be a polynomial of degree 46. Out of it a factor $(2z^2 + 2z + K^2)^4(z-1)^6 z^3$ can be factorized and at once discarded, for the roots of this factor depend neither on λ nor on $\operatorname{Im} E$, and are thus unphysical. This leaves a polynomial of degree 29.

For the small λ limit, the numerator of $\det_0$ provides a polynomial of degree 7, while the numerator of $\det_2$ provides a polynomial of degree 16 and a factor z^6. As already noticed, we have less interest in the latter 22 (16 + 6) roots, for they do not depend on $\operatorname{Im} E$. The numerical investigation, whose final results are abstracted in Sec. IV, further revealed that these 22 roots failed to satisfy the condition $\operatorname{Im}\omega > 0$. An

analytical, rather than numerical, proof of this failure could not be found, but in the following we focus on those branches of roots of det that are continuations of the seven roots of $\det_0$.

For the large λ limit, the numerator of $\det_6$ provides a polynomial of degree 42, which includes a trivial factor $(2z^2+2z+K^2)^6(z-1)^6z^3$ rather than $(2z^2+2z+K^2)^4 \times (z-1)^6z^3$ like det. This leaves seven triply degenerate roots as potentially "physical" candidates.

The difference between the total degree (42) obtained from $\det_6$, and the total degree (46) obtained from det or $\det_4$ corresponds thus to four additional roots ω that diverge when λ is very large and are seemingly of little interest.

We have not found an elegant physical justification for the triple degeneracy of the seven roots of $\det_6$ we retain for our analysis, but the numerical and physical results are quite unambiguous. As a matter of fact, we have studied numerically the 29 nontrivial roots of the full det as functions of λ, Im E, and K, and studied their connection with the "interesting" roots of $\det_0$ and the "interesting" roots of $\det_6$. It was critical that only one of these 29 roots be acceptable, for the linear problem at the beginning of our theory has a unique physical answer. Satisfactorily enough, the following selection and classification scheme for roots turned out to be available: a root ω is acceptable if and only if

(i) $\mathrm{Im}\,\omega > 0$,

(ii) $\mathrm{Im}\,\overline{D} < 0$,

when $\mathrm{Im}\,E > 0$, where $\overline{D}$ is that approximate amplitude generated by ω and the corresponding φ via Eq. (1.4),

(iii) $\mathrm{Im}\,\eta(\equiv \mathrm{Im}\,\omega^2) > 0$,

(iv) $\omega^2 \simeq K^2 + \mathrm{Im}\,E$, when $\mathrm{Im}\,E \gg K^2$,

(v) $\mathrm{Re}\,A > 0$ and $\mathrm{Im}\,A > 0$.

The first two rules, (i) and (ii), are mathematically rigorous, as seen from our choice of "upper" contour integrals and our search for retarded diagonal amplitudes, respectively.

The third rule is almost rigorous, in the sense that many-body retardation should be translated into single-particle retardation, hence a self-energy η with a positive imaginary part.

The fourth rule, (iv), is empirical, but fairly intuitive insofar as $\mathrm{Re}\,\eta$ ($\equiv \mathrm{Re}\,\omega^2$) is concerned for the source wave packet χ and the variational wave packet φ should have similar energies. For the imaginary part, this rule hints that retardation effects are similar for the single-particle and the many-particle propagations.

The first four rules were sufficient to eliminate all roots but one. The survivor root turns out also to be almost independent of λ, which almost decouples the self-energy from the interaction strength, an interesting approximation. The last rule is an empirical, and apparently specific property, of this survivor root. More details are given in Sec. IV.

IV. NUMERICAL RESULTS

We have solved Eq. (2.25) for a large range of parameters, namely λ ranging from 0.001 to 120 fm^{-4}, K from 1 to 5 fm^{-1} (which means, with traditional units, Re E from 80 to 2000 MeV), and Im E from 0 to 2000 MeV. Most roots are at once eliminated because they violate rule (i) and/or rule (ii). This leaves at most four to five roots to be tracked when the various parameters λ, Im E, and K evolve.

The "on-shell" case, namely Im $E = 0$, is of special interest, since it corresponds to a physical collision. As χ has been chosen a real wave function and since the functional F, Eq. (1.4), is formally real, the analyticity and formal reality of ϕ with respect to the parameters η and α generate pairs of conjugate solutions $(\omega, -\omega^*)$, hence pairs of conjugate self-energies (η, η^*). In other words, the variational principle systematically generates both an advanced and a retarded solution in this on-shell limit. Any advanced solution is eliminated under rules (i) and/or (ii), of course. Real roots for η (purely imaginary for ω), which do not appear in pairs, are more delicate to analyze, for they may be candidates for a principal part amplitude. We have scrutinized them one by one.

In order to sort the survivor roots, it is very useful to study the pure kinetic energy case, namely $\lambda = 0$. We know already from our past experience[3] that this model exhibits a well-defined physical root ω_1, which is a smooth function of K and Im E. A striking property of this root is that it fits rule (iv) very closely. An even more striking property of ω_1 is that the bare (kinetic) propagator amplitudes D (exact) and $\overline{D}$ (approximate) are very close to each other, while the $\overline{D}$ generated by any other competitor root in this kinetic model generates a much poorer approximation. Hence there is no doubt that in this kinetic limit ω_1 is that unique physical root to be found.

Our strategy is thus very simple.

First we study the "kinetic" model, namely the seven roots of $\det_0$ in the present case, as functions of only K and Im E. It is easy to isolate the physical root there. Its behavior is shown in Fig. 1, where we have also shown, for the sake of completeness, branches $(-\omega_2^*)$ of an "advanced" root that merge with the branches of the physical root ω_1 when Im $E = 0$. There is little comment to add, except that both Ref. 3, where there were five roots, and the present model

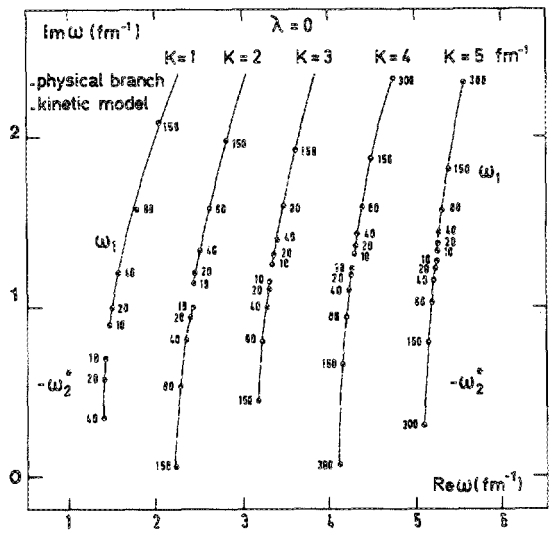


FIG. 1. Kinetic model (no interaction). Each point is labeled by Im E, in a system where $\hbar^2/2m = 20$ MeV fm^2. The upper branches correspond to the physical solution. The lower branches correspond to the complex conjugation of the opposite of another root.

show that there is no branch crossing as long as $K \geqslant 2\gamma$, namely that the average momentum exceeds the fluctuations in the wave packets χ_i.

Then we freeze Im $E = 0$, which is the physical situation, and let λ increase. We follow that root which is a continuation of ω_1 and also any other root whose branch may interfere. A typical situation is shown in Fig. 2, with $K = 3$. The root ω_1 hardly depends on λ, while two other roots tend to merge with ω_1 when λ increases. This merging remarkably confirms the triple degeneracy expected from our formal analysis in Sec. III. For a value of K ($= 3$) which is sufficiently large compared to that of γ ($= 1$), we find no branch crossing that could interfere with the identification of ω_1.

Some of the roots plotted on Fig. 2 are shown between parentheses, in order to indicate that they violate one or several of the rules (i)–(v). The physical branch seems to be the only one for which these rules are never violated.

Finally we freeze K and λ, the latter to a large value in order to have a nontrivial potential influence, and study the behavior of the roots as functions of Im E. This is shown in Fig. 3, in a typical situation $K = 3$ again. The root ω_1 is very well approximated by its qualitative estimate, rule (iv), competing branches show some illegal points (between parentheses), and again we find no branch crossing.

To summarize our long numerical analysis, whose details cannot be published extensively, there is indeed only one physical branch provided by this many-solution model.

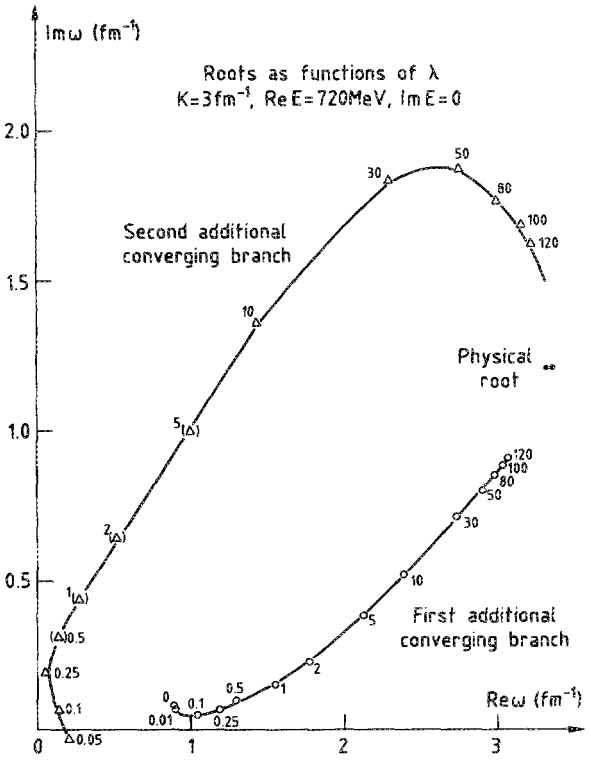


FIG. 2. Behavior of the roots as functions of the strength constant λ. The physical branch is reduced to a minute line element since ω_1 hardly depends on λ. Points are labeled by the value of λ. Some points, between parentheses, explicitly violate rule (ii) described at the end of Sec. III. Most points on the two "additional" branches violate rule (v).

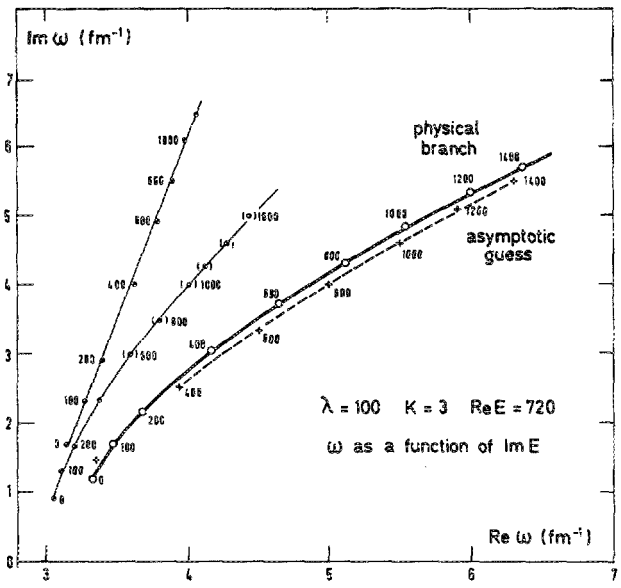


FIG. 3. Behavior of the roots as functions of their label, Im E. The asymptotic guess is generated by rule (iv).

V. DISCUSSION AND CONCLUSION

Our work would be complete if we could compare, for $\lambda \neq 0$, the exact amplitude D with the approximate $\bar{D}$ generated by ω_1. We have indeed easily calculated $\bar{D}$, but the calculation of D is another matter. This four-body problem, however, is not completely out of reach, for the separability of v leads to great simplifications. We are considering the comparison of D and $\bar{D}$ for a further paper.

The main result of the present paper is thus the rich nature of the mean-field approximation in this model, because of the occurrence of 29 competing solutions. Even though most of them can be rapidly ruled out, one day there should be a physical interpretation for most of them.

It is extremely satisfactory that finally only one among these roots survives as the undisputed physical solution, with great stability and simplicity properties (in particular, no branching). This gives great confidence in the validity of the variational method in those practical cases where no analytical solution is available and the mean-field equations, Eqs. (1.6), have to be solved via only a numerical iteration.

It may be pointed out, incidentally, that the present model could also be used to increase the efficiency of various iteration schemes and then define the attraction basins of the various solutions. In the case of Ref. 3 and in the present model, we have established that ω_1 has in general good stability under various iterative approximation schemes, but the solution $\omega_2 \simeq -\omega_1^*$ also has a non-negligible stability domain, hence rules (i)–(v) always retain their interest.

In summary, there exists in our mean-field theory of collisions a nontrivial and soluble model that provides all desirable tests for the validity of our theory.

ACKNOWLEDGMENT

It is a pleasure for two of the authors (SK and AW) to thank Service de Physique Théorique, Saclay, for its hospitality during this work.

[1]B. G. Giraud, M. A. Nagarajan, and I. J. Thompson, Ann. Phys. (NY) **152**, 475 (1984).
[2]B. G. Giraud, M. A. Nagarajan, and C. J. Noble, Phys. Rev. A **34**, 1034 (1986).
[3]B. G. Giraud, Physica D **19**, 112 (1986).
[4]B. G. Giraud, M. A. Nagarajan, and A. Weiguny, Phys. Rev. C **35**, 55 (1987).

Comment on the paper "Theta series and magic numbers for diamond and certain ionic crystal structures" [J. Math. Phys. 28, 1653 (1987)]

George L. Hall
Department of Physics, North Carolina State University, Raleigh, North Carolina 27695-8202

(Received 4 August 1987; accepted for publication 11 May 1988)

Sloane's work [J. Math. Phys. **28**, 1653 (1987)] is supplemented by published theorems that have been applied to the NaCl, CsCl, ZnS, and CaF_2 structures to give formulas for the radii of all occupied shells (i.e., all radii for which $S_n = 0$ are identified) and to give the species of atoms in each shell. In this paper the sort of information these theorems can provide is briefly reviewed and some predictions based on them are checked with the tabulated results of Sloane and others.

I. INTRODUCTION

When coupled with collaborative work with Teo,[1,2] a recent paper by Sloane[3] provides extensive information about the radii $(n)^{1/2}$ and coordination numbers S_n of "shells," and about nuclearity (magic numbers) of spherical clusters of lattice points associated with "spheres" in multidimensional lattices, with the spheres centered about various lattice and nonlattice points. Connections with theta functions are established and exploited in the enumerative and combinatorial problems. They discuss possible applications in many fields, but for detailed illustrations and tabulation of results they emphasize two-dimensional (2-D) and three-dimensional (3-D) lattices with the lattice points occupied by one or two species of atoms. They appear to be unaware of similar work for the 3-D case reported by Kihara and Koba,[4] Wood,[5] and Wiley and Seman,[6] but their tabulated results are more extensive than those of Refs. 4 and 5. However, the tabulated results of Ref. 6 for diamond and zinc blende are more extensive than those of Sloane.[3] Clearly, all this work possesses a collective importance that is furthered by the Sloane–Teo discussions of applications and the connections they make with theta functions. The same could be said about the additional supplementing work I shall address, which is of a rather different nature.

The main purpose of the present comment is to point out that the Sloane–Teo work for the 3-D case is supplemented by earlier work[7,8] providing six theorems that follow with a little analysis added to Waring's[9] theorem. For the particular cases studied, this work gives expressions for Sloane's $(n)^{1/2}$; if any S_n are zero, it gives a *formula* for the *infinite* set (i.e., necessary and sufficient conditions are given for the infinite set of *occupied* shells); and it discusses the sequences of spherical shells of atoms in the NaCl, CsCl, diamond, zinc blende, and CaF_2 structures. Because of the role of theta functions in the Sloane–Teo work, this earlier work also has bearing on the theory of theta functions, though that will not be explored here. I shall attempt to give the reader a little of the flavor of the work on sequencing and to show how the tabulated results of Refs. 1–6, for a finite number (but large in some cases) of shells, and the sequencing formulas, for an infinite number of shells, can serve as mutual checks.

II. THE FACE-CENTERED-CUBIC (fcc) LATTICE

In 1952 Kihara and Koba[4] published a table of coordination numbers and radii for 60 shells of neighbors to the origin (lattice site) in an fcc lattice. In the table there is a conspicuous absence of the values of 14, 30, 46, 56, and 62 for $(r/d)^2$, where r is the radius and d the first neighbor distance. Theorem 1 of Refs. 7(a) and 8, first reported in 1957 in Ref. 7(b), gives the radii of nonempty shells as

$$d(k)^{1/2}, \quad k = 0,1,2,3,\ldots,$$

$$k \neq 4^a(16b + 14), \quad a,b = 0,1,2,3,\ldots, \qquad (1)$$

showing that the above numbers simply correspond to the first few "empty" (Sloane's $S_n = 0$) shells of an infinite set. Sloane and Teo[1] give in their Table 11 information on a larger number of shells, including one additional empty shell predicted by my Eq. (1) with $a = 0$, $b = 4$, yielding $k \neq 78$, which in turn corresponds to $n \neq 156$ of Sloane and Teo.

Here the deviation from the sequencing of shells established by the first few shells starts at the fairly small value $k = 14$. In other cases, such as that of diamond and zinc blende, the initial pattern is maintained to surprisingly large radii.

III. THE NaCl AND CsCl STRUCTURES

The NaCl *structure* is fcc, which may be viewed in terms of a fcc lattice of positive (and one of negative) ions interpenetrating so as to fill a simple cubic (sc) lattice of points. Clearly, if d is replaced in Eq. (1) with c, used to denote the distance between two neighboring ions of like sign, then the modified equation gives the radii of positive (negative) ions about a positive (negative) site. However, another theorem is needed to give the full interlacing pattern. This is provided by Theorem 2 of Refs. 7(a) and 8 [first reported in Ref. 7(b)]. It shows that the shells of the negative (positive) ions about a positive (negative) site have radii given by

$$c(k - \tfrac{1}{2})^{1/2}, \quad k = 1,2,3,\ldots,$$

$$k \neq 4m, \quad m = 1,2,3,\ldots. \qquad (2)$$

Thus one sees immediately that no shell ever possesses both types of ions and that the initial sequencing, say, about a positive site, of $+ - + - + -$ is not continued. The pat-

 0022-2488/88/092090-02$02.50

tern of + 's and − 's becomes very irregular as the radius increases, but it is completely describable by Eqs. (1) and (2).

A consistency check on the above results is provided by Theorem 3 of Refs. 7(a) and 8. Whereas the proofs of Theorems 1 and 2 required Waring's[9] theorem plus a little analysis, Theorem 3 is simply Waring's theorem expressed in terms of a sc lattice. It states that every atom in a sc lattice lies on one of the set of nonempty shells, taken about a lattice point, whose radii are given by

$$d(k)^{1/2}, \quad k = 0,1,2,3,\ldots,$$
$$k \neq 4^a(8b+7), \quad a,b = 0,1,2,3,\ldots, \qquad (3)$$

where d is the first neighbor distance in the sc lattice. By separating the even and the odd values of k in Eq. (3), it is easy to show that the complete set of filled shells given by Eqs. (1) and (2) is identical to the set given by Eq. (3).

In a similar manner it is shown in Ref. 7(a) that the CsCl initial pattern of + − + + − + + is not continued indefinitely, and the precise pattern of the sequencing of shells of each type is given.

IV. DIAMOND AND ZINC BLENDE STRUCTURES

These structures may be viewed in terms of two interpenetrating fcc lattices with one displaced with respect to the other along, and one quarter of the length of, the principal diagonal of the basic cube (the cube containing four lattice points of each fcc sublattice). In the diamond structure both sublattices are occupied by one chemical species. In the zinc blende structure the two sublattices are occupied by different species, which I shall denote with A's and B's. Clearly, the diamond structure is simply a degenerate case of the zinc blende where the A's and B's have become identical.

Let the origin be taken first at a lattice site of the A sublattice. Denote the edge length of the basic cube of this sublattice by S so its first neighbor distance is given by $S/(2)^{1/2}$. Then Ref. 8 shows that the radii of the A shells about an A origin are given by

$$S(k/2)^{1/2}, \quad k = 0,1,2,3,\ldots,$$
$$k \neq 4^a(16b+14), \quad a,b = 0,1,2,3,\ldots, \qquad (4)$$

and the radii of B shells about an A origin are given by

$$(S/4)(8i-5)^{1/2}, \quad i = 1,2,3,\ldots. \qquad (5)$$

To connect these with the notation of Sloane's[3] Table I for diamond, rewrite these, respectively, as

$$(S/2)(n)^{1/2}, \quad n = 0,2,4,\ldots,$$
$$n \neq 4^a(32b+28), \quad a,b = 0,1,2,3,\ldots, \qquad (4')$$
$$(S/2)(n)^{1/2}, \quad n = 2j + (3/4), \quad j = 0,1,2,3,\ldots. \qquad (5')$$

Connections with the notation used in Table VIII of Wiley and Seman[6] may be similarly found.

Notice that the sequence of occupied shells about an A site goes as $ABABAB\cdots$ for a considerable distance before the first "forbidden" shell, given by Eq. (4') with $n = 28$, interrupts the sequence. It might be surprising to some readers that a sequence established by the first 27 shells is finally broken. Table VIII of Wiley and Seman displays the next two forbidden shells corresponding to $n = 60$ and $n = 92$ in Eq. (4') above.

V. THE CaF_2 STRUCTURE

This structure is somewhat related to, but much more complicated than, the zinc blende structure. It may be viewed as composed of three interpenetrating fcc sublattices with one containing A's and the two others containing B's. With Theorem 6 of Ref. 8 it is possible to sort out the sequencing of shells. Detailed expressions are given for the radii of filled A shells about an A site, B shells about an A site, A shells about a B site, B shells about a B site on the same sublattice, and B shells about B sites on a different sublattice. The reader is referred to Ref. 8 for details on the sequencing expressions, to Table XII of Wiley and Seman for extensive tabulated results, and to Sloane[3] for theta functions associated with this structure.

[1] N. J. A. Sloane and B. K. Teo, J. Chem. Phys. **83**, 6520 (1985).
[2] B. K. Teo and N. J. A. Sloane, Inorg. Chem. **25**, 2315 (1986).
[3] N. J. A. Sloane, J. Math. Phys. **28**, 1653 (1987).
[4] T. Kihara and S. Koba, J. Phys. Soc. Jpn. **7**, 348 (1952).
[5] V. E. Wood, Am. J. Phys. **33**, 632 (1965).
[6] J. D. Wiley and J. A. Seman, Bell Syst. Tech. J. **49**, 355 (1970).
[7] (a) G. L. Hall and D. O. Christy, Am. J. Phys. **34**, 526 (1966); (b) G. L. Hall, J. Phys. Chem. Solids **3**, 210 (1957).
[8] G. L. Hall, Am. J. Phys. **50**, 653 (1982). Erratum: A typographical error was made in the statement of Theorem 5 in transcribing it from its Ref. 3, which is Ref. 7(a) here. Above Eq. (A5) the interstitial position should be specified as $d(\hat{x}+\hat{y})/2$.
[9] G. H. Hardy and E. M. Wright, *Theory of Numbers* (Oxford University, London, 1954), 3rd ed., p. 310.

The Φ^4 equations of motion. I. A new constructive method: The Φ iteration

Marietta M. Manolessou
Institut des Hautes Etudes Scientifiques, 35, route de Chartres, 91440 Bures-sur-Yvette, France

(Received 5 November 1987; accepted for publication 6 April 1988)

A new method for the construction of a nontrivial Φ_4^4 model consistent with the general principles of a Wightman quantum field theory is proposed. The renormalized equations of motion for the connected Green's functions in the Euclidean momentum space are considered. Using a fixed-point method in an appropriate Banach space, the existence of a unique nontrivial solution of these equations is proved when the coupling constant Λ is fixed *positive* and smaller than a *finite* value. As a first step that avoids the difficulties of the renormalization operation, the corresponding problem in two dimensions is solved first. Also, in order to deal with all technical complexities that stem from the purely combinatorial nature of the equations, the corresponding zero-dimensional problem was studied previously. In all cases, it has been proved that the nontrivial solution exists inside particular subsets of the corresponding Banach space, characterized by the *alternating signs* and the factorization (or "splitting") properties of the Green's functions at zero external momenta. These properties first appeared "experimentally" by the iteration of the two-dimensional system of equations called "Φ iteration" in the paper, and have been crucial for the conservation of the norms and for the contractivity of the nonlinear mappings.

I. INTRODUCTION: A NEW CONSTRUCTIVE METHOD

In quantum field theory language, when a Lagrangian of four interacting scalar fields $\Phi(x)$ is considered, the dynamical equations that describe the interaction mathematically in the four-dimensional Minkowski space with coordinates $x \equiv \{\vec{x}\in\mathbb{R}^3,\ x^0\in\mathbb{R};\ \|x\| = [x^{02} - \vec{x}^2]^{1/2}\}$ are (a) a nonlinear differential equation ("equation of motion") of the form[1]

$$-(\Box + m^2)\Phi(x) = \frac{\Lambda}{\rho_1 + \rho_2}(:\Phi^3(x): - \rho_0\Phi(x)), \tag{1.1a}$$

and, simultaneously, (b) the conditions of "quantization" of the field $\Phi(x)$ expressed by the commutation relations

$$[\Phi(x),\Phi(y)] = [\dot{\Phi}(x),\dot{\Phi}(y)] = 0, \tag{1.1b}$$

$$[\Phi(x),\dot{\Phi}(y)] = [i\rho_1/(\rho_1 + \rho_2)]\delta_3(\vec{x} - \vec{y}), \quad x^0 = y^0. \tag{1.1c}$$

Here m and Λ are the physical mass and coupling constant of the interaction model, and ρ_0, ρ_1, ρ_2 are physically well-defined given quantities associated to this model, the so-called renormalization constants. For the precise definition of the normal product $:\Phi^3(x):$, we refer the reader to Ref. 1.

From these equations (1.1) one can formally derive an equivalent infinite system of equations of motion for the Green's functions (or the "vacuum expectation values") of the theory.[1]

The present paper is the first of a series of three papers concerning the study of the Φ^4 equations of motion for the Green's functions in their Euclidean version. These papers constitute the second step of a general program we started some years ago in order to derive a new method which could ensure the existence of a Φ_4^4 nontrivial Wightman quantum field theory.[1] This method is different in approach from the work done in the constructive quantum field theory framework of Glimm–Jaffe[2] and others.[3]

Several authors have already treated the problem of existence or other crucial features of a field theory by the study of the equations of motion of particular models of interacting fields. The corresponding theoretical framework has not always been the same. There exist several attempts and results in both constructive[4,5] and perturbative[6,7] quantum field theory. In particular, the first idea of an existence proof by fixed-point methods applied on the equations of motion comes from Taylor.[8]

The fundamental elements of the present study are the infinite sequences of C^∞ functions on

$$\mathscr{E}^{rn}_{(q)} \times \mathbb{R},\ H \equiv \{H^{n+1}(q,\Lambda)\}_{n=2k+1,\ k\in N},$$

which formally represent the connected, completely amputated Green's functions (Schwinger functions). Here

$$q = \{q_i = (\mathbf{q}_i\in\mathbb{R}^{r-1}, q_i^0\in\mathbb{R}),\ \|q_i\| = |q_i^{02} + \mathbf{q}_i^2|^{1/2};\ 1\leqslant i\leqslant n\}$$

is the set of independent r momenta, and $\mathscr{E}^{rn}_{(q)}$ means the corresponding Euclidean momentum space; $\Lambda\in\mathbb{R}$ with Λ the coupling constant.

We can briefly describe our approach as follows: Using a fixed-point method in a appropriate Banach space, we try to demonstrate (under some sufficient conditions on the coupling constant Λ) the existence of a unique nontrivial solution $\overline{H} \equiv \{\overline{H}^{n+1}(q,\Lambda)\}_{n=1,3,\ldots}$ of the Φ_4^4 system of equations of motion established in terms of H sequences. Then we intend to prove that this solution $\overline{H}$ verifies all the Osterwalder–Schrader axioms,[9] which are the Euclidean equivalent system of the Wightman axioms in complex Minkowski space. This last step ensures (in view of the reconstruction theorem[9,10]) that the sequence $\overline{H}$ is no longer formal but a well-defined infinite sequence of Green's functions equivalent to a nontrivial axiomatic quantum field theory (AQFT). In order to accomplish the above purposes we have developed a program that is composed of the following four main steps.

 0022-2488/88/092092-14$02.50

(1) (a) We started from the definition of the renormalized normal product (RNP) in the AQFT framework by using the synthesis of two previously elaborated ideas: the perturbative composite operators of Zimmermann[11] and Lowenstein,[12] and the renormalized G-convolution product of Refs. 13 and 14. We then established (formally) finite equations of motion for the connected Euclidean Green's functions, in $0\leqslant r\leqslant 4$ dimensions, for a Lagrangian Φ^4 model [cf. (1.1)] where this RNP has been supposed to represent the corresponding interaction term.

(b) The consistency of this construction with the general AQFT principles has also been ensured by proving the conservation of linear AQFT properties, in $0\leqslant r\leqslant 4$ dimensions, by the extension of the previously mentioned equations in complex Minkowski momentum space.

These results (a) and (b) have been published in Refs. 15 and 16, respectively.

(2) The studies in step (1) are encouraging as far as the consistency of the scheme is concerned, but they remain formal in this sense: They do not give any answer to the fundamental question concerning the existence of a nontrivial Φ^4 Wightman theory defined by an infinite sequence $\overline{H}\equiv\{\overline{H}^{n+1}\}_n$ of Green's functions which is the solution of the dynamical equations. So, before trying to look for any answer in four dimensions, we carried out the study of the Φ^4 equations of motion in smaller dimensions ($0\leqslant r\leqslant 2$) in order to avoid any difficulties due to the renormalization operation, which is nontrivial for $3\leqslant r\leqslant 4$.[15] In other words, the second step of the program is the formulation of the fixed-point method through the study of the zero- and two- (or one-)dimensional systems of equations, presented in the three papers we shall designate I (the present one), II (Ref. 17), and III (Ref. 18).

The final result we obtain can be summarized as follows.

(a) The construction of a nontrivial solution for each one of the corresponding systems of equations is achieved.

(b) This solution is uniquely defined in terms of characteristic sign, namely, $(-1)^{(n-1)/2}$, and the factorization properties (the so-called splitting properties) of the Schwinger functions H^{n+1} and all their convolution products, at zero external momenta. We can describe these properties in a unifying and simple way. When $0<\Lambda\lesssim 0.006$,

$$H^2(q=0,\Lambda)=1+\delta_1(\Lambda)\Lambda, \quad \text{with } \delta_1(\Lambda)\sim 6\Lambda, \tag{1.2}$$

and $\forall n=3,5,\ldots$,

$$H^{n+1}(q=0,\Lambda) \sim -\delta_n(\Lambda)n^2H^{n-1}(q=0,\Lambda)H^2(q=0,\Lambda), \tag{1.3}$$

with

$$\delta_1(\Lambda)\leqslant\delta_n(\Lambda)\leqslant\delta_\infty^\Lambda<\infty$$

and (1.3a)

$$\lim_{\Lambda\to 0}[\delta_n(\Lambda)/\Lambda]\sim n^2.$$

The sequences $\delta\equiv\{\delta_n(\Lambda)\}$, called the splitting constant sequences, are the building blocks of the method. The bounds automatically implied by these properties are

$$\begin{aligned}&|H^2(q,\Lambda)|(q^2+1)^{-1}\leqslant c_0, \quad 2\leqslant c_0<\infty,\\ &|H^{n+1}(q,\Lambda)|\leqslant[n!]^2c_0^n, \quad \forall n\geqslant 3.\end{aligned} \tag{1.4}$$

(2[I]) The intuition that guided us to the method developed in I–III was the following: The information concerning the specific features of a dynamical system of interacting fields should be most naturally obtainable through an iteration of the renormalized equations of motion with the free solution as starting point. This is what we have called the "Φ iteration." The investigation of the properties of this iteration procedure in the two-dimensional case (trivial renormalization, nontrivial convolution integrals), these properties being presented in Sec. II of the present paper I, has pointed out to us what specific features must characterize the space of $\{H^{n+1}\}_n$ sequences in which the fixed-point method had to be constructed. In other words, the exploration of the detailed structure of the Φ iteration has brought out the particular properties of the different global terms of the equations at every order ν, constructed in terms of $H_\nu^{n+1}(q,\Lambda)$ functions. These properties, which essentially are signs, splitting properties, and bounds at zero external momenta, form a self-consistent system of conditions conserved by the Φ iteration. They imply precise norms of the Green's functions [cf. (1.4)] of H sequences, which in turn are conserved and automatically ensure the convergence of the Φ iteration to the solution.

(2[II]) and (2[III]) Once the information concerning the detailed structure of the space of H sequences was obtained, the answer to the problem of how to construct the fixed-point method became clear. It would be sufficient at first to define a Banach space using the norms provided by the Φ iteration. One should then seek a fixed point of the equations of motion, by application of the contractive mapping principle, inside characteristic subsets of this space, which exactly imitate the fine structure of the Φ iteration at zero external momenta.

A detailed study of the combinational features of the global terms of the equations at fixed external momenta appeared then as a necessary starting point for the whole technique. For this reason, we based the mathematical formalism of the approach on the study of the zero-dimensional problem, where there is no momentum dependence. In this case, the Green's functions depend only on Λ and n, i.e., $H_0\equiv\{H_0^{n+1}(\Lambda)\}_n$, all complications due to the convolution integrals are absent, and the problem becomes one of a purely combinatorial character (at fixed $\Lambda\in\mathbb{R}^+$). Related to this adequately simpler form of the zero-dimensional case is the introduction of a basic combinatorial notion provided by the study of the zero-dimensional problem. It deals with the sequences of the so-called sweeping factors $\beta_n(\Lambda)$, $\alpha_n(\Lambda)$. These quantities "sweep" all terms of the two ordered sums representing the global terms of the equations, called C^{n+1} and B^{n+1}, respectively [cf. Eqs. (3.1)], and replace them by only one term proportional to the dominant contribution. The combinatorial information they carry about the terms of the sums they sweep is reflected in their explicit dependence on n^2 and n, respectively. In terms of these sweeping factors (in particular, β_n's) the splitting properties and the appro-

priate subsets are formulated in a simpler and mathematically consistent way. In other words, by using extensively the definition and the properties of the sweeping factors, we give in paper II the detailed presentation of the novel technique (cf. Sec. II of II), the solution of the zero-dimensional problem (cf. Sec. III of II), and the solution of the two-dimensional problem (cf. Sec. IV of II). On the other hand, paper III contains the proof of all the above-mentioned combinatorial properties of the sweeping factors β_n and α_n, respectively.

Another useful aspect of the zero-dimensional problem is the fact that it provides a direct way to test the validity of the method. More precisely, a comparison is possible between (i) the solution we obtained using the system of equations derived through the RNP definition (in zero dimensions), and (ii) the solution of the corresponding system of zero-dimensional equations that are obtained by the standard functional integral formalism.[19] Such a comparison was performed numerically, and the result is the perfect coincidence of the two solutions associated with the two different methods. This is a first confirmation of the method. In Appendix V of III we present precisely these verifications, which are essentially based on the numerical results of Voros.[20]

(3) In the third step of the program we apply the method previously developed in (2[I–III]) in order to solve the four- (and three-)dimensional system of equations of motion. This means that we look again for a nontrivial fixed point of the equations considered as a nonlinear mapping on the appropriate Banach space. This fixed point satisfies also all the nice properties of signs and splitting (mentioned above) at zero external momenta and norms.

However, the renormalization procedure now provides some new ingredients that present two different aspects.

(a) The structure of the global terms containing *a priori* divergent convolutions and, therefore, nontrivial renormalization operators [automatically present in view of the definition of RNP in step (1)] needs a finer construction of the subsets characterized by the signs and splitting properties. In other words, in view of the nontrivial renormalization some supplementary conditions appear, which imply a deeper structure of the particular subsets. These conditions are necessary in order to ensure the stability of the subsets under the action of the mapping and, consequently, the convergence of the mapping to a good solution.

(b) The renormalization parameters defined formally in Ref. 15 associated to the solution of the Φ_4^4 system have to be described and interpreted as well-defined physical quantities.

The study concerning this step and the results obtained will appear in Ref. 21.

(4) As we mentioned at the beginning, the verification by the solution of the Osterwalder–Schrader axioms[9] must be the step that will complete the program towards the construction of an AQFT model. Except for the positivity and cluster properties all the other axioms—the distribution property, Euclidean covariance, and symmetry—are directly implied by the renormalized equations of motion Φ_4^4 (cf. also Ref. 16). Concerning the positivity property there exist only some partial results of the author relative to the four-point function, which makes one optimistic for a (nontrivial) extension. This subject is under investigation.

(4′) Once these results on the Osterwalder–Schrader positivity are established, one should try to understand the reason for the "contradiction" between our nontrivial solution of Φ_4^4 obtained in step (3) and the arguments of Aizenmann and Fröhlich[22] concerning the triviality of analogous Φ_4^4 models obtained as limits to the continuum of lattice field theory models.

The most apparent reason for this contradiction seems, at this moment, to be the fundamental difference in the nature of the two approaches. In our method there is no "artificial" regularization in terms of a lattice approximation. The renormalization (and all information concerning the continuum!) is introduced in a natural way as an initial condition at the starting point of the whole construction [cf. step (1)]. We give an example that makes evident this character of the method. In two (or one) and zero dimensions one notices that the presence of the renormalization implies the absence of the terms proportional to Green's functions represented by Fig. 1.

These terms, *a priori* divergent even in two dimensions, are annihilated thanks to the renormalization procedure.[15] On the other hand, the same terms are present in the corresponding zero-dimensional equations of motion automatically obtained by the standard functional integral formula.[19] (The reader is referred to Appendix V of III for a more detailed discussion of this point.) These last considerations complete the presentation of the total program.

We now conclude the introduction with the plan for this paper (I).

In Sec. II we introduce the $\Phi_{1,2}^4$ system of equations of motion [derived from step (1)] and all general convolution products (the so-called Φ convolutions) generated by an arbitrary iteration of the fundamental operations appearing in these equations. Next we describe in detail the Φ-iteration procedure presented in step (2[I]) above and explain why it converges. We then show the relation of the sum representing H_ν^{n+1} (fixed n) at a given order ν of the Φ-iteration procedure with the corresponding sum up to order ν of the perturbative series, in order to give an answer to the following problem.

Several years ago Jaffe[23] proved the divergence of the perturbative series of a Φ_2^4 model. So a two-part question immediately arises: Does the Φ iteration generate the perturbation series exactly? If yes, then is there any contradiction between the divergence of the perturbative expansion and the convergence of the Φ iteration?[24]

Following the "experimental" results coming from the first three orders of the Φ iteration (cf. Appendix Φ of III), and following our discussion in Sec. II, which generalizes these results (cf. Proposition 2.2), the total sum of the per-

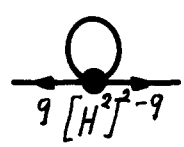


FIG. 1. Graphical representation of the divergent terms that are annihilated thanks to the renormalization procedure.

turbative series up to order ν is present in the expansion of H_ν^{n+1} by the Φ iteration. But apart from this there is another sum added to the first, with the appropriate signs and coefficients, that contains larger orders of the perturbative series. It is then clear that the fundamental difference between the two approaches comes from the different way in which the polynomials in Λ are arranged and summed in each of the two approximations. So automatically there is no contradiction if the Φ iteration converges to a nontrivial solution, despite the divergence of perturbation theory.

Section III is devoted to the presentation of the technical basis of our method. We first describe the zero-dimensional problem and introduce the fundamental combinatorial notions mentioned in step (2), namely, the sweeping factors. Then the splitting properties appearing through the Φ iteration are redefined in terms of the sweeping factors for both zero- and two- (or one-)dimensional cases.

Finally, by means of two theorems presented without proof (their demonstration being the aim of II) we state our results concerning the $\Phi^4_{0,1,2}$ equations of motion. These results constitute in fact a mathematically more elegant (and simpler!) reformulation of the proof of the Φ iteration's convergence to the solution.

II. THE Φ-CONVOLUTION PRODUCTS AND THE Φ ITERATION

This section contains essentially the presentation of the Φ iteration (mentioned in the Introduction) with all the detailed information we obtained "experimentally" using this procedure, together with its relation to the perturbation theory.

We shall distinguish two parts to this section.

(i) We start with some fundamental notation and definitions concerning the equations of motion of a Φ^4_2 model and the corresponding (formal) Schwinger functions $H^{n+1}(q)$ ($q\in\mathscr{E}^{2n}_{(q)}$; $\mathscr{E}^{2n}_{(q)}\equiv$Euclidean two-momentum space) presented in detail in Ref. 15, and define the appropriate space $\mathscr{B}$ of the sequences $H\equiv\{H^{n+1}\}_n$ with $n=1,3,5,\ldots$. Furthermore, we introduce the notion of Φ-convolution products. We define the three elementary operations of Φ^4 type that are exactly the special convolutions coming from the global terms of the equations of motion and their respective graphical representations.

By repeated application of these operations, we obtain recursively an arbitrary composition of these fundamental convolutions and arrive at general Φ-convolution products (ΦC's) with their corresponding graphs G_Φ. Finally, we define sequences of ΦC's appropriate for the convergence proofs of the method, which we call coherent sequences of ΦC's (cf. Sec. III and II).

(ii) In the second part of the section we shall describe the structure and the features of the Φ iteration in terms of the above-mentioned ΦC's at zero external momenta.

We end the section by showing recurrently (cf. Proposition 2.2) that at a given order ν of the Φ iteration, and for fixed $n=1,3,5,\ldots$, all terms up to order $\sim\Lambda^\nu$ of the corresponding perturbative series are contained in the corresponding expression H_ν^{n+1}, and we discuss this point.

A. Basic notation

In all of what follows, $n=2k+1$, $k\in\mathbb{N}$, denotes the characteristic number of every Green's function.

Definition 2(a): In the space of C^∞ functions $\eta^{n+1}(q)$, $q\in\mathscr{E}^{2n}_{(q)}$, and $n=2k+1$, $k\in\mathbb{N}$, which formally represent the connected Euclidean Green's functions in momentum space and satisfy all the Osterwalder–Schrader axioms,[9] we consider the equations of motion established in Ref. 15 (cf. Eq. 2.40 of Ref. 15). We set $\beta=0$, $\gamma=1$, and $m_0^2=m^2-\Lambda\alpha\overset{\text{def}}{\equiv}1$. (Here m_0^2 is the bare mass, m^2 the physical mass, Λ the coupling constant, and $\Lambda\alpha\equiv\delta m^2$ the mass renormalization.) We have

$$\begin{aligned}(P^2_{(n)}+1)\eta^{(n+1)}(P_{(n)},q) &= -\Lambda\Big\{\int\eta^{n+3}(q,k_1,k_2)dk_1\,dk_2 \\ &\quad+3\sum_{\omega_n(J)}\int\eta^{(j_2+2)}(q_{j_2},k)dk\,\eta^{(j_1+1)}(q_{j_1}) \\ &\quad+6\sum_{\omega_n(I)}\prod_{l=1}^{3}\eta^{(i_l+1)}_{(q_{i_l})}\Big\}+\delta_{1,n},\end{aligned}\tag{2.1}$$

where

$$P_{(n)}=\sum_{p=1}^{n}q_i,\quad q_i\in\mathscr{E}^2_{(q)},\quad \Lambda\in\mathbb{R}.\tag{2.1a}$$

Here $\{k_1,k_2\}\in\mathscr{E}^4_{(k)}$ (resp. $k\in\mathscr{E}^2_{(k)}$) means the set of integration variables in the first integral (resp. in the second integral), and $\omega_n(J)$ [resp. $\omega_n(I)$] denotes the following partitions of the set $q=\{q_1,\ldots,q_n\}\in\mathscr{E}^{2n}_{(q)}$:

$$\omega_n(J)=\Big\{q_{j_l}\subset q,\ l=1,2,\ j_1=2k+1,\ k\in\mathbb{N};\ q_{j_l}\cap q_{j_{l'}}=\emptyset,\ \bigcup_{l=1}^{2}q_{j_l}=q\Big\},\quad |q_{j_l}|=j_l,\quad (j_1+j_2=n),\tag{2.2a}$$

$$\omega_n(I)=\Big\{q_{i_{\bar l}}\subset q,\ i_{\bar l}=2k+1,\ k\in\mathbb{N},\ 1\leqslant\bar l\leqslant 3,\ q_{i_{\bar l}}\cap q_{i_{\bar l'}}=\emptyset\ (\bar l\neq\bar l'),\ \bigcup_{\bar l=1}^{3}q_{i_{\bar l}}=q\Big\},\quad \Big\{|q_{i_{\bar l}}|=i_{\bar l},\ \sum_{\bar l=1}^{3}i_{\bar l}=n\Big\}.\tag{2.2b}$$

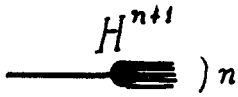


FIG. 2. Graph of H^{n+1}.

Moreover, we shall use below the notation $\omega_n(J) \equiv (j_1 j_2)$, $\omega_n(I) \equiv (i_1 i_2 i_3)$, and

$$Q_{J_l} \equiv \sum_{q_l \in q_{j_l}} q_l, \quad Q_{I_l} \equiv \sum_{q_{l'} \in q_{i_l}} q_{l'}. \tag{2.2c}$$

We now consider the (formal) completely amputated Green's functions with respect to the free propagators (expressed in the bare mass), denoted by

$$H^{n+1}(q) = \prod_{i=1}^{n+1} \Delta_F^{-1}(q_i)\eta^{n+1}(q), \tag{2.3}$$

where

$$q_{n+1} = \sum_{i=1}^{n} q_i, \quad \Delta_F(q_i) \equiv 1/(1+q_i^2) \tag{2.3}$$

(cf. Fig. 2). We then write Eqs. (2.1) in an equivalent way:

$$H^2(q,\Lambda) = -\Lambda[N_3 H^4](q,\Lambda) + q^2 + 1, \tag{2.4a}$$

and

$$H^{n+1}(q,\Lambda) = -\Lambda\Big\{[N_3 H^{n+3}] + 3\sum_{\omega_n(J)} [N_2 H^{j_2+2}][N_1 H^{j_1+1}] + 6\sum_{\omega_n(I)} \prod_{l=1}^{3} [N_1 H^{i_l+1}]\Big\}(q,\Lambda), \quad \forall n \geqslant 3. \tag{2.4b}$$

Here we introduce slightly abbreviated notation for the connected part of orders 3, 2, and 1, respectively, of the normal product (see Ref. 15, Proposition 2.2):

$$-\Lambda[N_3 H^{n+3}](q,\Lambda) \equiv -\Lambda\int H^{n+3}(q,k_1,k_2\Lambda) \prod_{l=1}^{3} \Delta_F(l_i)\,dk_1 \cdot dk_2 \tag{2.5}$$

(cf. Fig. 3),

$$-\Lambda[N_2 H^{j_2+2}][N_1 H^{j_1+1}](q,\Lambda) \equiv -\Lambda\Big\{\int H^{j_2+2}(q_{j_2}(k,q),\Lambda) \prod_{j=1}^{2} \Delta_F(l_j)\,dk\Big\} \times H^{j_1+1}(q_{j_1})\Delta_F(Q_{J_1}) \tag{2.6}$$

(cf. Fig. 4),

$$-\Lambda \prod_{l=1}^{3} [N_1 H^{i_l+1}](q,\Lambda) \equiv -\Lambda \prod_{l=1}^{3} H^{i_l+1} \Delta_F(Q_{I_l}) \tag{2.7}$$

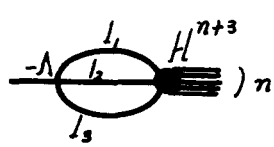


FIG. 3. Graph of $-\Lambda[N_3 H^{n+3}]$.

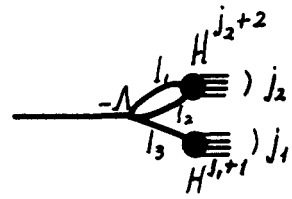


FIG. 4. Graph of $-\Lambda[N_2 H^{j_2+2}][N_1 H^{j_1+1}]$.

(cf. Fig. 5). We shall call "Φ^4-type operations" the convolutions defined by Eqs. (2.5)–(2.7).

In Figs. 2–5 the internal lines are associated with a free propagator $\Delta_F(l_i)$ [cf. Eq. (2.3)], where $l_i = l_i(q,k)$ is the momentum carried by the line i following the corresponding momentum assignment. Moreover, each "bubble" vertex (with $n+1$ incident lines—external or internal) is associated with one H^{n+1} Green's function, and the constant four-vertex is associated with the constant $(-\Lambda)$.

Definition 2(b) (The space $\mathscr{B}$ and the nonlinear mapping $\mathscr{M}$): We introduce the space $\mathscr{B}$ of the sequences $H \equiv \{H^{n+1}(q,\Lambda)\}_n$ of C^∞ functions H^{n+1} (with respect to $q \in \mathscr{C}^{2n}_{(q)}$ and $\Lambda \in \mathbb{R}$), which is described by the following bounds: $\forall H \in \mathscr{B}$, there exists a finite positive constant C_H such that, $\forall n \geqslant 1$,

$$|H^{n+1}(q,\Lambda)| \leqslant [n!]^2 C_H^n M_{n+1}(q), \quad M_{n+1}(q) = \begin{cases} q^2+1, & \text{if } n=1, \\ 1, & \text{if } n \geqslant 3. \end{cases} \tag{2.8}$$

The system of equations (2.4) defines a nonlinear mapping of $\mathscr{B}$ onto itself, $\mathscr{M}: \mathscr{B} \to \mathscr{B}$.

Definition 2(c) [The Φ-convolution products (ΦC's)]: Given any sequence $H \equiv \{H^{n+1}(q,\Lambda)\} \in \mathscr{B}$, we associate with it an infinite family $\mathscr{F}$ of sequences—convolutions that we call Φ-convolution products (ΦC's)—together with their corresponding graph representations, called Φ graphs. We shall define them recursively and call the final Φ graph G_Φ. We first construct the "primitive" sequences of ΦC's (resp. the corresponding disconnected primitive Φ graph) by application on H of one or more Φ^4-type operations defined by (2.5)–(2.7) (resp. represented by Figs. 3–5). Then, for every n, starting from a set of disconnected primitive convolution products (resp. disconnected primitive Φ graphs, which are subgraphs of G_Φ—cf. the Appendix), we label a subset of the corresponding external momenta (resp. a subset of external lines). By using these labeled external momenta as integration variables (resp. independent loops), we apply successively in a unique way one or more uniquely chosen Φ^4-type operations. This uniqueness characterizes

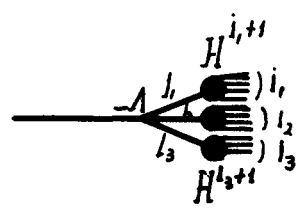


FIG. 5. Graph of $-\Lambda\Pi[N_1 H^{i_l+1}]$.

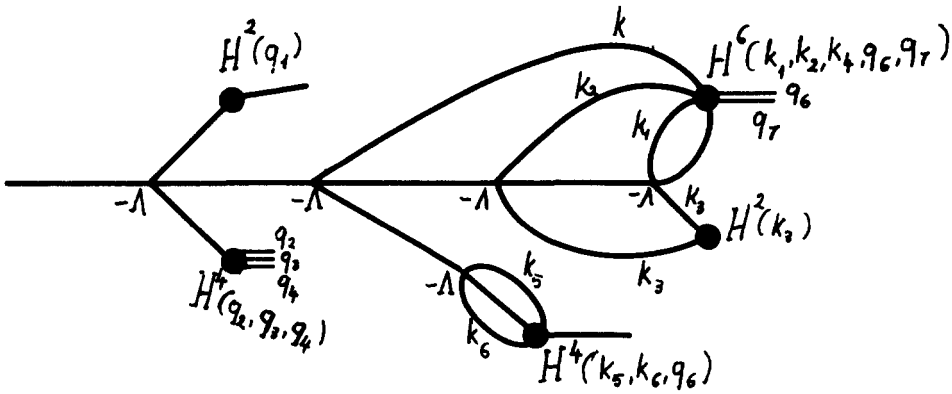


FIG. 6. Graphical representation of the ΦC $\Phi^5(H)H^6$.

each particular ΦC. In the Appendix we illustrate such a construction for an example (cf. the first through fifth steps successively). So from the initially disconnected components (resp. the primitive disconnected Φ graphs) we obtain (a) one more and more complex Φ-convolution product (resp. one more and more complicated Φ graph, a subgraph of G_Φ), and (b) the remaining disconnected initial primitive ΦC's (resp. Φ graphs).

The final step yields the general ΦC denoted by $\Phi^{(n)}(H)H^{n+1}$ (resp. the corresponding Φ graph G_Φ) by composition of the last one or two primitive disconnected components with the Φ-convolution product of the preceding step.

The general form of the G_Φ graph presents the following characteristics.

(i) Every vertex of G_Φ is either a simple four-vertex associated with the constant $(-\Lambda)$, called a "dynamical four-vertex," or a "bubble" vertex associated either with H^{n+1} itself or with one of the other H^{i_l+1} ($i_l = 2k+1$, $k\in\mathbb{N}$, $l = 1,...,m$, with m a fixed integer) Green's functions of the sequence H. These generalized vertices of G_Φ belong to one or more subgraphs of G_Φ, which are defined at a given step of the above recursive construction.

(ii) Given the set $q = \{q_1,...q_n\}\in\mathscr{C}^{2n}_{(q)}$ of independent momenta of $H^{n+1}(q)$ we consider partitions $\omega_n(\tilde{I},\tilde{J})$ of q defined by

$$\omega_n(\tilde{I},\tilde{J}) = \{q_{\tilde{I}}\subset q,\ k_{\tilde{J}}\subset q,\ q_{\tilde{I}}\cup k_{\tilde{J}} = q,\ q_{\tilde{I}}\cap k_{\tilde{J}} = \varnothing\}. \tag{2.9}$$

These partitions are such that the lines (i.e., momenta) *not* labeled at the starting point of the above recursive construction of G_Φ correspond to the subset $q_{\tilde{I}}$ (i.e., the remaining external momenta). The subset $k_{\tilde{J}}$ corresponds to the labeled lines and represents the set of integration independent variables (loops of G_Φ) of $\Phi^n(H)H^{n+1}$. Using a precise momentum assignment (cf. also Ref. 15) we define the corresponding momentum $\tilde{l}_i$ carried by each internal line i of G_Φ:

$$\tilde{l}_i \equiv \tilde{l}_i(q_{\tilde{I}},k_{\tilde{J}}). \tag{2.10}$$

So every internal line i of G_Φ corresponds to the free propagator $\Delta_F(\tilde{l}_i) = [\tilde{l}_i^2+1]^{-1}$. Finally, the above detailed description of G_Φ can be expressed in terms of the explicit recurrent synthesis of all corresponding operations, as is illustrated by the following formula for the ΦC $\Phi^{(5)}(H)H^6$, represented in Fig. (6):

$$\begin{aligned}&[\Phi^{(5)}(H)H^6](q_{\tilde{I}},\Lambda)\\ &\quad\equiv[N_1H^2][N_1[N_2[N_3[N_2H^6][N_1H^2]]]]\\ &\qquad\times[N_1[N_3H^4]][N_1H^4](q_{\tilde{I}},\Lambda).\end{aligned} \tag{2.11}$$

Evidently these analytical expressions are not always simple to use, so for practical reasons we often summarize them in the following global formula of the ΦC:

$$\begin{aligned}&[\Phi^{(n)}(H)H^{n+1}](q_{\tilde{I}},\Lambda)\\ &\quad\equiv(-\Lambda)^{|V|}\int_{G_\Phi}H^{n+1}(q_{\tilde{I}},k_{\tilde{J}})\Phi(H)\prod\Delta_F(\tilde{l}_i)dk_{\tilde{J}},\end{aligned} \tag{2.12}$$

where

$$\Phi(H) = \prod_{l=1}^{m}H^{i_l+1}(q_{\tilde{I}},k_{\tilde{J}},\Lambda),$$

where m = some fixed integer, which means the number minus one of all bubble vertices of G_Φ, and where V (resp. $|V|$) is the set of "dynamical" four-vertices associated with G_Φ (resp. the number of the Φ^4 operations defining the ΦC).

Sometimes, when a ΦC results from the composition of a primitive Φ^4-type operation with another more complex ΦC, then it is necessary to write it explicitly. For example, we write $\Phi^{(n)}[N_3^{(n+2)}H^{n+3}]$ or $\Phi^{(n)}[N_2^{(n-1)}H^{n-1}]\times[N_1H^2]$, etc.

Definition 2(d) (The coherent sequences of ΦC's): For later use, it is necessary that we introduce the so-called coherent sequences of ΦC's. We define them by reference to a fixed n, and we shall use the notation $\{\Phi^{(\bar{n},n)}(H)H^{n+1}\}_n$.

Let us consider an arbitrary ΦC constructed above [by Definition 2(c) and described by (2.12)] such that the first Φ^4-type operation that acts on the $H^{\bar{n}+1}$ function is of type $[N_2^{(\bar{n}-1)}H^{\bar{n}+1}][N_1H^2]$ (exactly like the example $[N_2^{(4)}H^6][N_1H^2]$ of Fig. (6). We then define the two types of coherent sequences (CS) of ΦC's associated with a fixed $\bar{n}\geqslant 3$ as follows.

CS type (1):

$$\text{(a)}\quad \Phi^{(\bar n,n)}_{(1)}(H)H^{n+1}\equiv\Phi^{(\bar n)}(H)\Big[\prod_{l=1}^{2}N_1^{(i_l)}H^{i_l+1}N_1^{(n)}H^{n+1}\Big],\quad \forall n\leqslant\bar n-2,\quad n+i_1+i_2=\bar n, \tag{2.13a}$$

$$\text{(b)}\quad \Phi^{(\bar n,\bar n)}_{(1)}(H)H^{\bar n+1}\equiv\Phi^{(\bar n)}(H)[N_2^{(\bar n-1)}H^{\bar n+1}][N_1H^2], \tag{2.13b}$$

$$\text{(c)}\quad \Phi^{(\bar n,n)}_{(1)}(H)H^{n+1}\equiv\Phi^{(\bar n)}(H)[N_3^{(\bar n+2)}[N_3^{(\bar n+4)}[\cdots[N_3^{(n-2)}[N_3^{(n)}H^{n+1}]]\cdots]]],\quad \forall n\geqslant\bar n+2. \tag{2.13c}$$

CS type (2):

$$\text{(a)}\quad \Phi^{(\bar n,n)}_{(2)}(H)H^{n+1}\equiv\Phi^{(\bar n)}(H)N_2{}^{(j_2)}H^{j_2+2}N_1^{(j_1)}H^{j_1+1},\quad \forall n\leqslant\bar n-2,\quad j_2+1=n,\quad \text{or } j_1=n \text{ and } j_1+j_2=\bar n, \tag{2.14a}$$

$$\text{(b)}\quad \Phi^{(\bar n,\bar n)}_{(2)}(H)H^{\bar n+1}\equiv\Phi^{(\bar n)}(H)[N_2^{(\bar n-1)}H^{\bar n+1}][N_1H^2], \tag{2.14b}$$

$$\text{(c)}\quad \Phi^{(\bar n,n)}_{(2)}(H)H^{\bar n+1}\equiv\Phi^{(\bar n)}(H)[N_3^{(\bar n+2)}[N_3^{(\bar n+4)}[\cdots[N_3^{(n-2)}[N_3^{(n)}H^{n+1}]]\cdots]]],\quad \forall n\geqslant\bar n+2. \tag{2.14c}$$

We notice that in the above definitions, except for the first Φ^4-type operation, the remaining global ΦC is the same throughout the total coherent sequence: $\Phi^{(\bar n)}(H)$. Moreover, the only difference that exists between the two types of sequences concerns their corresponding parts with $n\leqslant\bar n-2$.

Given a sequence $H\in\mathscr{B}$, we consider the corresponding infinite family $\mathscr{F}$ of coherent sequences of ΦC's. We can easily verify that every nontrivial subgraph of G_Φ associated with each of the ΦC's $\Phi^{(\bar n,n)}(H)H^{n+1}$ is absolutely convergent. More precisely, for every elementary subgraph of G_Φ with one loop, the "most divergent" configuration is that described by two free propagators (or two internal lines that link one bubble and one constant four-vertex). Let $\tilde\alpha(S)$ denote the asymptotic index of slow increase at infinity (the reader is referred to the precise Weinberg definitions in Ref. 13 or Ref. 14) of the integrand with respect to a subspace $S\subset\mathscr{C}^2_{(k)}$ ($k=$ the integral variable). In view of the assumptions (2.8) for every $H\in\mathscr{B}$, we always have

$$\tilde\alpha(S)\leqslant -4, \tag{2.15a}$$

so

$$\dim S+\tilde\alpha(S)\leqslant 2-4<0. \tag{2.15b}$$

Following Weinberg's criterion, (2.15b) ensures the convergence of the integral. An analogous result follows for more-than-one-loop subgraphs and G_Φ itself, which means that the absolute convergence of every ΦC is verified.

Moreover, by application of the second part of Weinberg's theorem, we can obtain directly the precise asymptotic behavior of every ΦC, as has been shown in detail in Ref. 15 for the primitive Φ convolutions. We state the above results in the following proposition.

Proposition 2.1: When $H\in\mathscr{B}$, every ΦC element of the corresponding family $\mathscr{F}$ of the first and second types of coherent sequences [(2.13) and (2.14)] converges absolutely and belongs to a Weinberg class of slow increasing functions in the Euclidean space of the set $q_{\bar I}$ of the external variables.

Definition 2(e) (Reductions): When some or all of the H^{n+1},H^{i_l+1} functions of definition (2.12) (resp. some or all of the bubble vertices of G_Φ) are replaced by the constant 1 (resp. by a simple point vertex), then we obtain the *partially* or *completely reduced* ΦC's (resp. the partially or completely reduced graph) and denote it by $\tilde\Phi^{(\bar n,n)}$ or $[\Phi^{(\bar n)}(H)N_3]\tilde{}\,[\Phi^{(\bar n)}N_2N_1]\tilde{}$, etc. (resp. $\tilde G_\Phi$).

Definition 2(f) (The Φ iteration): Starting from the "free" two-point function $H^2_f(q)=q^2+1$, we apply Eqs. (2.4) successively. In this way we define an iteration procedure that we call the Φ iteration. The order ν of the Φ iteration is given by the equations

$$H^2_\nu(q,\Lambda)=A^2_{\nu-1}(q,\Lambda)+q^2+1, \tag{2.16a}$$

$$H^{n+1}_\nu(q,\Lambda)=A^{n+1}_{\nu-1}(q,\Lambda)+B^{n+1}_{\nu-1}(q,\Lambda)+C^{n+1}_{\nu-1}(q,\Lambda), \tag{2.16b}$$

where the global terms $A^{n+1}_{\nu-1}$, $B^{n+1}_{\nu-1}$, and $C^{n+1}_{\nu-1}$ are defined in terms of the corresponding Φ^4 operations [cf. (2.5)–(2.7)] as follows:

$$\begin{aligned}
A^{n+1}_{\nu-1}&=-\Lambda[N_3H^{n+3}_{\nu-1}](q,\Lambda),\\
B^{n+1}_{\nu-1}&=-3\Lambda\sum_{\omega_n(J)}[N_2H^{j_2+2}_{\nu-1}][N_1H^{j_1+1}_{\nu-1}](q,\Lambda),\\
C^{n+1}_{\nu-1}&=-6\Lambda\sum_{\omega_n(J)}\prod_{l=1}^{3}[N_1H^{j_l+1}_{\nu-1}](q,\Lambda).
\end{aligned} \tag{2.17}$$

Let us denote by $\nu_0(n)$ the order of the Φ iteration at which the function $H^{n+1}(q,\Lambda)$ starts to be $\neq 0$. At a given order ν of the Φ iteration there is a finite sequence of functions H^{n+1} corresponding to the same $\nu_0(n)$, i.e.,

$$\nu=\nu_0(n), \tag{2.18}$$

if n is such that $3^{\nu-1}<n\leqslant 3^\nu$. Moreover, we define the characteristic order $\overline{\nu_0(n)}$ of a function H^{n+1} by

$$\overline{\nu_0(n)}=\begin{cases}\nu, & \text{if } n=n_{\min},\\ \nu+1, & \text{if } n>n_{\min},\end{cases}\quad \text{in } 3^{\nu-1}<n\leqslant 3^\nu, \tag{2.19}$$

where $n_{\min}$ means the minimal n of the sequence (2.18).

The study of the first three orders of the Φ iteration up to $n=11$, i.e., $\{H^{n+1}_\nu,\ 1\leqslant n\leqslant 11,\ 1\leqslant\nu\leqslant 3\}$, displays interesting properties of signs, bounds, and particular factorization properties. They are explicitly obtained in Appendix Φ of III. We have proved[25] that these properties are more generally satisfied for every n and at every order ν of the Φ iteration. We present them below without proof.

B. Properties conserved by the Φ iteration

There exist, $\forall \nu \geqslant \overline{\nu_0(n)}$ and when $0<\Lambda\leqslant 0.01$, well-defined positive bounded sequences of continuous functions of Λ—$\delta_{\nu,n}(\Lambda)$, $\hat{\delta}_{\nu,n}(\Lambda)$, $\hat{\epsilon}_{\nu,n}(\Lambda)$, $\epsilon_{\nu,n}(\Lambda)$, $d_n(\Lambda)$, $\phi_n(\Lambda)$, $\tilde{\phi}_n(\Lambda)$—such that for every coherent sequence of ΦC's the following properties are verified, $\forall n = 1,3,\ldots$.

Φ(1). Signs:

$$(a)\ (-1)^{(n-1)/2} s_\Phi [\Phi^{(\bar{n},n)}(H) H^{n+1}]_{q_I=0} > 0, \tag{2.20a}$$

$$s_\Phi = \prod_{l=1}^{m} (-1)^{(i_l-1)/2}$$

[cf. Eq. (2.12)],

$$(b)\ H_\nu^2(q,\Lambda)\Delta_F(q) > 1, \quad q \in \mathscr{C}^2_{(q)}, \qquad H_\nu^4(q,\Lambda) < 0, \quad q \in \mathscr{C}^6_{(q)}. \tag{2.20b}$$

Φ(2). Simple splitting: For all $n\geqslant 3$, $\exists 0<\delta_\infty<+\infty$, $0<\gamma_0(\Lambda)<1$,

$$|\Phi^{(\bar{n})}(H)[N_2 H^{n+1}][N_1 H^2]|_{q_I=0} \leqslant \delta_{\nu,n}(\Lambda)|\Phi^{(\bar{n})}(H)\Delta_\nu^{n+1}(H)[N_1 H^2]|_{q_I=0}, \tag{2.21}$$

with

$$\gamma_0(\Lambda)\leqslant\delta_{\nu,n}(\Lambda)\leqslant\delta_\infty(\Lambda) \quad \text{and} \quad \lim_{\Lambda\to 0}\delta_{\nu,n}/\Lambda = \text{const} > 0. \tag{2.21a}$$

Here

$$\Delta_\nu^{n+1} = \sum_{\mathscr{P}(\bar{I})} \prod_{l=1}^{2} H^{\bar{I}_l+1}\Delta_F. \tag{2.21b}$$

For a graphical representation of the simple splitting, cf. Fig. 7 and Appendix Φ, Fig. Φ.5) in III.

Φ(3). Double splitting: For all $n\geqslant 3$, $\exists 0<\hat{\delta}_\infty(\Lambda)<\infty$, $0<\hat{\gamma}_0(\Lambda)<1$,

$$\Lambda|\Phi^{(\bar{n})}(H)[N_3 H_\nu^{n+3}]|_{q_I=0} \leqslant \hat{\delta}_{\nu,n+2}(\Lambda)|\Phi^{(\bar{n})}(H)C_\nu^{n+1}|_{q_I=0}, \tag{2.22}$$

with

$$\hat{\gamma}_0(\Lambda)\leqslant\hat{\delta}_{\nu,n}(\Lambda)\leqslant\hat{\delta}_\infty(\Lambda), \quad \lim_{\Lambda^2\to 0}\hat{\delta}_{\nu,n}(\Lambda)/\Lambda^2 = \text{const} > 0, \tag{2.22a}$$

and C_ν^{n+1} defined by (2.17) ("tree" terms).

Φ(4). Convergence properties: For all $n\geqslant 1$,

$$|\Phi^{(\bar{n},n)}(H)H_\nu^{n+1}|_{q_I=0} \leqslant (1+\hat{\epsilon}_{\nu,n}(\Lambda))|\Phi^{(\bar{n},n)}(H)H_{\nu-1}^{n+1}|_{q_I=0},$$

$$\lim_{\nu\to\infty}\epsilon_{\nu,n}(\Lambda) = 0, \tag{2.23a}$$

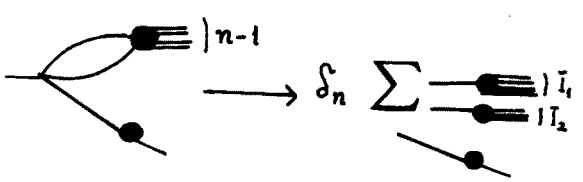


FIG. 7. Graphical representation of the simple splitting.

$$|\Phi^{(\bar{n},n)}(H)H_\nu^{n+1}|_{q_I=0} \geqslant (1-\epsilon_{\nu,n}(\Lambda))|\Phi^{(\bar{n},n)}(H)H_{\nu-1}^{n+1}|_{q_I=0},$$

$$\lim_{\nu\to\infty}\epsilon_{\nu,n}(\Lambda) = 0. \tag{2.23b}$$

Φ(5). Tree dominance: For all $n\geqslant 3$,

$$|\Phi^{(\bar{n},n)}(H)B_\nu^{n+1}|_{q_I=0} \leqslant d_{\nu,n}(\Lambda)|\Phi^{(\bar{n},n)}(H)C_\nu^{n+1}|_{q_I=0}, \tag{2.24}$$

with

$$0<d_{\nu,n}(\Lambda)<1. \tag{2.24a}$$

Φ(6). Zero momentum dominance: For all $n\geqslant 1$,

$$|\Phi^{(\bar{n},n)}(H)H^{n+1}| \leqslant (1+\varphi_{\nu,n}(\Lambda))|\Phi^{(\bar{n},n)}(H)H^{n+1}|_{q_I=0}, \qquad \varphi_{\nu,n}(\Lambda) < +\infty. \tag{2.25}$$

Φ(7). The norms:

$$|\Phi^{(\bar{n},n)}(H)H^{n+1}|_{q_I=0} \leqslant (1+\tilde{\varphi}_{\nu,n}(\Lambda))^{(n-1)/2} n!\,\Lambda^{(n-1)/2}|\Phi^{(\bar{n},n)}(H)|_{q_I=0}, \qquad \tilde{\varphi}_{\nu,n}(\Lambda) < \tilde{\varphi}_0 < \infty. \tag{2.26}$$

C. Remarks: The convergence to the solution of the Φ iteration and the divergence of the perturbative series

Remark 1: By inspection of the above properties we first notice that the characteristic signs, for all $n\geqslant 1$, hold *at zero external momenta*, while the positive (resp. negative) sign of two-point (resp. four-point) functions are valid for every value of the external momenta.

Moreover, the factorization properties that we called simple (resp. double) splitting, in terms of lower $\bar{n}$ ($\bar{n}\leqslant n-2$) Green's functions, will be formulated more globally later (in Sec. III and in II) in terms of the dominant contribution $H^{n-1}H^2$ (resp. $H^{n-1}H^2H^2$). Concerning the "tree dominance" properties [the name comes from the tree terms C_ν^{n+1} of (2.16) and (2.17)], they are intimately related to the zero momentum dominance and norm properties Φ(6) and Φ(7). So they constitute the crucial features for the conservation of the norms in $\mathscr{B}$ space (cf. also below in remark 2, and in II where the fixed-point method is developed).

Remark 2: The proof of the Φ iteration's convergence to the solution of Eqs. (2.4), given in Ref. 25, depends crucially on the above properties. In other words, the conservation of the signs and splitting properties is equivalent to the convergence of the Φ iteration. We do not present here the detailed demonstration; nevertheless, it is worthwhile to explain schematically how the above features Φ(1)–Φ(7) are combined in the general strategy to yield the final answer.

The argument goes recursively. We suppose that Φ(1)–Φ(7) are true up to order $\nu-1$ and for all $1\leqslant n\leqslant 3^{\nu-1}$. Then the sign properties Φ(1) yield the good sign $(-1)^{(n-1)/2}$ for the global terms $A_{\nu-1}^{n+1}$ and $C_{\nu-1}^{n+1}$ of (2.16), and the opposite sign for $B_{\nu-1}^{n+1}$. Due to the tree dominance property Φ(5) (up to order $\nu-1$) and the previous result, we obtain the good sign of every H_ν^{n+1} at zero external momenta to-

gether with the "positivity" (resp. negativity) of H_ν^2 (resp. H_ν^4) when $0<\Lambda\leqslant 0.01$. Furthermore, by application of the convergence properties $\Phi(4)$, a secondary recursion (over $\bar{n}$, with $1\leqslant\bar{n}\leqslant n-2$ and ν fixed) is established to show the simple double splitting as well as the tree dominance $\Phi(2)$, $\Phi(3)$, and $\Phi(5)$, always under the condition $0<\Lambda\lesssim 0.01$.

These results, in turn, are necessary in order to prove the convergence properties $\Phi(4)$ and the zero momentum dominance $\Phi(6)$. Then these last properties directly imply the conservation of the norms $\Phi(7)$; so the total "loop" of the recurrence hypothesis is closed. By a simple argument of convergent infinite products [combining $\Phi(4)$ with $\Phi(7)$] we finally obtain the complete answer.

By this brief presentation one realizes that the convergence of the Φ iteration has been possible only thanks to these properties. On the other hand, the solution we found [for the system (2.4)] as a limit point of the Φ iteration, when $\nu\to\infty$, satisfies all of them.

These are precisely the reasons that motivated us to construct our method by formulating the structure of the Φ iteration, described by $\Phi(1)$–$\Phi(7)$, in terms of the appropriate subsets in the space $\mathscr{B}$. Moreover, the existence of the solution is proved inside these particular subsets (cf. Sec. III and Paper II, Sec. 4).

Remark 3: Another interesting aspect of the Φ iteration is the fact that, at fixed n and for every ν sufficiently large ($\nu\geqslant\nu_0(n)$), it contains all terms up to the order ν of the perturbative expansion. More precisely, following the details of Appendix Φ of III, we can verify that in the first three orders, for every fixed n, all terms (graphically, all Feynmann graphs) of the corresponding perturbative series proportional to $\sim\Lambda^{\bar{\nu}}$, with $\bar{\nu}\leqslant 3$, are present in the $A_{\bar{\nu}-1}^{n+1}$, or the $B_{\bar{\nu}-1}^{n+1}$, or the $C_{\bar{\nu}-1}^{n+1}$ terms (graphically, the G_Φ graphs of the global terms) of the Φ iteration. Moreover, supplementary terms of the perturbative expansion of order larger than $\nu\geqslant 3$ appear even for $\nu=2$ (see, graphically, the example of Fig. Φ.3 in Appendix Φ of III).

Using these properties as the starting point, we establish a recurrence hypothesis up to order $\bar{\nu}\leqslant\nu-1$. We verify them for the order ν by simple application of this recurrence hypothesis on the global terms $A_{\nu-1}^{n+1}$, $B_{\nu-1}^{n+1}$, $C_{\nu-1}^{n+1}$ of (2.16) and (2.17). (Notice that Λ multiplies $H_{\nu-1}^{n-1}H_{\nu-1}^2H_{\nu-1}^2$, etc., or $H_{\nu-1}^{n+1}H_\nu^2$, etc., or $H_{\nu-1}^{n+3}$.)

We state this result in the following proposition.

Proposition 2.2: Let n be fixed. At every order ν of the Φ iteration such that $\nu\geqslant\bar{\nu}_0(n)$ [cf. Eq. (2.19)], all terms of the corresponding perturbative series proportional to $\sim\Lambda^{\bar{\nu}}$, with $0\leqslant\bar{\nu}\leqslant\nu$, are contained in the rhs of the expansion of H_ν^{n+1} of Eqs. (2.16) and (2.17). Moreover, supplementary terms of the perturbative expansion proportional to $\Lambda^{\tilde{\nu}}$, with $\tilde{\nu}\geqslant\nu+1$, exist distributed inside the global terms $A_{\nu-1}^{n+1}, B_{\nu-1}^{n+1}, C_{\nu-1}^{n+1}$ of Eq. (2.17).

In other words, we have proved that at every fixed n, the expansion provided by the Φ iteration is a (convergent) reordered summation of the perturbative series in the form

$$\lim_{\nu\to\infty} H_\nu^{n+1}=\lim_{\nu\to\infty}\left\{\sum_{\bar{\nu}=1}^{\nu}{}^{\text{pert}}c_{\bar{\nu}}\Lambda^{\bar{\nu}}+\sum_{\bar{\nu}=\nu+1}^{\infty}c_{\bar{\nu}}\Lambda^{\bar{\nu}}\right\}.$$

So finally, as we already argued in the Introduction, our Remarks 1–3 above eliminate the contradiction between the convergence of the Φ iteration and the divergence of the perturbation theory[24] in the two-dimensional Euclidean space.

III. THE ZERO-DIMENSIONAL PROBLEM, THE SWEEPING FACTORS, AND THE SPLITTING CONSTANTS

In this section we shall present the fundamental elements of our formalism, which is developed in Papers II and III. These elements are reformulations of the notions and properties revealed by the Φ iteration presented in Sec. II.

A. The zero-dimensional problem

We start with the description of the zero-dimensional problem and the corresponding space, $\mathscr{B}_0$. We also explain how the zero-dimensional study provided us with the basic combinatorial tools for the technique of the method. We introduce the sweeping factors β_n by the so-called sweeping procedure, and then define in terms of these factors the splitting procedure and the corresponding sequences $\{\delta_n(\Lambda)\}_n$. The splitting constants $\delta_n(\Lambda)$ have already appeared in the splitting property $\Phi(2)$ of the Φ iteration in Sec. II. We mentioned there (cf. remarks 1 and 2) the role of the factorization, or splitting, property in terms of bounded positive functions of Λ, $\delta_n(\Lambda)$, for the self-consistency of the overall Φ-iteration structure.

In the second part of the section, we translate the previous definitions in terms of the elements of coherent sequences of ΦC's (in the space $\mathscr{B}$) at zero external momenta, i.e., we introduce the sweeping factors β_n and the splitting constants δ_n to be used extensively in II for Eqs. (24).

Finally, using these definitions, the Φ iteration's convergence to the solution (thanks to the signs, splitting, and norms) is reformulated in terms of two theorems, Theorems 3.1 and 3.2. These theorems state the results of our method concerning the zero- and two-dimensional problems, and their proof constitutes the purpose of Paper II. We close this section and the paper by some remarks concerning the four- (and three-)dimensional problem.[21]

Definition 3(a) (The subspace $\mathscr{B}_0$ of $\mathscr{B}$): We consider the subspace $\mathscr{B}_0\subset\mathscr{B}$ of sequences $H_0\equiv\{H_0^{n+1}(\Lambda)\}_{n=1,3,\ldots}$ of continuous functions of $\Lambda\in\mathbb{R}^+$, characterized by increasing properties with respect to n described as follows: For all $H_0\in\mathscr{B}_0$ there exist finite positive constants c_H such that, $\forall n\geqslant 1$,

$$|H_0^{n+1}(\Lambda)|\leqslant n!\,c_H^{(n+1)/2}.$$

Definition 3(b) (The zero-dimensional problem): We define the nonlinear mapping $\mathscr{M}_0$: $\mathscr{B}_0\xrightarrow{\mathscr{M}_0}\mathscr{B}_0$ by the system of equations

$$H_0^2(\Lambda)=-\Lambda H_0^4(\Lambda)+1, \tag{3.1a}$$

$$H_0^{n+1}(\Lambda)=A_0^{n+1}(\Lambda)+B_0^{n+1}(\Lambda)+C_0^{n+1}(\Lambda). \tag{3.1b}$$

Here

$$A_0^{n+1}(\Lambda) = -\Lambda H_0^{n+3}(\Lambda), \tag{3.2a}$$

$$B_0^{n+1}(\Lambda) = -3\Lambda \sum_{\omega(J)} \theta^n_{j_1 j_2} H_0^{j_2+2}(\Lambda) H_0^{j_1+1}(\Lambda), \tag{3.2b}$$

$$C_0^{n+1}(\Lambda) = -6\Lambda \sum_{\omega(I)} \theta^n_{i_1 i_2 i_3} \prod_{l=1}^{3} H_0^{i_l+1}(\Lambda), \tag{3.2c}$$

with

$$\theta^n_{j_1 j_2} \equiv n!/j_1!\, j_2, \quad \theta^n_{i_1 i_2 i_3} \equiv [\, n!/i_1!\, i_2!\, i_3!\, \sigma_{\mathrm{sym}}(i_1 i_2 i_3)\,],$$

$$\sigma_{\mathrm{sym}}(i_1 i_2 i_3) = \begin{cases} 1!, & \text{if } i_1 \neq i_2 \neq i_3, \\ 3!, & \text{if } i_1 = i_2 = i_3, \\ 2!, & \text{otherwise,} \end{cases} \tag{3.3}$$

and the following set of partitions $\omega(J)$, $\omega(I)$ of n:

$$\omega(J) = \{j_1 = 2k+1,\ k \in \mathbf{N},\ j_2 = n - j_1\},$$

and equivalent notation $(j_1 j_2)$, (3.4a)

$$\omega(I) = \left\{ i_l = 2k+1,\ k \in \mathbf{N},\ l = 1,2,3,\ \sum_{l=1}^{3} i_l = n \right\},$$

and equivalent notation $(i_1 i_2 i_3)$. (3.4b)

Remark 1: The study of the above system (3.1) is called the zero-dimensional problem, and it is carried out in detail in II. We have obtained this system directly from the system (2.4) of Sec. II by eliminating all momentum dependence and by identification of every Φ convolution with a factorization of the constant 1.

Following Sec. II the system (2.4) has been derived, in turn, from the definition of the renormalized normal product,[15] which constitutes the starting point of the method. So as we explained in the Introduction, a crucial test of the method should be to verify the equivalence of Eqs. (3.1) with the analogous system one obtains starting from the standard functional integral formula.[19] Under an appropriate renormalization condition on the two-point function, the equivalence, together with the identification of the corresponding solutions of the two zero-dimensional systems, has been established thanks to numerical calculations of Voros.[20]

In Appendix V of III, we present in detail the procedure of this verification, as well as the corresponding numerical results.

Remark 2: The absence of the complexity due to the ΦC's in two (or one) dimensions makes the zero-dimensional problem particularly attractive. Its simplicity has been appropriate in order to formulate all properties of the Φ iteration (cf. Sec. II) in terms of notions more elegant mathematically.

In other words, this particular simple form of the system (3.1) stimulated us to explore combinatorial properties of the global terms C_0^{n+1}, B_0^{n+1} of the system, which in turn allowed us to develop completely the formalism (splitting norms) in zero dimensions.

Thus the extension of the technique in the dimensional cases has been directly through the study of a family (cf. Paper II, Sec. 4) of zero-dimensional-type systems.

B. The sweeping factors

Let us first introduce the sequences of the so-called sweeping factors that play a fundamental role in the combinatorial structure of the system (3.1).

Definition 3(c) (The sweeping procedure): For every $n \geqslant 3$ and $\Lambda \in \mathbf{R}^+$, we define the sequences β_n [resp. $\beta^n_{i_1 i_2 i_3}$ for every triplet $(i_1 i_2 i_3)$ with $(i_1 \geqslant i_2 \geqslant i_3)$] of sweeping factors by the following:

$$\beta_3 = \beta_5 = 1,$$

$$|C_0^{n+1}| = 3\Lambda n(n-1)\beta_n |H_0^{n-1}|\, |H_0^2|^2, \quad \forall n \geqslant 7, \tag{3.5}$$

with C_0^{n+1} defined by (3.2c). (From now on we shall very often omit the argument Λ for simplicity.) Respectively, for every fixed partition $\omega(\bar{I})$ with $\bar{i}_1 \geqslant \bar{i}_2 \geqslant \bar{i}_3$,

$$\left| \sum_{\omega(I)} \theta^n_{i_1 i_2 i_3} \prod_{l=1}^{3} h_0^{i_l+1} \right|$$
$$= \beta^n_{\bar{i}_1 \bar{i}_2 \bar{i}_3} \theta^n_{\bar{i}_1 \bar{i}_2 \hat{i}_3} \prod_{l=1}^{3} \left| H_0^{\bar{i}_l+1} \right|$$
$$+ \beta^n_{\bar{i}_1 - 2, \hat{i}_2, \hat{i}_3} |H_0^{\bar{i}_1 - 1}|\, |H_0^{\bar{i}_2+1}|\, |H_0^{\hat{i}_3+1}|,$$
$$i_1 \leqslant \bar{i}_1, \quad i_2 \leqslant \bar{i}_2(i_1), \quad \text{if } \bar{i}_3 > \hat{i}_3 \equiv i_3 \min(\bar{i}_1 - 2), \tag{3.5a}$$

$$\left| \sum_{\omega(I)} \theta^n_{i_1 i_2 i_3} \prod_{l=1}^{3} H_0^{i_l+1} \right|$$
$$= \beta^n_{\bar{i}_1 \bar{i}_2 \bar{i}_3} \theta^n_{\bar{i}_1 \bar{i}_2 \hat{i}_3} \prod_{l=1}^{3} \left| H_0^{\bar{i}_l+1} \right|,$$
$$i_1 \leqslant \bar{i}_1, \quad i_2 \leqslant \bar{i}_2(i_1), \quad \text{if } \bar{i}_3 \leqslant \hat{i}_3, \tag{3.5b}$$

where $\hat{i}_3 \equiv i_3 \min(i_1 - 2)$ means the minimal value of i_3 among the triplets $\omega(\bar{I}) \equiv (\bar{i}_1 - 2, \bar{i}_2 \bar{i}_3)$.

From the above definitions one realizes the qualitative role of the sequences β_n (resp. $\beta^n_{i_1 i_2 i_3}$). Starting from $i_1 \min \equiv [n/3]$ they "sweep" all the terms—all possible different configurations $(i_1 i_2 i_3)$—of the sum C_0^{n+1}, and they replace them by an expression in terms of only the "largest" contribution $|H_0^{n-1}||H_0^2|^2$ (resp. $|H_0^{\bar{i}_1 - 1}||H_0^{\bar{i}_2+1}||H_0^{\hat{i}_3+1}|$ and $\Pi_{l=1}^3 |H^{\bar{i}_l+1}|$). For this reason we called this operation the sweeping procedure.

In Appendix $\mathcal{T}$ of III we have evaluated the precise number of different terms (partitions) inside the total ordered tree sum C_0^{n+1} and denoted it by $\bar{\mathcal{T}}_n$; we found that

$$\bar{\mathcal{T}}_n \approx (n-3)^2/48 + (n-3)/4. \tag{3.6a}$$

On the other hand, using the sweeping procedure, we prove in III that every sweeping factor β_n is proportional (up to a factor smaller than 1) to $\bar{\mathcal{T}}_n$. Schematically, for sufficiently large n, we can write

$$\beta_n \sim \bar{\mathcal{T}}_n \sim n^2. \tag{3.6b}$$

So the sweeping procedure yields by (3.5) a form for the global term C_0^{n+1} of Eqs. (3.1), which simplifies the studies in II and III, and simultaneously carries precise combinatorial information about the tree term.

We also note that in II, analogous sweeping factors—denoted by α_n—will be introduced by a corresponding sweeping procedure inside the global term B_0^{n+1} of Eqs. (3.1), which are proportional to $\alpha_n \sim n$. Due to their advantages explained above, the sweeping factors constitute the basis for the definitions and proofs of the method presented below and, essentially, in II and III.

C. The splitting constants

Definition 3(d) (The space $\mathscr{B}_\delta$ and the splitting procedure in $\mathscr{B}_0$): We consider the space of infinite sequences of C^∞ functions of $\Lambda\in\mathbb{R}^+$ denoted by $\delta\equiv\{\delta_n(\Lambda)\}_n$ and satisfying the following increase properties: For all δ there exists a finite positive constant c_δ such that $|\delta_n|\leqslant C_\delta^{(n-1)/2}$, $\forall n$. We call this space $\mathscr{B}_\delta$. With every sequence $H_0\in\mathscr{B}_0$ we associate a sequence $\delta\in\mathscr{B}_\delta$ by the following recursion:

$$|H_0^2| = 1+\delta_1(\Lambda)\Lambda,\quad |H_0^4| = \delta_3(\Lambda)|H_0^2|^2,$$
$$|H_0^6| = 4\delta_5(\Lambda)|H_0^4|\,|H_0^2|, \tag{3.7a}$$

and, $\forall n\geqslant 7$,

$$|H_0^{n+1}(\Lambda)| = \delta_n(\Lambda)\beta_n(\Lambda)|H_0^{n-1}|\,|H_0^2|. \tag{3.7b}$$

Here the sequences $\beta_n(\Lambda)$ are the same as those introduced by the sweeping procedure of Definition 3(c). This kind of factorization (by analogy to the corresponding property of Φ iteration) of every $H_0^{n+1}(\Lambda)$ in terms of the "neighboring" $H_0^{n-1}(\Lambda)$ and $H_0^2(\Lambda)$ we call the splitting procedure or, simply, "splitting." The associated sequence $\{\delta_n(\Lambda)\}$ is the sequence of splitting constants.

The above definitions can be extended in the dimensional cases as follows.

Definitions 3(e) (The sweeping factors ${}^\Phi\beta_n$ and the splitting in $\mathscr{B}$: (i) Letting $H_0\in\mathscr{B}_0$ where a coherent sequence of ΦC's$\in\mathscr{F}$ (cf. Sec. II) acts on H_0, we trivially verify that the corresponding completely reduced coherent sequence of ΦC's factorizes out; in other words, $\forall n = 1,3,\ldots$,

$$\Phi^{(\bar n,n)}(H_0)H_0^{n+1} = \prod_{l=1}^{m} H_0^{i_l+1}H_0^{n+1}\tilde\Phi^{(\bar n,n)}. \tag{3.8}$$

Due to this factorization property one can define sweeping and splitting procedures analogous to the ones of the zero-dimensional case above for every coherent sequence of ΦC's. To be precise, we introduce the tree terms of zero-dimensional type for a given coherent sequence of ΦC's, $\Phi\in\mathscr{F}$, by

$${}^\Phi C_0^{n+1} \overset{\text{def}}{\equiv} -6\Lambda\sum_{\omega(I)}\prod_{l=1}^{3} H_0^{i_l+1}\left[\Phi^{(\bar n,n)}\prod_{l=1}^{3}\widetilde{N_1^{(i_l)}}\right]_{q_l=0}. \tag{3.9}$$

We then define the corresponding sweeping factor ${}^\Phi\beta_n$ as follows:

$$|{}^\Phi C_0^{n+1}| = 3\Lambda\,{}^\Phi\beta_n n(n-1)|H_0^{n-1}|\,|H_0^2|^2 \times\left[|\Phi^{(\bar n,n)}\prod_{l=1}^{3}\widetilde{N_1^{(\hat i_l)}}|\right]_{q_l=0}$$
$$(\hat i_1 = n-2,\ \hat i_2=\hat i_3=1). \tag{3.10}$$

Moreover there exist expressions analogous to (3.5a) and (3.5b), which define (at fixed n) a general sweeping factor ${}^\Phi\beta^n_{i_1i_2i_3}$ corresponding to an arbitrary fixed partition $\omega(I)$ of n and the associated partial ordered sum of $|{}^\Phi C_0^{n+1}|$ (cf. II).

(ii) For every $H\in\mathscr{B}$, and for a given coherent sequence of ΦC's, $\{\Phi^{(\bar n,n)}(H)\}_n\in\mathscr{F}$, we introduce the corresponding splitting sequence ${}^\Phi\delta\in\mathscr{B}_\delta$ by the following recursive procedure:

$$|\Phi^{(\bar n,1)}(H)H^2|_{q_l=0} = (1+{}^\Phi\delta_1\Lambda)|\tilde\Phi^{(\bar n,1)}(H)|_{q_l=0},\quad \text{and define}\quad {}^\Phi H_0^2\equiv 1+{}^\Phi\delta_1\Lambda; \tag{3.11a}$$
$$|\Phi^{(\bar n,3)}(H)H^4|_{q_l=0} = {}^\Phi\delta_3(\Lambda)[{}^\Phi H_0^2]^2|\tilde\Phi^{(\bar n,3)}(H)|_{q_l=0},\quad \text{and define}\quad {}^\Phi H_0^4\equiv{}^\Phi\delta_3[{}^\Phi H_0^2]^2; \tag{3.11b}$$
$$|\Phi^{(\bar n,5)}(H)H^6|_{q_l=0} = 4{}^\Phi\delta_5(\Lambda)|{}^\Phi H_0^4|\,|{}^\Phi H_0^2|,\quad \text{and define}\quad {}^\Phi H_0^6\equiv 4{}^\Phi\delta_5|{}^\Phi H_0^4|\,|{}^\Phi H_0^2|; \tag{3.11c}$$

and, $\forall n\geqslant 7$ recurrently, taking into account the definition (3.10) of ${}^\Phi\beta_n$ in terms of ${}^\Phi H_0^{\bar n+1}$, $\forall\bar n\leqslant n-2$,

$$|\Phi^{(\bar n,n)}(H)H^{n+1}|_{q_l=0} = {}^\Phi\delta_n(\Lambda)\,{}^\Phi\beta_n(\Lambda)|{}^\Phi H_0^{n-1}|\,|{}^\Phi H_0^2|\,|\tilde\Phi^{(\bar n,n)}(H)|_{q_l=0},\quad \text{and define}\quad |{}^\Phi H_0^{n+1}|\equiv{}^\Phi\delta_n\,{}^\Phi\beta_n|{}^\Phi H_0^{n-1}|\,|{}^\Phi H_0^2|. \tag{3.11d}$$

We call the above procedure the splitting procedure in $\mathscr{B}$.

We notice that the sequence ${}^\Phi H_0\equiv\{{}^\Phi H_0^{n+1}\}_n$ defined by (3.11a)–(3.11d) belongs to $\mathscr{B}_0$. That means the splitting procedure in $\mathscr{B}$ yields sequences of zero-dimensional type by factorization of the (partially) reduced ΦC's. This fact implies that it is possible, once the fixed-point method is formulated to resolve the zero-dimensional problem (cf. Sec. 3 of II), to extend it in two (or one) dimensions.

The solution of the two- (or one-)dimensional problem, presented in Sec. 4 of II, is precisely the realization of such a nontrivial extension.

Using the above definitions we finally present, by two theorems proved in II and III, our results concerning the zero- and two-dimensional problems. These two theorems translate into a more mathematically consistent language the positive answer we obtained by the Φ iteration for the existence of a solution for Eqs. (2.4) under the *conservation of signs, splitting,* and *norms* in $\mathscr{B}$ space.

Theorem 3.1: when $0<\Lambda\leqslant 0.01$, the system (3.1) [cf. Definition 3(b)] has at least one nontrivial solution $\bar H_0\in\mathscr{B}_0$ that satisfies the following properties.

(i) Signs and splitting:

$$\bar H_0^2 = 1+\bar\delta_1\Lambda;\quad \bar H_0^4 = -\bar\delta_3[\bar H_0^2]^2,$$
$$\bar H_0^6 = -4\delta_5\bar H_0^4\bar H_0^2, \tag{3.12a}$$
$$\bar H_0^{n+1} = -\bar\delta_n\bar\beta_n H^{n-1}. \tag{3.12b}$$

Here (a) the splitting sequence $\bar\delta=\{\bar\delta_n\}_n$ is a well-defined positive bounded sequence in $\mathscr{B}_\delta$ [cf. Definition 3(d)] with precise slow increase properties (cf. II), and also

$$\lim_{\Lambda\to 0}\bar{\delta}_n(\Lambda)/\Lambda = \text{const}, \quad \forall n = 1,3,...; \tag{3.13}$$

and (b) the sequence of sweeping factors $\bar{\beta}_n$ [defined by (3.4)] can be explicitly expressed recurrently in terms of all $\bar{\delta}_{\bar{n}}$'s with $\bar{n}\leqslant n-2$, and satisfies [cf. Definition 3(c) and II]

$$\bar{\beta}_n(\Lambda)\sim\mathscr{T}_n\sim n^2/48.$$

(ii) The above solution $\bar{H}_0$ coincides with the standard solution of the equations analogous to (3.1), obtained directly by the functional integral method (cf. III).

We also note that in II the uniqueness of $\bar{H}_0$ is established inside a restricted subset of $\mathscr{B}_0$ characterized by (i)(a) above and some order relations of the splitting sequences $\{\delta\}$.

Theorem 3.2: When Λ satisfies $0<\Lambda\lesssim 0.006$, then the system (2.4) [cf. Definition 2(a)] has a unique nontrivial solution $\bar{H}\in\mathscr{B}$ that satisfies the following signs and splitting properties at zero external momenta: For any coherent sequence of ΦC's $\{\Phi^{(\bar{n},n)}(H)\}_n$ there exists ${}^{\Phi}\bar{H}_0\in\mathscr{B}_0$ such that

$$\{\Phi^{(\bar{n},n)}(H)H^{n+1}\}_{q_j=0} = {}^{\Phi}\bar{H}_0^{n+1}\{\bar{\Phi}^{(\bar{n},n)}(H)\}_{q_j=0}, \tag{3.14}$$

with

$${}^{\Phi}\bar{H}_0^2 = 1+{}^{\Phi}\bar{\delta}_1\Lambda, \quad {}^{\Phi}\bar{H}_0^4 = -{}^{\Phi}\bar{\delta}_3(\Lambda)[{}^{\Phi}\bar{H}_0^2]^2,$$
$${}^{\Phi}\bar{H}_0^6 = -4{}^{\Phi}\bar{\delta}_5(\Lambda){}^{\Phi}\bar{H}_0^4{}^{\Phi}\bar{H}_0^2, \tag{3.14a}$$

and

$${}^{\Phi}\bar{H}_0^{n+1} = -{}^{\Phi}\bar{\delta}_n{}^{\Phi}\bar{\beta}_n{}^{\Phi}\bar{H}_0^{n-1}{}^{\Phi}\bar{H}_0^2, \quad \forall n\geqslant 7.$$

Here (a) the splitting sequence ${}^{\Phi}\bar{\delta}\equiv\{{}^{\Phi}\bar{\delta}\}_n$ is a uniquely defined positive bounded sequence in $\mathscr{B}_\delta$ (cf. Definition 3(d)] depending on the coherent sequence of ΦC's, satisfying precise slow increase properties (with respect to n) and

$$\lim_{\Lambda\to 0}{}^{\Phi}\bar{\delta}_n(\Lambda)/\Lambda = \text{const}>0; \tag{3.15}$$

and (b) the sweeping factors ${}^{\Phi}\bar{\beta}_n$ [defined by (3.10)] can be written in explicit recursive expressions in terms of all ${}^{\Phi}\bar{\delta}_{\bar{n}}$'s with $\bar{n}\leqslant n-2$, and they satisfy

$${}^{\Phi}\bar{\beta}_n\sim\bar{\mathscr{T}}_n\sim n^2/48$$

[cf. Definition 3(c) and II]. We again notice that the uniqueness has to be understood in the sense that there is no other solution of Eqs. (2.4) in $\mathscr{B}$ satisfying the above signs and splitting properties at zero external momenta.

IV. FINAL REMARK

Analogous results are obtained in four dimensions (or three dimensions) for the corresponding system of equations derived by the definition of the renormalized normal product [cf. Eqs. (3.6) of Ref. 15]. The supplementary ingredients in this case are (a) the renormalization operators that are nontrivial now and have to be taken into account, and the particular consequences of this nontriviality concerning the signs and splitting properties, in terms of the renormalized Φ convolutions; and (b) the interpretation of the physical renormalization parameters (defined in Ref. 15) in terms of this solution.

ACKNOWLEDGMENTS

This program of research began when the author was invited to the University of Bielefeld (Institute für Theoretische Physik). She would like to express her gratitude to Professor G. Sommer, Professor P. Blanchard, Professor S. Albeverio, and Professor O. Steinmann for their kind hospitality and several interesting discussions and criticisms.

This first paper was completed while the author was visiting the University of Marseille, Luminy, and the Institut des Hautes Etudes Scientifiques (IHES) at Bures-sur-Yvette. At both places she has profited from the stimulating environment created by the presence of Professor M. Sirugue-Collin, Professor A. Rouet, Professor J. Bellissard, Professor L. Michel, Professor J. Fröhlich, Professor H. Epstein, Professor E. Seiler, and Professor J. Ginibre.

The author wishes to express her sincere thanks to Professor A. Wightman, Professor A. Jaffe, and Professor K. Osterwalder for their interest in her work, their private communications, and their helpful suggestions.

She is also indepted to IHES and, in particular, to F. Breiner and V. Meyer for the material presentation of the paper.

APPENDIX: RECURSIVE CONSTRUCTION OF G_Φ GRAPH (EXAMPLE FOR SEC. II)

We are given

$$\tilde{H}^8_\Phi \equiv \int H^8 \prod_{i=1}^{24}\Delta_F(\tilde{l}_i)\prod_{l=1}^{6}H^{i_l+1}\,dk_{\bar{j}},$$

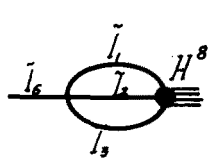
i.e.,

$$\Phi(H) = \prod_{l=1}^{6}H^{i_l+1}.$$

In all the steps below the double arrow $\Rightarrow$ signifies graphical representation of the function that appears to the left of the arrow.

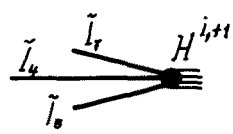
Zeroth step (choice of primitive subgraphs of G_Φ):

$$\tilde{H}_{\gamma_1}\equiv[N_3H^8]\left(\prod_{i=1}^{3}\Delta_F(\tilde{l}_i)\right)\Rightarrow$$

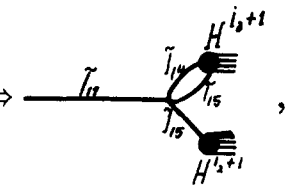


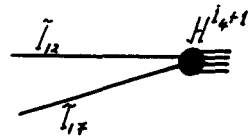
, (γ_1)

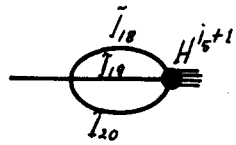
$$\tilde{H}_{\gamma_2}\equiv H^{i_1+1}\Rightarrow$$

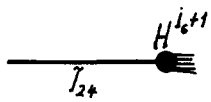


, (γ_2)

$$\tilde{H}_{\gamma_3} \equiv [N_2 H^{i_3+1}] \left(\prod_{i=13}^{14} \Delta_F(\tilde{l}_i) \right) H^{i_2+1} \Delta_F(\tilde{l}_{15}) \Rightarrow$$  , (γ_3)

$$\tilde{H}_{\gamma_4} \equiv H^{i_4+1} \Rightarrow$$  , (γ_4)

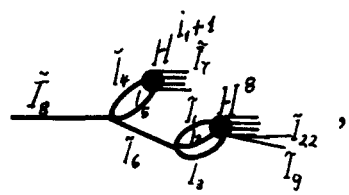
$$\tilde{H}_{\gamma_5} \equiv [N_3 H^{i_5+1}] \left(\prod_{i=18}^{20} \Delta_F(\tilde{l}_i) \right) \Rightarrow$$  , (γ_5)

$$\tilde{H}_{\gamma_6} \equiv H^{i_6+1} \Rightarrow$$  . (γ_6)

[The constant $(-\Lambda)$ has been omitted for simplicity, but it has to be understood at every constant four-vertex, i.e., associated with every operation N_3, N_1N_2, or $N_1N_1N_1$.]

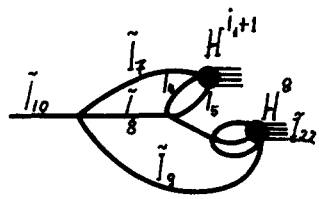
Remark: Notice that in this example the primitive subgraphs (γ_2), (γ_4), and (γ_6) of G_Φ are "simple bubbles."

First step:

$$\tilde{H}_{\gamma_7} \equiv [N_2 H_{\gamma_2}] \left(\prod_{i=4,5} \Delta_F(l_i) \right) \times \tilde{H}_{\gamma_1} \Delta_F(\tilde{l}_6) \Rightarrow$$  , (γ_7)

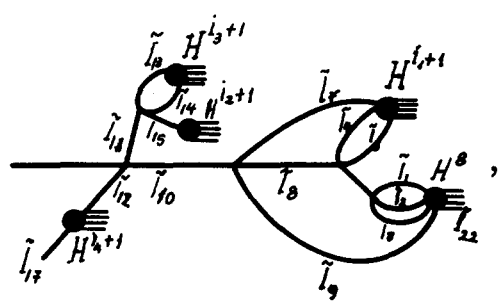
and $\tilde{H}_{\gamma_3} \tilde{H}_{\gamma_4} \tilde{H}_{\gamma_5} \tilde{H}_{\gamma_6} \Rightarrow$ disconnected (γ_3) (γ_4) (γ_5) (γ_6) as in zeroth step).

Second step:

$$\tilde{H}_{\gamma_8} \equiv [N_3 H_{\gamma_7}] \left(\prod_{i=7}^{9} \Delta_F(\tilde{l}_i) \right) \Rightarrow$$  , (γ_8)

and $\tilde{H}_{\gamma_3} \tilde{H}_{\gamma_4} \tilde{H}_{\gamma_5} \tilde{H}_{\gamma_6} \Rightarrow$ disconnected (γ_3) (γ_4) (γ_5) (γ_6) as in zeroth and first step.

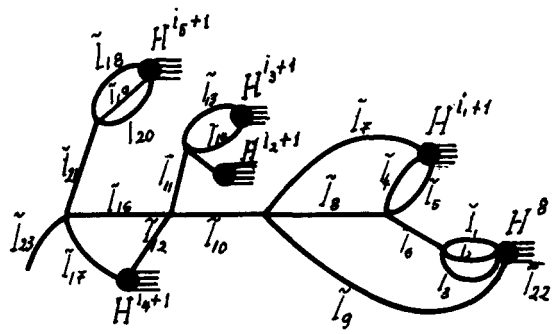
Third step:

$$\tilde{H}_{\gamma_9} \equiv \tilde{H}_{\gamma_8} \Delta_F(\tilde{l}_{10}) \times \tilde{H}_{\gamma_3} \Delta_F(\tilde{l}_{11}) \times \tilde{H}_{\gamma_4} \Delta_F(\tilde{l}_{12}) \Rightarrow$$  , (γ_9)

and $\tilde{H}_{\gamma_5}\tilde{H}_{\gamma_6} \Rightarrow$ disconnected (γ_5) (γ_6) as in second step.

Fourth step:

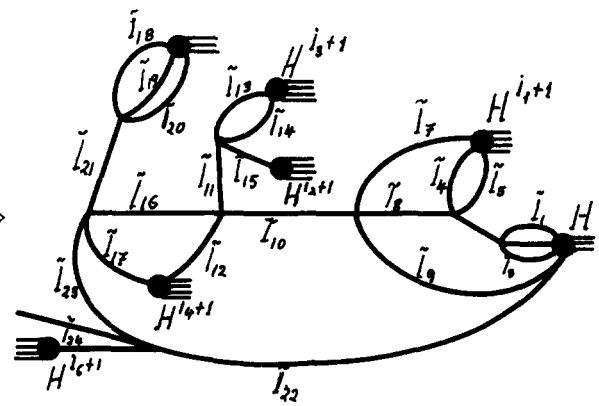
$$\tilde{H}_{\gamma_{10}} \equiv [N_2 \tilde{H}_{\gamma_6}] \left(\prod_{i=16}^{17} \Delta_F(\tilde{l}_1) \right) \times \tilde{H}_{\gamma_5} \Delta_F(\tilde{l}_{21}) \Rightarrow \qquad (\gamma_{10})$$



and $\tilde{H}_{\gamma_6}$ (as in third step) $\Rightarrow$ (γ_6) disconnected.

Fifth step:

$$\tilde{H}^{8}_{(G_\Phi)} \equiv [N_2 \tilde{H}_{\gamma_{10}}] \left(\prod_{i=22}^{23} \Delta_F(\tilde{l}_i) \right) \times H^{i_6+1} \Delta_F(\tilde{l}_{24}) \Rightarrow$$



This is the final Φ graph G_Φ with which the Φ-convolution product $\tilde{H}^8$ is associated.

[1]W. Zimmermann, Commun. Math. Phys. **6**, 161 (1967); **10**, 325 (1968); N. N. Bogoliubov, A. A. Logunov, and I. T. Todorov, *Introduction to the Axiomatic Quantum Field Theory* (Benjamin, New York, 1975); R. Jost, *The General Theory of Quantized Fields* (American Mathematical Society, Providence, RI, 1965); N. N. Bogoliubov and D. V. Shirkov, *Introduction to the Theory of Quantized Fields* (Interscience, New York, 1968).

[2]J. Glimm and A. Jaffe, Phys. Rev. **176**, 1945 (1968); Commun. Math. Phys. **11**, 9 (1968); Bull. Am. Math. Soc. **76**, 407 (1969); Ann. Math. **91**, 362 (1970); Acta Math. **125**, 203 (1970); *Statistical Mechanics and Quantum Field Theory*, Les Houches, 1970 (1-108) (Gordon and Breach, New York, 1971); Commun. Math. Phys. **22**, 253 (1971).

[3]J. Glimm, A. Jaffe, and T. Spencer, *Constructive Quantum Field Theory, Lecture Notes in Physics Vol. 25*, edited by G. Velo and A. Wightman (Springer, Berlin, 1973);T. Spencer, Commun. Math. Phys. **30**, 63 (1974); B. Simon, *The $P(\Phi)_2$ Euclidean (Quantum) Field Theory* (Princeton U.P., Princeton, NJ, 1974); J. Feldman, Commun. Math. Phys. **37**, 93 (1974); J. Magnen and R. Seneor, Ann. Inst. H. Poincaré **24**, 95 (1976); J. Feldman and K. Osterwalder, Ann. Phys. **97**, 80 (1976).

[4]K. Hepp, Commun. Math. Phys. **1**, 59 (1956); K. Symanzik, *Lectures on High Energy Physics Vol. 485* (Federal Nuclear Energy Commission of Yugoslavia, Zagreb, 1961).

[5]T. T. Wu, Phys. Rev. **125**, 1436 (1962); J. Feldman and R. Raczka, Ann. Phys. **10**, 212 (1977).

[6]O. Steinmann, Ann. Phys. (NY) **29**, 76 (1964); **36**, 267 (1966); J. Valatin, Proc. R. Soc. London Ser. A **225**, 535 (1954); **226**, 254 (1954).

[7]K. G. Wilson, Cornell University Report, 1964; R. A. Brandt, University of Maryland Technical Report 646, 1966.

[8]J. G. Taylor, J. Math. Phys. **7**, 1720 (1966); Proc. Phil. Soc. **63**, 335 (1967).

[9]K. Osterwalder and R. Schrader, Commun. Math. Phys. **42**, 281 (1975).

[10]R. Streater and A. Wightman, *PCT Spin Statistics and All That* (Benjamin, New York, 1964).

[11]W. Zimmermann, Ann. Phys. (NY) **77**, 536 (1973); *Lectures on Elementary Particles and Quantum Field Theory, Brandeis Summer Institute*, 1970 (Massachussetts Institute of Technology, Cambridge, MA, 1971).

[12]J. Lowenstein, Phys. Rev. D **4**, 2281 (1971).

[13]J. Bros and M. Manolessou, Commun. Math. Phys. **72**, 175, 207 (1980).

[14]M. Manolessou, Ann. Phys. (NY) **122**, 297 (1979).

[15]M. Manolessou, Ann. Phys. (NY) **152**, 327 (1984).

[16]M. Manolessou and B. Ducomet, Ann. Inst. H. Poincaré **40**, 389 (1984).

[17]M. Manolessou, IHES preprint P/87/49, July 1986.

[18]M. Manolessou, IHES preprint P/87/50, July 1986.

[19]A. Wightman (private communication).

[20]A. Voros (private communication).

[21]M. Manolessou, "The ϕ^4_4 equations of motion, the existence of a nontrivial solution" (unpublished).

[22]M. Aizenmann, Phys. Rev. Lett. **47**, 1 (1981); J. Fröhlich, Nucl. Phys. B **200**, 281 (1982).

[23]A. Jaffe, Commun. Math. Phys. **1**, 127 (1965).

[24]A. Wightman and A. Jaffe (private communication).

[25]M. Manolessou, "The Φ iteration and the nontrivial solution of the Φ^4_2 equations of motion" (unpublished).

The existence of quantum fields for local nets of algebras of observables

Manfred Wollenberg
Karl-Weierstrass-Institut für Mathematik, Akademie der Wissenschaften der DDR, Berlin, German Democratic Republic

(Received 29 October 1987; accepted for publication 4 May 1988)

A necessary and sufficient criterion for the existence of quantum fields related to a given local net of algebras of observables is proved. This criterion is written in terms of the "conjugation operators" $S_{\mathcal{O}}$ of the local algebras $\mathcal{A}(\mathcal{O})$. It shows that the modular structure of the local algebras essentially determines the corresponding quantum field theory. In the case where the local net of algebras is conformally covariant the criterion gives the result that the existence of quantum fields depends only on the representation of the conformal group and does not depend on the concrete form of the algebras. For the proof of the criterion the fact that a quantum field related to a local net of algebras can be essentially characterized by one vector is used.

I. INTRODUCTION

In the last years the relation between the general theory of local quantum fields and the theory of local nets of algebras of observables has been studied and some progress has been made on this question.[1-4] For instance, the following problem has been solved[1-3]: Suppose we have a local net of algebras of observables and a set of quantum fields related to this net. How can we recover these quantum fields (and all other quantum fields related to this net) from the given algebras?

If we consider the local net of algebras of observables as the primary object then a natural question is whether, for a given net, there exist quantum fields describing the same theory. It is difficult to answer this question. Probably the answer is yes only under additional assumptions on the local net of algebras (assumptions that go beyond the usual minimal assumptions). The aim of this paper is to prove a necessary and sufficient criterion for the existence of quantum fields in the case where a local net of algebras of observables is given. Roughly speaking the criterion says the following: Assume there is a vector ω_1 (from an energetic weighted space) invariant under the action of the "conjugation operators" $S_{\mathcal{O}_n}$ [$S_{\mathcal{O}}: A\omega \to A^*\omega, A \in \mathcal{A}(\mathcal{O}), \omega$ is the vacuum] for a sequence $\mathcal{O}_n \subset \mathbb{R}^4$ shrinking to the point zero. Then a quantum field affiliated to the given local net exists. Conversely, if a quantum field affiliated to the given local net exists, then we obtain such a vector ω_1 with the properties above [the case that $\omega_1 = c\omega$ gives the "trivial quantum field" $A(f) = c_f 1$].

This criterion can be applied to a local net of algebras which is conformally covariant. In this case all quantum fields affiliated to the local net of algebras can be described by eigenvectors of operators given by the representation of the conformal group. Thus the question of the existence of quantum fields is determined by the group representation alone and does not depend on the concrete form of the local algebras of observables.

For the proof of the criterion we use the fact that a quantum field (related to a local net) is essentially characterized by one vector $\omega_A := A(0)\omega$, where $A(0)$ is the quantum field at the space-time point zero.

II. LOCAL NETS AND QUANTUM FIELDS

First some necessary definitions are given. In particular we define what a local net of algebras and a quantum field are to be in this paper.

We assume that there is a strongly continuous unitary representation T_x of the group $\mathbb{R}^4 \ni x$ on a separable Hilbert space $\mathcal{H}$. Further we assume that there is a vacuum vector $\omega \in \mathcal{H}$, i.e., $T_x\omega = \omega$ for all $x \in \mathbb{R}^4$. We assume that the spectrum of T_x is contained in the closed forward cone $V_+ := \{x \in \mathbb{R}^4; x_0 \geqslant 0, x_0^2 - \bar{x}^2 \geqslant 0\}$, with $x = \{x_0, \bar{x}\}$, where x_0 is the time component of x. The self-adjoint generators of T_x are denoted by $P = \{H, P_1, P_2, P_3\}$. We use the short notation $R := (1 + H)^{-1}$ (note $H \geqslant 0$) and $C^\infty(H) := \cap_{n=0}^{\infty} R^n \mathcal{H}$.

A four-tuple $\{\mathcal{H}, T_x, \omega, \mathcal{A}\}$ is called a local net of algebras if it satisfies the following assumptions.

(A1) To each open set $\mathcal{O} \subset \mathbb{R}^4$ there is assigned a von Neumann algebra $\mathcal{A}(\mathcal{O})$ on $\mathcal{H}$ and $\mathcal{A} = \cup_{\mathcal{O}} \mathcal{A}(\mathcal{O})$.

(A2) $\mathcal{O}_1 \subseteq \mathcal{O}_2$ implies $\mathcal{A}(\mathcal{O}_1) \subseteq \mathcal{A}(\mathcal{O}_2)$ (isotony).

(A3) $T_x \mathcal{A}(\mathcal{O}) T_{-x} = \mathcal{A}(\mathcal{O} + x), x \in \mathbb{R}^4$ (translation covariance).

(A4) ω is cyclic and separating for each algebra $\mathcal{A}(\mathcal{O})$, where $\mathcal{O}$ and its causal complement $\mathcal{O}_c$ (the interior of the set $\{x \in \mathbb{R}^4; (x_0 - y_0)^2 - (\bar{x} - \bar{y})^2 < 0, y \in \mathcal{O}\}$) are nonempty.

(A5) Let $\mathcal{O}_n$ be a sequence of open sets of $\mathbb{R}^4$ such that $\mathcal{O}_n \supseteq \mathcal{O}_{n+1}$, $\cap_n \mathcal{O}_n \subseteq \overline{\mathcal{O}}$. Then $\cap_n \mathcal{A}(\mathcal{O}_n) \subseteq \mathcal{A}(\mathcal{O})$ (continuity from the outside).

We call the local net of algebras causal if $\mathcal{O}_2 \subset \mathcal{O}_{1,c}$ implies $A_1 A_2 = A_2 A_1$ for all $A_1 \in \mathcal{A}(\mathcal{O}_1)$ and $A_2 \in \mathcal{A}(\mathcal{O}_2)$. We call the local net of algebras Poincaré covariant if there is a representation by unitary operators $U(g)$ of the restricted Poincaré group $\mathcal{P}^{\uparrow}_{+} \ni g$ such that $T_x = U(\{x, 1\})$ and (A3) is true with T_x, x replaced by $U(g), g$. The local net of algebras is called CPT covariant if there is a CPT operator θ ($\theta\omega = \omega, \theta^2 = 1, \theta T_x = T_{-x}\theta, \theta$ is antiunitary) such that $\theta \mathcal{A}(\mathcal{O}) \theta = \mathcal{A}(-\mathcal{O})$. Causality, CPT covariance, and Poincaré covariance are not included in the list of assump-

 0022-2488/88/092106-06$02.50

tions on the local net because we do not need to make these assumptions to prove the results in the following sections.

Next we introduce smooth subalgebras of $\mathcal{A}(\mathcal{O})$ and $\mathcal{A}(\mathcal{O})'$ [the commutant of $\mathcal{A}(\mathcal{O})$]. We write $A\in\mathcal{A}_s(\mathcal{O})'$ [$A\in\mathcal{A}_s(\mathcal{O})$] if $A\in\mathcal{A}(\mathcal{O})'$ [$A\in\mathcal{A}(\mathcal{O})$], $AC^\infty(H)\subseteq C^\infty(H)$, $A^*C^\infty(H)\subseteq C^\infty(H)$, R^mAR^{-m}, and $R^mA^*R^{-m}$ are bounded for all $m=0,1,2,\ldots$. Further we introduce

$$\mathcal{C}(x):=\bigcup_{\mathcal{O}\ni x}\mathcal{A}(\mathcal{O})',\quad \mathcal{C}_s(x):=\bigcup_{\mathcal{O}\ni x}\mathcal{A}_s(\mathcal{O})'.$$

Using (A2), (A3), and (A5) we obtain that $\mathcal{A}_s(\mathcal{O})'$ is strongly $*$-dense in $\mathcal{A}(\mathcal{O})'$ and therefore $\mathcal{C}_s(x)$ is strongly $*$-dense in $\mathcal{C}(x)$. Naturally the algebras $\mathcal{A}_s(\mathcal{O})$, $\mathcal{A}_s(\mathcal{O})'$, $\mathcal{C}_s(x)$, and $\mathcal{C}(x)$ transform covariantly under the action of the translation group.

Now we introduce the notion of a quantum field. Our notion is very wide. In particular the domain of the field operators $A(f)$ can depend on the support of f. This is in general necessary to have a tight connection to the local net of algebras.

A five-tuple $\{\mathcal{H}, T_x, \omega, \mathcal{D}(\cdot), A(\cdot)\}$ is called a quantum field if it satisfies the following assumptions.

(Q1) For each open bounded set $\mathcal{O}\subset\mathbf{R}^4$ there is a dense set $\mathcal{D}(\mathcal{O})$ in $\mathcal{H}$ such that $\mathcal{D}(\mathcal{O})\subseteq C^\infty(H)$, $T_a\mathcal{D}(\mathcal{O})=\mathcal{D}(\mathcal{O}+a)$, and $\omega\in\mathcal{D}(\mathcal{O})$.

(Q2) $A(\cdot): C_0^\infty(\mathbf{R}^4)\ni f\to A(f)$ is a linear map into the set of closable operators on $\mathcal{H}$. For all $f\in C_0^\infty(\mathbf{R}^4)$ and all open sets $\mathcal{O}$ such that $\overline{\mathcal{O}}\supseteq\operatorname{supp} f$ we have $\mathcal{D}(\mathcal{O})\subseteq\operatorname{dom}A(f)$ and $A(f)^*\restriction\mathcal{D}(\mathcal{O})=A(\bar f)\restriction\mathcal{D}(\mathcal{O})$. We can put $\operatorname{dom}A(f)=\mathcal{D}(\mathcal{O}_f)$ with $\overline{\mathcal{O}}_f=\operatorname{supp} f$.

(Q3) For fixed $\mathcal{O}\subseteq\mathbf{R}^4$ and $u\in\mathcal{D}(\mathcal{O})$ the map $C_0^\infty(\mathcal{O})\ni f\to A(f)u$ is strongly continuous. For vectors $v\in\bigcap_{\mathcal{O}}\mathcal{D}(\mathcal{O})$ the map can be extended to a strongly continuous map $\mathcal{S}(\mathbf{R}^4)\ni f\to A(f)v$.

(Q4) Let $a\in\mathbf{R}^4$, $f\in C_0^\infty(\mathcal{O})$, and $u\in\mathcal{D}(\mathcal{O}+a)$. Then $T_aA(f)T_{-a}u=A(f_a)u$ with $f_a(x)=f(x-a)$. For vectors $u\in\bigcap_{\mathcal{O}}\mathcal{D}(\mathcal{O})$ this holds also for $f\in\mathcal{S}(\mathbf{R}^4)$.

We call the quantum field causal if

$$(A(f_1)u,A(f_2)v)=(A(\bar f_2)u,A(\bar f_1)v),$$

for all vectors $u,v\in\mathcal{D}(\mathcal{O}_1)\cap\mathcal{D}(\mathcal{O}_2)$ with $\operatorname{supp}f_i\subset\mathcal{O}_i$ and $\mathcal{O}_2\subset\mathcal{O}_{1,c}$. We call the quantum field Poincaré covariant if there is a strongly continuous unitary representation $U(g)$ of the restricted Poincaré group $\mathcal{P}^\uparrow_+\ni g$ such that $T_x=U(\{x,1\})$ and (Q4) is true with T_x,x replaced by $U(g),g$. The quantum field is called CPT covariant if there is a CPT operator θ with $\theta\mathcal{D}(\mathcal{O})\subseteq\mathcal{D}(-\mathcal{O})$ and $\theta A(f)\theta=A(f_\theta)$, where $f_\theta=\overline{f(-x)}$.

A special class of quantum fields is of considerable interest. A quantum field is called energetic bounded if there is a natural number s such that $\|R^sA(f)R^s\|\leq|f|_\alpha$, where $|\cdot|_\alpha$ is some Schwartz norm.[1] In this case we can assume that $\mathcal{D}(\mathcal{O})=C^\infty(H)$ for all open bounded sets $\mathcal{O}\subseteq\mathbf{R}^4$ without loss of generality.

If a quantum field is to belong to the same theory as the local net of algebras then we expect the following relation for all $f\in C_0^\infty(\mathcal{O})$, $C\in\mathcal{A}_s(\mathcal{O})'$, and $u,v\in\mathcal{D}(\mathcal{O})$, where $\mathcal{O}$ is an arbitrary open bounded set of $\mathbf{R}^4$:

$$(A(\bar f)v,Cu)=(v,CA(f)u). \tag{1}$$

We say the quantum field $\{\mathcal{H}, T_x, \omega, \mathcal{D}(\cdot), A(\cdot)\}$ is relatively local to $\{\mathcal{H}, T_x, \omega, \mathcal{A}\}$ if (1) is fulfilled. In this case we shortly say that $A(\cdot)$ belongs to $Q_{\mathcal{A}}$.

Now we introduce the sets $\tilde{\mathcal{D}}_{\mathcal{A}}(\mathcal{O}):=\{u=Bv;\ v\in\mathcal{D}(\mathcal{O}), B\in\mathcal{A}_s(\mathcal{O})'\}$ and $\mathcal{D}_{\mathcal{A}}(\mathcal{O}):=\mathcal{A}_s(\mathcal{O})'\omega$. Then we have the relations $\mathcal{D}(\mathcal{O})\subseteq\tilde{\mathcal{D}}_{\mathcal{A}}(\mathcal{O})\supseteq\mathcal{D}_{\mathcal{A}}(\mathcal{O})$. It is not difficult to verify that $\{\mathcal{H}, T_x, \omega, \tilde{\mathcal{D}}_{\mathcal{A}}(\cdot), \tilde A(\cdot)\}$ and $\{\mathcal{H}, T_x, \omega, \mathcal{D}_{\mathcal{A}}(\cdot), \tilde A(\cdot)\restriction\mathcal{D}_{\mathcal{A}}(\cdot)\}$ are quantum fields [in the case where $A(\cdot)$ belongs to $Q_{\mathcal{A}}$], where $\tilde A(f):=A(\bar f)^*\restriction\tilde{\mathcal{D}}_{\mathcal{A}}(\mathcal{O})$ for $\operatorname{supp}f=\overline{\mathcal{O}}$. The proof follows easily from (1) (see the proof of Proposition 1 in Ref. 3). The three quantum fields differ in their domains. The main point is that the extensions $\tilde A(f)$ and $\tilde A(f)\restriction\mathcal{D}_{\mathcal{A}}(\mathcal{O})$, where $\operatorname{supp}f=\overline{\mathcal{O}}$, are affiliated to the von Neumann algebra $\mathcal{A}(\mathcal{O})$. Thus if the local net is causal then the quantum fields belonging to $Q_{\mathcal{A}}$ are also causal.

For the algebras $\mathcal{A}(\mathcal{O})$ we can introduce the conjugation operators $S_{\mathcal{O}}$ defined by

$$S_{\mathcal{O}}A\omega:=A^*\omega,\quad A\in\mathcal{A}(\mathcal{O}). \tag{2}$$

If $\mathcal{O}$ and $\mathcal{O}_c$ are nonempty, then ω is cyclic and separating for $\mathcal{A}(\mathcal{O})$ by assumption (A2). Thus the operators $S_{\mathcal{O}}$ are closable and antilinear. The polar decomposition $S_{\mathcal{O}}=J_{\mathcal{O}}\Delta_{\mathcal{O}}^{1/2}$ defines the modular operator $\Delta_{\mathcal{O}}$, the modular conjugation $J_{\mathcal{O}}$, and the modular group $\Delta_{\mathcal{O}}^{it}$, $t\in\mathbf{R}$. The conjugation operator for the algebra $\mathcal{A}(\mathcal{O})'$ is $S^*_{\mathcal{O}}\restriction\mathcal{A}(\mathcal{O})'\omega$ (for details on these operators see Ref. 5).

Let $\mathcal{O}_n$ be a sequence of open bounded sets of $\mathbf{R}^4$ such that $\mathcal{O}_n$ shrinks to the point zero. Then we call $\mathcal{O}_n$ a shrinking sequence. A sequence $f_n\in\mathcal{S}(\mathbf{R}^4)$ such that $f_n(x)=\overline{f_n(x)}$ and $f_n*\varphi$ tends to φ in the Schwartz topology for each $\varphi\in\mathcal{S}(\mathbf{R}^4)$ is called a δ-function approximating sequence. We set $f_{n,x}(y):=f_n(y-x)$. Note that for each shrinking sequence $\mathcal{O}_n$ there is a δ-function approximating sequence f_n such that $\operatorname{supp}f_n\subseteq\mathcal{O}_n$.

III. ENERGETIC WEIGHTED SPACES

Let n be a natural number. Then we introduce scalar products and norms on $C^\infty(H)$ by

$$(u,v)_n:=(R^{-n}u,R^{-n}v),\quad \|u\|_n:=\|R^{-n}u\|,$$
$$(u,v)_{-n}:=(R^nu,R^nv),\quad \|u\|_{-n}:=\|R^nu\|.$$

The corresponding Hilbert spaces (energetic weighted spaces) are denoted by $\mathcal{H}_n$ and $\mathcal{H}_{-n}$, respectively. In the following, operators on $\mathcal{H}$, $\mathcal{H}_n$, and $\mathcal{H}_{-n}$ are denoted by the same symbol if they agree on a dense set $\mathcal{D}\subseteq\mathcal{H}_n\subset\mathcal{H}\subset\mathcal{H}_{-n}$ (dense in each of these spaces).

It is obvious that T_x is also a unitary representation of $\mathbf{R}^4\ni x$ on $\mathcal{H}_n,\mathcal{H}_{-n}$. The operators C from $\mathcal{A}_s(\mathcal{O})'$, $\mathcal{A}_s(\mathcal{O})$, $\mathcal{C}_s(x)$ are bounded operators on $\mathcal{H}_n,\mathcal{H}_{-n}$. If we have a strongly continuous unitary representation $U(g)$ on $\mathcal{H}$ of the restricted Poincaré group $\mathcal{P}^\uparrow_+$ with $U(\{x,1\})=T_x$ then the operators $U(g)$ are also bounded operators on $\mathcal{H}_n,\mathcal{H}_{-n}$. This follows from the commutation relation with $\exp\{-itH\}$.

Proposition 1: Let $\mathcal{O}\subset\mathbf{R}^4$ be an open set such that $\mathcal{O}$ and $\mathcal{O}_c$ are nonempty. Then the sets $\mathcal{D}_{\mathcal{A}}(\mathcal{O})=\mathcal{A}_s(\mathcal{O})'\omega$ and

$\mathscr{A}_s(\mathscr{O})\omega$ are dense in $\mathscr{H}_n$, $\mathscr{H}_{-n}$ for each natural number n.

Proof: (1) $\mathscr{H}_{-n}$. Let $\mathscr{B}$ be a dense set in $\mathscr{H}$. Then $\mathscr{B}$ is also contained in $\mathscr{H}_{-n}$ and is dense in $\mathscr{H}_{-n}$ because of the simple estimate $\|u\|_{-n} = \|R^n u\| \leqslant \|u\|$, for $u \in \mathscr{H} \subseteq \mathscr{H}_{-n}$, and the fact that $\mathscr{H}$ is dense in $\mathscr{H}_{-n}$. Since $\mathscr{A}(\mathscr{O})'\omega$ is dense in $\mathscr{H}$ and $\mathscr{A}_s(\mathscr{O})'$ is dense (strongly $*$) in $\mathscr{A}(\mathscr{O})'$ we obtain that $\mathscr{A}_s(\mathscr{O})'\omega$ is dense in $\mathscr{H}$. This shows that $\mathscr{A}_s(\mathscr{O})'\omega$ is dense in $\mathscr{H}_{-n}$. For $\mathscr{O}$ there is an open nonempty set $\mathscr{O}_1 \subset \mathscr{O}$ and $\epsilon > 0$ such that $\{\mathscr{O}_1 + x;\ |x| < \epsilon\} \subset \mathscr{O}$. Then for each element $C \in \mathscr{A}(\mathscr{O}_1)$ there is a sequence $C_n \in \mathscr{A}_s(\mathscr{O})$ such that C_n tends strongly to C [take

$$C_n = \int f_n(x) T_x C T_{-x}\, dx$$

with f_n a δ-function approximating sequence]. Since $\mathscr{A}(\mathscr{O}_1)\omega$ is dense in $\mathscr{H}$ we find that $\mathscr{A}_s(\mathscr{O})\omega$ is dense in $\mathscr{H}$. Thus $\mathscr{A}_s(\mathscr{O})\omega$ is dense in $\mathscr{H}_{-n}$.

(2) $\mathscr{H}_n$. Assume $u \in \mathscr{H}_n$ and $(u, C\omega)_n = 0$ for all $C \in \mathscr{A}_s(\mathscr{O})'$. For $\mathscr{O}$ there exist a set $\mathscr{O}_1$ and $\epsilon > 0$ such that $\mathscr{O}_{1,c}$ is nonempty and $\{\mathscr{O} + x;\ |x| < \epsilon\} \subset \mathscr{O}_1$. Thus

$$0 = (u, T_x C T_{-x}\omega)_n = (T_{-x}u, C\omega)_n\,,$$

for all $|x| < \epsilon$ and $C \in \mathscr{A}_s(\mathscr{O}_1)'$. Here $T_{-x}u$ is strongly continuous in x. By integration we get

$$(2\pi)^{-2} \int f(x)(T_{-x}u, C\omega)_n\, ds = (\hat{f}(P)u, C\omega)_n = 0\,,$$

for all $f \in C_0^\infty(\mathscr{O}_\epsilon)$, $\mathscr{O}_\epsilon = \{x;\ |x| < \epsilon\}$. Thus $(R^{-2n}\hat{f}(P)u, C\omega) = 0$. Since the set of vectors $\mathscr{A}_s(\mathscr{O}_1)'\omega$ is dense in $\mathscr{H}$ we obtain $R^{-2n}\hat{f}(P)u = 0$. We consider a function f such that $\hat{f}(p) \neq 0$, for all $p \in \mathbb{R}^4$, and $\operatorname{supp} f \subset \mathscr{O}_\epsilon$ (such a function exists). Thus for this function $R^{-2n}\hat{f}(P)u = 0$ is only possible if $u = 0$. This implies that $\mathscr{A}_s(\mathscr{O})'\omega$ is dense in $\mathscr{H}$. Similarly we can prove that $\mathscr{A}_s(\mathscr{O})'\omega$ is dense in $\mathscr{H}_n$. ■

From the proof of this proposition it also follows that the sets $R^m \mathscr{A}_s(\mathscr{O})'\omega$, $R^m \mathscr{A}_s(\mathscr{O})\omega$ are dense in $\mathscr{H}$, $\mathscr{H}_n$, and $\mathscr{H}_{-n}$ for all $m = 0, \pm 1, \pm 2, \ldots$ and $n = 1, 2, \ldots$. Further we immediately see that the sets $\mathscr{A}(\mathscr{O})\omega$ and $\mathscr{A}(\mathscr{O})'\omega$ are dense in $\mathscr{H}_{-n}$.

Proposition 2: Let $\mathscr{O} \subset \mathbb{R}^4$ be an open set such that $\mathscr{O}$ and its causal complement $\mathscr{O}_c$ are nonempty. Then $S_{\mathscr{O}} \restriction \mathscr{A}(\mathscr{O})\omega$ is a closable operator on $\mathscr{H}_{-n}$ for each natural number n.

Proof: Since $S_{\mathscr{O}} A\omega = A^*\omega \in \mathscr{H} \subset \mathscr{H}_{-n}$, for all $A \in \mathscr{A}(\mathscr{O})$, and $\mathscr{A}(\mathscr{O})\omega$ is dense in $\mathscr{H}_{-n}$ the operator $S_{\mathscr{O}}$ is well defined in $\mathscr{H}_{-n}$. Further we have for all $B \in \mathscr{A}_s(\mathscr{O})'$

$$\begin{aligned} R^{-n} S_{\mathscr{O}}^* R^n R^{-n} B\omega &= R^{-n} S_{\mathscr{O}}^* B\omega \\ &= R^{-n} B^* \omega = R^{-n} B^* R^n \omega\,. \end{aligned}$$

Since the set $R^{-n}\mathscr{A}_s(\mathscr{O})'\omega$ is dense in $\mathscr{H}$ (see the remark after Proposition 1) the operator $R^{-n} S_{\mathscr{O}}^* R^n$ is well defined on a dense set in $\mathscr{H}$ for each natural number n. Now it is not hard to show that the adjoint operator $S_{\mathscr{O}}^{(*)}$ in $\mathscr{H}_{-n}$ is connected with the adjoint operator $S_{\mathscr{O}}^*$ in $\mathscr{H}$ by the relation $S_{\mathscr{O}}^{(*)} = R^{-2n} S_{\mathscr{O}}^* R^{2n}$ and $S_{\mathscr{O}}^{(*)}$ exists on a dense set in $\mathscr{H}_{-n}$ if $R^{-n} S_{\mathscr{O}}^* R^n$ exists on a dense set in $\mathscr{H}$. Thus we obtain that the adjoint of $S_{\mathscr{O}}$ in $\mathscr{H}_{-n}$ exists and therefore $S_{\mathscr{O}}$ is a closable operator in $\mathscr{H}_{-n}$. ■

We denote the closure of $S_{\mathscr{O}} \restriction \mathscr{A}(\mathscr{O})\omega$ in $\mathscr{H}_{-n}$ by the symbol $S_{\mathscr{O},n}$.

IV. GENERATING VECTORS FOR QUANTUM FIELDS

In this section it will be shown that a quantum field relatively local to a given local net of algebras is essentially given by one vector. This vector is, roughly speaking, the quantum field at the point zero applied to the vacuum. The vector is invariant under the conjugation operators $S_{\mathscr{O}}$ with $\overline{\mathscr{O}} \ni 0$. Conversely we will see that a vector with such invariance properties defines a quantum field related to the given local net. If the local net satisfies causality then this quantum field is also causal.

Proposition 3: Let $\{\mathscr{H}, T_a, \omega, \mathscr{D}(\cdot), B(\cdot)\}$ be a quantum field belonging to $Q_{\mathscr{A}}$. Then there is a natural number n and a vector $\omega_B \in \mathscr{H}_{-n}$ such that

$$\|B(f_m)\omega - \omega_B\|_{-n} \to 0 \quad \text{as } m \to \infty\,, \tag{3}$$

for each δ-function approximating sequence f_m.

Proof: From the strong continuity of the map $\mathscr{S}(\mathbb{R}^4)$ $\ni f \to B(f)\omega$ it follows that $\|B(f)\omega\| \leqslant |f|_\alpha$ for some Schwartz norm $|\cdot|_\alpha$. This implies that the map $f \to B(f)\omega$ can be continuously extended to the space of functions with the norm $|\cdot|_\alpha$. Using the properties of the Schwartz norm $|\cdot|_\alpha$ and of the sequence f_m we find that there are some positive numbers n_0, n_1, n_2, n_3 such that $|h|_\alpha < c < \infty$ with

$$\hat{h}(p) := (2\pi)^{-2} \prod_{j=0}^{3} (i + p_j)^{-n_j}$$

and $f_m * h$ tends in $|\cdot|_\alpha$ norm to h. This implies

$$\|B(h * f_m)\omega - B(h)\omega\| \to 0 \quad \text{as } m \to \infty\,.$$

From spectrality we obtain that $R^n \Pi_{j=0}^3 (i + P_j)^{n_j}$ is bounded for $n \geqslant n_0 + n_1 + n_2 + n_3$. This gives the estimate

$$\begin{aligned} &\left\| R^n B(f_m)\omega - R^n \prod_{j=0}^{3} (i + P_j)^{n_j} B(h)\omega \right\| \\ &\leqslant \left\| R^n \prod_{j=0}^{3} (i + P_j)^{n_j} \right\| \\ &\quad \times \left\| \prod_{j=0}^{3} (i + P_j)^{-n_j} B(f_m)\omega - B(h)\omega \right\| \\ &\leqslant \left\| R^n \prod_{j=0}^{3} (i + P_j)^{n_j} \right\| \| B(h * f_m)\omega - B(h)\omega \|\,, \end{aligned}$$

where we have used the relation $\hat{h}(P)B(f)\omega = B(h * f)\omega$, which follows from the strong continuity of the vector function $B(f)\omega$ and translation covariance. Setting

$$\omega_B := \prod_{j=0}^{3} (i + P_j)^{n_j} B(h)\omega\,,$$

we obtain the result. ■

The smallest natural number n such that Proposition 3 holds is called the energetic dimension of the quantum field; ω_B is called the generating vector of the quantum field.

Proposition 4: Let n be the energetic dimension and let ω_B be the generating vector of the quantum field $\{\mathscr{H}, T_a, \omega, \mathscr{D}(\cdot), B(\cdot)\}$ from $Q_{\mathscr{A}}$. Then we have the following.

(i) If the quantum field is CPT covariant, then ω_B is CPT invariant, i.e., $\theta\omega_B = \omega_B$.

(ii) If the quantum field is Poincaré covariant, then ω_B is Lorentz invariant, i.e., $U(\{0,\Lambda\})\omega_B = \omega_B$, $\{0,\Lambda\}\in\mathscr{P}^{\uparrow}_{+}$.

(iii) Let $\mathscr{O}\subset\mathbb{R}^4$ be an open set such that $\mathscr{O}$ and $\mathscr{O}_c$ are nonempty and $\overline{\mathscr{O}}\ni 0$. Then $\omega_B\in\operatorname{dom} S_{\mathscr{O},n}$ and $S_{\mathscr{O},n}\omega_B = \omega_B$.

Proof: (i) We have $\theta\omega = \omega$ and $\theta B(f)\omega = B(f_\theta)\omega$ with $f_\theta(x) = \overline{f(-x)}$. Further θ commutes with R. Thus θ is also an antiunitary operator in $\mathscr{H}_{-n}$. If f_m is a δ-function approximating sequence, then $f_{m,\theta}$ is also a δ-function approximating sequence. Thus

$$\|\theta B(f_m)\omega - \omega_B\|_{-n} = \|B(f_{m,\theta})\omega - \omega_B\|_{-n}\to 0$$

as $m\to\infty$. On the other side $\theta B(f_m)\omega$ tends to $\theta\omega_B$ because θ is a bounded operator in $\mathscr{H}_{-n}$. This implies $\theta\omega_B = \omega_B$.

(ii) The reasoning is similar to that in (i). We note that $U(\{0,\Lambda\})B(f_m)\omega = B(f_{m,\Lambda})\omega$, where $f_{m,\Lambda}$ is again a δ-function approximating sequence. Further, $U(\{0,\Lambda\})$ is a bounded operator in $\mathscr{H}_{-n}$. This implies that $U(\{0,\Lambda\})\omega_B = \omega_B$.

(iii) Because of $\overline{\mathscr{O}}\ni 0$ there is a δ-function approximating sequence f_m such that $\operatorname{supp} f_m\subset\mathscr{O}$. The operators $\tilde{B}(f_m)$ are affiliated to $\mathscr{A}(\mathscr{O})$ (see Sec. II). Let $E_m(\Delta)$ be the spectral projections of $|\tilde{B}(f_m)^*|$. Further let $\tilde{B}(f_m)^* = U_m|\tilde{B}(f_m)^*|$ be the polar decomposition of $\tilde{B}(f_m)^*$. Let $W_m(\Delta) := U_m E_m(\Delta)U_m^*$. Then a simple calculation shows that the operators

$$B_m(\Delta) := \tfrac{1}{2}[E_m(\Delta)\tilde{B}(f_m) + W_m(\Delta)\tilde{B}(f_m)^*]$$

are self-adjoint and belong to $\mathscr{A}(\mathscr{O})$. Choosing a suitable sequence $\Delta_m\to\mathbb{R}$ one easily proves that

$$\|B_m(\Delta_m)\omega - B(f_m)\omega\|_{-n}\to 0 \quad \text{as } m\to\infty\,.$$

Here we have used that $\tilde{B}(f_m)\omega = B(f_m)\omega = \tilde{B}(f_m)^*\omega$ and $E_m(\Delta_m)$ and $W_m(\Delta_m)$ tend strongly to 1. Using $B_m(\Delta_m)\in\mathscr{A}(\mathscr{O})$ and $\mathscr{A}(\mathscr{O})\omega\subset\operatorname{dom} S_{\mathscr{O},n}$ this gives

$$S_{\mathscr{O},n}B_m(\Delta_m)\omega = B_m(\Delta_m)\omega\to\omega_B\,.$$

Since $S_{\mathscr{O},n}$ is closed in $\mathscr{H}_{-n}$ this proves $\omega_B\in\operatorname{dom} S_{\mathscr{O},n}$ and $S_{\mathscr{O},n}\omega_B = \omega_B$. ■

Next we need a technical result.

Lemma 1: Let $\omega_1\in\mathscr{H}_{-n}$. Assume that for a shrinking sequence $\mathscr{O}_m$, $\omega_1\in\operatorname{dom} S_{\mathscr{O}_m,n}$ and $S_{\mathscr{O}_m,n}\,\omega_1 = \omega_1$. Then the functional $\Omega_1(\cdot)$ on $\mathscr{C}_s(0)$ defined by

$$\Omega_1(A) := (R^n\omega_1, R^{-n}A\omega),\quad A\in\mathscr{C}_s(0)\,,$$

is Hermitian, i.e., $\Omega_1(A) = \overline{\Omega_1(A^*)}$.

Proof: Let $A\in\mathscr{C}_s(0)$. Thus there is some $\mathscr{O}\ni 0$ such that $A\in\mathscr{O}_s(\mathscr{O})'$. Since $\mathscr{O}_m$ is a shrinking sequence there is an index j such that $\mathscr{O}_j\subset\mathscr{O}$. This implies that $S^*_{\mathscr{O}_j}\supset S^*_{\mathscr{O}}$ and $S^*_{\mathscr{O}_j}A\omega = A^*\omega$. Further we use that the adjoint of $S_{\mathscr{O}_j,n}$ in $\mathscr{H}_{-n}$ is defined and given by $R^{-2n}S^*_{\mathscr{O}_j,n}R^{2n}$ on the set $R^{-2n}\mathscr{A}_s(\mathscr{O}_j)\omega\ni R^{-2n}A\omega$. This gives

$$\begin{aligned}
\Omega_1(A) &= (R^n\omega_1, R^{-n}A\omega)\\
&= (R^nS_{\mathscr{O}_j,n}\omega_1, R^{-n}A\omega)\\
&= (S_{\mathscr{O}_j,n}\omega_1, R^{-2n}A\omega)_{-n}\\
&= (\overline{\omega_1, S^{(*)}_{\mathscr{O}_j,n}R^{-2n}A\omega})_{-n}\\
&= \overline{(R^n\omega_1, R^{-n}S^*_{\mathscr{O}_j}A\omega)}\\
&= \overline{(R^n\omega_1, R^{-n}A^*\omega)} = \Omega_1(A^*)\,.
\end{aligned}$$ ■

Now we come to the main result of this section.

Theorem 1: Let $\omega_1\in\mathscr{H}_{-n}$ for some natural number n. Assume there is a shrinking sequence $\mathscr{O}_m$ such that $\omega_1\in\operatorname{dom}\mathscr{S}_{\mathscr{O}_m,n}$ and $\mathscr{S}_{\mathscr{O}_m,n}\omega_1 = \omega_1$. Then the corresponding operator $A(0)$ from $\mathscr{H}_n$ into $\mathscr{H}_{-n}$ defined by

$$A(0)C\omega := C\omega_1,\quad C\in\mathscr{C}_s(0)\,, \tag{4}$$

is closable and defines a quantum field $\{\mathscr{H}, T_a, \omega, \mathscr{D}_{\mathscr{A}}(\cdot), A(\cdot)\}$ with

$$A(f)u := \int f(x)T_xA(0)T_{-x}u\,dx$$

(where the integral exists in strong sense in $\mathscr{H}_{-n}$ and gives a vector from $\mathscr{H}$). This quantum field belongs to $Q_{\mathscr{A}}$.

(i) If the local net of algebras is CPT covariant with the CPT operator θ and $\theta\omega_1 = \omega_1$, then the quantum field is CPT covariant.

(ii) If the local net of algebras is Poincaré covariant with the group representation $U(g)$ and $U(\{0,\Lambda\})\omega_1 = \omega_1$, $\{0,\Lambda\}\in\mathscr{P}^{\uparrow}_{+}$, then the quantum field is Poincaré covariant.

(iii) The quantum field is energetic bounded if and only if there is some natural number $n'\geqslant n$ such that

$$\|R^{n'}C\omega_1\|\leqslant b\,\|R^{-n'}C\omega\|,\quad C\in\mathscr{C}_s(0)\,,$$

for some constant b.

Proof: (1) Because of Lemma 1 the functional $\Omega_1(C) = (R^n\omega_1, R^{-n}C\omega)$ is Hermitian. Further, the operator $A(0)$ is densely defined because $\mathscr{C}_s(0)\omega$ is dense in $\mathscr{H}_n$. From the fact that ω is separating for all $\mathscr{A}(\mathscr{O})'\supset\mathscr{A}_s(\mathscr{O})'$ and $\mathscr{C}_s(0) = \cup_{\mathscr{O}\ni 0}\mathscr{A}_s(\mathscr{O})'$ we obtain that $C\omega = 0$ implies $C = 0$ for $C\in\mathscr{C}_s(0)$. Thus $A(0)C\omega = C\omega_1 = 0$ if $C\omega = 0$. Therefore $A(0)$ is well defined. Correspondingly $A(x) := T_xA(0)T_{-x}$ is well defined on $\mathscr{C}_s(x)\omega = T_x\mathscr{C}_s(0)\omega$ and $A(x)C\omega$ is strongly continuous in $x\in\mathscr{O}$ for $C\in\mathscr{A}_s(\mathscr{O})'\subset\cap_{x\in\mathscr{O}}\mathscr{C}_s(x)$. The closability of $A(0)$ follows from the fact that the adjoint of $A(0)$ exists and is given by $A(0)^* = R^{2n}A(0)R^{2n}$ on the set $R^{-2n}\mathscr{C}_s(0)\omega$, which is dense in $\mathscr{H}_{-n}$. For the proof of this fact we use the chain of equalities

$$\begin{aligned}
&(A(0)C_2\omega, R^{-2n}C_1\omega)_{-n}\\
&\quad= (R^{2n}C_2\omega_1, R^{-2n}C_1\omega)\\
&\quad= (R^{2n}\omega_1, R^{-2n}C_2^*C_1\omega)\\
&\quad= \Omega_1(C_2^*C_1) = \overline{\Omega_1(C_1^*C_2)}\\
&\quad= (R^{-2n}C_1^*C_2\omega, R^{2n}\omega_1)\\
&\quad= (R^{-2n}C_2\omega, R^{2n}A(0)C_1\omega)\\
&\quad= (C_2\omega, R^{2n}A(0)R^{2n}R^{-2n}C_1\omega)_n\,.
\end{aligned}$$

(2) Let $f\in C_0^\infty(\mathcal{O})$. The operators $A(x)$, $x\in\mathcal{O}$, have the common domain $\mathcal{A}_s(\mathcal{O})'\omega$ and are strongly continuous in $x\in\mathcal{O}$ on this domain. Thus

$$A(f)C\omega := \int f(x)A(x)C\omega\,dx = C\int f(x)A(x)\omega\,dx$$
$$= (2\pi)^2 C\hat{f}(P)\omega_1$$

is well defined. Since $\hat{f}(P)\omega_1\in\mathcal{H}$, the vector $A(f)C\omega$ is also from $\mathcal{H}$. For vectors $v\in\cap_{\mathcal{O}}\mathcal{A}_s(\mathcal{O})'\omega$ this equation can be extended to all functions $f\in\mathcal{S}(\mathbb{R}^4)$. That the five-tuple $\{\mathcal{H},T_x,\omega,\mathcal{D}_{\mathcal{A}}(\cdot),A(\cdot)\}$ defines a quantum field is an easy consequence of this equation and the properties of $A(x)$ (translation covariance, continuity in x, special form of the adjoint operator). The commutation relation $A(f)Cu = CA(f)u$ for vectors $u\in\mathcal{D}_{\mathcal{A}}(\mathcal{O})$, $C\in\mathcal{A}_s(\mathcal{O})'$, and $f\in C_0^\infty(\mathcal{O})$ follows from the definition of $A(f)$ and implies that the quantum field belongs to $Q_{\mathcal{A}}$.

(3) We omit the simple proofs of (i) and (ii).

(4) Assume that $A(\cdot)$ is energetic bounded. This implies that $R^n A(x)R^n$ is bounded as an operator on $\mathcal{H}$.[1] Thus

$$\|R^{n'}C\omega_1\| = \|R^{n'}A(0)C\omega\| = \|R^{n'}A(0)R^{n'}R^{-n'}C\omega\|$$
$$\leqslant \|R^{n'}A(0)R^{n'}\|\,\|R^{-n'}C\omega\|\,.$$

Conversely, assume that $\|R^{n'}C\omega_1\|\leqslant c\|R^{-n'}C\omega\|$, for $n'\geqslant n$. This gives

$$\|R^{n'}A(0)R^{n'}R^{-n'}C\omega\| = \|R^{n'}CR^{-n'}R^{n'}A(0)\omega\|$$
$$= \|R^{n'}C\omega_1\| \leqslant c\|R^{-n'}C\omega\|\,.$$

Thus $R^{n'}A(0)R^{n'}$ is a bounded operator on $\mathcal{H}$. Further, this implies that the corresponding quantum field $A(f)$ is energetic bounded (use similar arguments as in Ref. 1). ■

It remains to show that the generating vectors define their quantum fields via the construction described in Theorem 1.

Theorem 2: Let $\{\mathcal{H},T_x,\omega,\mathcal{D}(\cdot),A(\cdot)\}$ be a quantum field from $Q_{\mathcal{A}}$. Let ω_A be its generating vector. Then the quantum field $\{\mathcal{H},T_x,\omega,\mathcal{D}_{\mathcal{A}}(\cdot),B(\cdot)\}$ generated by ω_A in the meaning of Theorem 1 coincides with the quantum field $\{\mathcal{H},T_x,\mathcal{D}_{\mathcal{A}}(\cdot),\tilde{A}(\cdot)\restriction\mathcal{D}_{\mathcal{A}}(\cdot)\}$ (see Sec. II), i.e., $\overline{\tilde{A}(f)\restriction\mathcal{D}_{\mathcal{A}}(\mathcal{O})} = \overline{B(f)}$ for $\operatorname{supp} f = \overline{\mathcal{O}}$.

Proof: Let f_m be a δ-function approximating sequence and $f_{m,x}(y) = f_m(x-y)$. Then from the definition of $B(x)$ it follows that

$$\operatorname*{s-lim}_{m\to\infty} A(f_{m,x})C\omega$$
$$= C\operatorname*{s-lim}_{m\to\infty} A(f_{m,x})\omega = CT_x \operatorname*{s-lim}_{m\to\infty} A(f_m)\omega$$
$$= CB(x)\omega = B(x)C\omega, \quad C\in\mathcal{C}_s(x)\,.$$

Both $B(x)C\omega$ and $A(f_{m,x})C\omega$ are continuous in $x\in\mathcal{O}$ for $C\in\mathcal{A}_s(\mathcal{O}_m)'$ with $\mathcal{O}_m\supset\mathcal{O}$ shrinking to $\mathcal{O}$ as $m\to\infty$. This gives

$$\operatorname*{s-lim}_{m\to\infty}\int g(x)A(f_{m,x})C\omega\,dx = \int g(x)B(x)C\omega\,dx$$
$$= B(g)C\omega \quad \text{in } \mathcal{H}_{-n},$$

for $g\in C_0^\infty(\mathcal{O})$, $C\in\mathcal{A}_s(\mathcal{O}_m)'$. Now

$$\int g(x)A(f_{m,x})C\omega\,dx = CA(g*f_m)\omega$$

tends to $CA(g)\omega$ because $g*f_m$ goes to g in Schwartz norm. This means that $CA(g)\omega = A(g)C\omega = B(g)C\omega$, for $C\in\mathcal{A}_s(\mathcal{O}_m)'$ and all m. The continuity of the net $\mathcal{A}(\overline{\mathcal{O}})$ from the outside gives the same result for $C\in\mathcal{A}_s(\mathcal{O}')$. ■

From this theorem we learn that the quantum field $A(\cdot)$ is given by its generating vector ω_A up to different extensions. The next assertion says that if the quantum field is energetic bounded then there is only one extension, i.e., in this case the quantum field is uniquely determined by its generating vector.

Proposition 5: Let $\{\mathcal{H},T_x,\omega,C^\infty(H),A(\cdot)\}$ be an energetic bounded quantum field from $Q_{\mathcal{A}}$. Then its restriction to the domains $\mathcal{D}_{\mathcal{A}}(\mathcal{O})\subset C^\infty(H)$ determines uniquely the quantum field, i.e., $\overline{A(f)\restriction\mathcal{D}_{\mathcal{A}}(\mathcal{O})} = \overline{A(f)}$, for $\operatorname{supp} f\subset\overline{\mathcal{O}}$.

Proof: Let $u\in C^\infty(H)$. Because of Proposition 1 there is a sequence $u_m\in\mathcal{A}_s(\mathcal{O})'\omega$ such that $R^{-2n}(u_m-u)\to 0$ in $\mathcal{H}$. We have $\|R^nA(f)R^n\|<\infty$. This implies that $\|A(f)R^{2n}\|<\infty$.[1] We obtain the estimates

$$\|A(f)Cu - A(f)Cu_m\|$$
$$= \|C(A(f)u - A(f)u_m)\|$$
$$= \|CA(f)R^{2n}(R^{-2n}u - R^{-2n}u_m)\|$$
$$\leqslant \|C\|\,\|A(f)R^{2n}\|\,\|R^{-2n}(u-u_m)\|\,.$$

The right side goes to zero as $m\to\infty$. Thus we obtain that $\mathcal{A}_s(\mathcal{O})'\omega = \mathcal{D}_{\mathcal{A}}(\mathcal{O})$ is a core for $A(f)$ with $\operatorname{supp} f\subseteq\overline{\mathcal{O}}$. Furthermore we obtain from the estimate that the commutation relations on the set $\mathcal{D}_{\mathcal{A}}(\mathcal{O})$ imply the commutation relations on the set $C^\infty(H)$. ■

V. CONCLUDING REMARKS

It is not difficult to check that the results of this paper can be generalized. First, we note that the technical assumption of the strong continuity of the map $\mathcal{S}(\mathbb{R}^4)\ni f\to A(f)u$ for vectors $u\in\cap_{\mathcal{O}}\mathcal{D}(\mathcal{O})$ can be omitted. Then we obtain Theorems 1 and 2 with the modification that the spaces $\mathcal{H}_m,\mathcal{H}_{-m}$ have to be replaced by more general "energetic weighted spaces" $\mathcal{H}_n\sim\hat{u}(H)\mathcal{H}$, where $u\in C_0^\infty(\mathbb{R}^4)$ and $\hat{u}$ is invertible. Second, it is not necessary that the local net of algebras is continuous from the outside for all open bounded regions. It is enough to have this property for double cones (the continuity from the outside for double cones follow from the duality for double cones). Third, we can extend the discussion to quantum fields that are not localizable at points but are localizable in strings or other geometrical objects.

Since we have not used causality for the local net of algebras the results can also be applied to field algebras.

In the case where the local net of algebras is conformally covariant and is CPT covariant with CPT operator θ, and

the conjugation operator S_W for the right wedge has the form given by the Bisognano/Wichmann result,[6] the conjugation operators $S_\mathcal{O}$ are given by $S_\mathcal{O} = \theta G_\mathcal{O} = \tilde{G}_\mathcal{O}\theta$, where $G_\mathcal{O}$ and $\tilde{G}_\mathcal{O}$ are determined by the representation of the conformal group.[7] A simple calculation shows that the set of vectors ω_A from $\mathcal{H}_{-m}$ satisfying $S_{\mathcal{O}_n}\omega_A = \omega_A$ for a shrinking sequence $\mathcal{O}_n$ is uniquely determined by the set of vectors $\omega_1 \in \mathcal{H}_{-m}$ with $G_{\mathcal{O}_n}\omega_1 = \tilde{G}_{\mathcal{O}_n}\omega_1 = \pm\omega_1$. Using Theorems 1 and 2 this proves that the set of all quantum fields belonging to the given local net of algebras is determined (in the form of their generating vectors) only by the representation of the conformal group. Details of the simple calculations are omitted. Since only the behavior of the algebras near the space-time point zero has played a role in the arguments we can expect that a kind of asymptotically conformal covariance is sufficient to obtain these results.

ACKNOWLEDGMENT

It is a pleasure to thank Professor D. Buchholz for discussions about some of the topics of this paper.

[1] K. Fredenhagen and J. Hertel, Commun. Math. Phys. **80**, 555 (1981).
[2] J. Rehberg and M. Wollenberg, Math. Nachr. **125**, 1 (1986).
[3] M. Wollenberg, Rep. Math. Phys. **22**, 409 (1985).
[4] W. Driessler, S. J. Summers, and E. H. Wichmann, Commun. Math. Phys. **105**, 49 (1986).
[5] O. Bratteli and D. W. Robinson, *Operator Algebras and Quantum Statistical Mechanics I* (Springer, New York, 1979).
[6] J. J. Bisognano and E. H. Wichmann, J. Math. Phys. **16**, 985 (1975).
[7] P. D. Hislop and R. Longo, Commun. Math. Phys. **84**, 71 (1982).

A power counting formula for short distance singularities in quantum field theory

T. R. Hurd
Department of Mathematics, University of British Columbia, Vancouver, British Columbia V6T 1Y4, Canada

(Received 13 January 1988; accepted for publication 20 April 1988)

A power counting formula for the short distance singularities in quantum field theory is presented. This configuration space version of Weinberg's theorem applies to any renormalized Feynman amplitude, and is proved by a simple extension of the tree expansion method.

I. INTRODUCTION

Weinberg's theorem[1] is a power counting formula that measures the high momentum behavior of a convergent Feynman amplitude. Weinberg's theorem, and the refinement including logarithmic dependence due to Fink,[2] were widely applied in high energy physics even before the divergent case was sorted out in the 1970's by Bergère and Lam.[3]

In Refs. 3 and 4 the asymptotics of a general Feynman amplitude $G(a_k)$ are determined when a subset $\{a_m\} \subseteq \{a_k\}$ of momentum space invariants of the form $(\Sigma p)^2$ or m_i^2 is rescaled to infinity while $\{a_n\} = \{a_k\}/\{a_m\}$ is held fixed. They show that an asymptotic expansion of the form

$$G(\lambda a_m, a_n) \sim \sum_{p=p_{\max}}^{-\infty} \lambda^p \sum_{q=0}^{q_{\max}(p)} \ln^q \lambda \, g_{pq}(a_m, a_n)$$

holds for large λ and gives methods for finding $p_{\max}$, $q_{\max}$, and $g_{pq}(a_m, a_n)$.

In the present paper, I consider a general renormalized Feynman amplitude $V(x_1, x_2, \ldots, x_n)$ in d-dimensional Euclidean configuration space, and investigate directly the leading behavior of the singularity as two points coincide (e.g., $x_1 \to x_2$). A straightforward extension of the "tree expansion" bound for V, as developed in Refs. 5 and 6, leads to power counting formulas (3.5a)–(3.5d) for the power law and logarithmic dependence of the singularity. The proof renders explicit the following intuition: the behavior of $V(x_1, x_2, \ldots, x_n)$ when $|x_1 - x_2| \sim 0$ can be measured by examining the large H dependence of

$$\int d^d x_1 \exp - M^H |x_1 - x_2| \; V(x_1, x_2, \ldots, x_n) \quad (M > 1 \text{ fixed}).$$

Integrals such as this are easy to estimate by the method of Refs. 5 and 6. By the same token, the short distance behavior when a cluster of points $\{x_1, \ldots, x_m\} \subseteq \{x_1, \ldots, x_n\}$ coincide can be measured by connecting the points $\{x_1, \ldots, x_m\}$ with variable-strength exponential springs. This more general argument will be presented in a future paper.

Section II of this paper provides a thumbnail sketch of the tree expansion and bounds for a BPHZ renormalized Feynman amplitude. Section III sets up the main power counting formula (Theorem 3), which is then proved in Sec. IV by use of a straightforward extension of the bound described in Sec. II. The approach presented here is very different in flavor from the method of Ref. 4. With the tree expansion method, each Feynman amplitude V is decomposed into a (finite) sum of terms, and each term is bounded separately. Similarly, the power counting formula derived here applies to each term separately and, to determine the short distance behavior of V itself, all terms must be checked. The drawbacks of this piecemeal approach are compensated by the directness of the term-by-term estimates.

II. THE RENORMALIZED TREE EXPANSION

Let G be a connected Feynman graph in dimension $d > 2$, which may be ultraviolet divergent and whose lines $l \in L(G)$ denote propagators C_l of dimension δ_l and mass $m \geqslant 1$. Such propagators behave like $|x|^{-\delta_l}$ for small separations $|x|$ and admit a scale decomposition:

$$C_l = \sum_{k=0}^{\infty} C_l^{(k)}, \tag{2.1}$$

where the slice propagator $C_l^{(k)}$ and its derivatives have bounds

$$|\partial_x^j C_l^{(k)}(x)| < O(1) M^{(\delta_l + |j|)k} \exp - M^k |x|, \quad |j| = 0,1,\ldots, \quad k = 0,1,\ldots. \tag{2.2}$$

Throughout this paper $M > 1$ is fixed, and $O(1)$ is used to denote numbers independent of any choice of scale but which may depend on other things (e.g., M, G, $|j|$, or l). I suppose that the interaction vertices $v \in V(G)$ are all dimensionless (i.e., strictly renormalizable)

$$\delta_v \equiv d - \sum_{\lambda \in \Lambda(v)} \delta_\lambda = 0, \tag{2.3}$$

where δ_λ denotes the dimension of the half-line λ emerging from v (including $+1$ for each derivative). This assumption is for simplicity only: The optimal power counting is somewhat messier when super-renormalizable vertices (with $\delta_v > 0$) are present, but there is no difficulty.

The truncated amplitude associated to G, renormalized at zero external momentum according to the BPHZ prescription, is $V(G) = \int R(G)$, where the integrand $R(G)$ is given by the forest formula (see, e.g., Ref. 7, p. 392)

$$R(G) = \sum_{\substack{\text{proper divergent} \\ \text{forests } F}} \prod_{f \in F} (-\mathscr{T}_f) I(G). \tag{2.4}$$

In the α-parametric representation of $I(G)$, $\mathscr{T}_f$ denotes a Taylor operation of order δ_f, where $\delta_f \equiv d - \Sigma_{\lambda \in \Lambda(G_f)} \delta_\lambda$ is the superficial degree of divergence of the subgraph $G_f \subseteq G$.

 0022-2488/88/092112-05$02.50

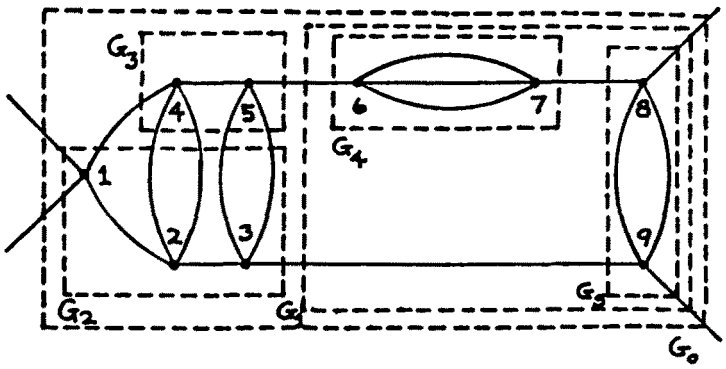
The above formula for $V(G)$, while useful, is not finite term-by-term and cannot be estimated directly. The renormalized tree expansion (Ref. 6, Sec. II) yields an alternative formula for $V(G)$, each term of which is finite:

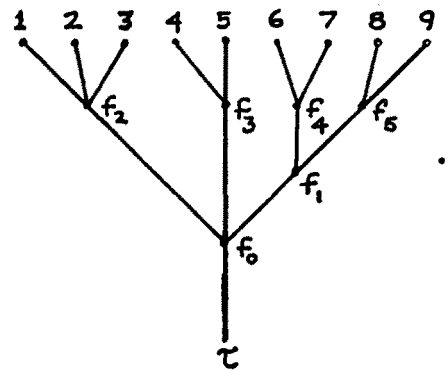
$$V(G) = \sum_{\substack{\text{connected} \\ \text{forests } \sigma}} \sum_{\substack{R/C \text{ labelings} \\ \vec{\rho}}} V(G,\sigma,\vec{\rho}) \,. \tag{2.5}$$

Here

$$\sigma = \{G_{f_0} \equiv G, G_{f_1}, \ldots, G_{f_n}\}$$

is a family of connected, not necessarily superficially divergent subgraphs which satisfies the forest property: If $i<j$ then either $G_{f_j} \subset G_{f_i}$ or $G_{f_i} \cap G_{f_j} = \phi$. Given σ, one can construct a rooted, planted tree τ by identifying the leaves of τ with the vertices of G, and the forks f_i of τ with the subgraphs G_{f_i}. Here is an example of a labeled graph G and the corresponding tree τ:





To each fork $f \in F(\tau)$ (equivalently, to each subgraph G_f) is assigned either an R (renormalization) or a C (counterterm): $\rho_f = R$ or $\rho_f = C$.

An allowed scale assignment $\vec{h} \in H(\tau,\vec{\rho})$ is defined to be an assignment $\vec{h} = (h_{f_0},\ldots,h_{f_n})$ of a scale h_f to each fork f of τ subject to the conditions

$$\begin{aligned} &h_{\pi(f)} + 1 \leqslant h_f < \infty, \quad \text{if } \rho_f = R \,, \\ &0 \leqslant h_f \leqslant h_{\pi(f)}, \quad \text{if } \rho_f = C \,. \end{aligned} \tag{2.6}$$

Here $\pi(f)$ denotes the fork which lies immediately beneath f, and $h_{\pi(f_0)} = -1$. Now

$$V(G,\sigma,\vec{\rho}) = \sum_{\vec{h} \in H(\tau,\vec{\rho})} V(G,\sigma,\vec{\rho},\vec{h}) \,. \tag{2.7}$$

The rules for calculating $V(G,\sigma,\vec{\rho},\vec{h})$, explained in detail in Ref. 6, Sec. II, are as follows: $V(G,\sigma,\vec{\rho},\vec{h})$ is a renormalized Feynman amplitude associated to G, where (i) each line l denotes either the "hard" propagator $C_l^{(h)}$ or the "soft" propagator $C_l^{(<h_l)}$, where h_l is the scale of the highest fork for which $l \in L(G_f)$; (ii) the possible hard/soft assignments are summed over, subject to the condition that each G_f be connected by hard lines; and (iii) each subgraph G_f is acted on by either the renormalization operator R if $\rho_f = R$ or the counterterm operator C if $\rho_f = C$.

For any bounded continuous n-point function $V(x_1,\ldots,x_n) \in (C(R^d)^n)$ let

$$\|V\| = \sup\left\{ \left| \int \left[\prod_{i=1}^{n} dx_i \, f_i(x_i) \right] V(x_1,\ldots,x_n) \right| : f_i \in C \right\}, \tag{2.8}$$

where

$$C \equiv \bigcup_{x \in R^d} \{ f \in C_0^\infty(B_1(x)) : \sup |\delta^m f(y)| \leqslant 1, \ |m| < d \} \,.$$

This norm is straightforward to deal with and $\tilde{S}_n \equiv$ completion $(C(R^d)^n)$ is contained in $S'((R^d)^n)$, the tempered distributions.

Theorem 1 (Gallavotti–Nicoló): $V(G,\sigma,\vec{\rho}) \in \tilde{S}_n$, where n = number of external vertices of G.

Proof: This result follows directly from the basic bound [Ref. 6, Eq. (2.86)]:

$$\|V(G,\sigma,\vec{\rho},\vec{h})\| \leqslant O(1) \prod_f M^{\delta(G_f)(h_f - h_{\pi(f)})} \,, \tag{2.9}$$

where the *renormalized* degrees of divergence are such that $\delta(G_f) \leqslant -1$, if $\rho_f = R$ and $\delta(G_f) \geqslant 0$, if $\rho_f = C$. In view of (2.6), we see that the sum over each h_f is a convergent geometric series. □

III. THE POWER COUNTING FORMULA

It is natural to say that the n-point kernel $V \in \tilde{S}_n$ behaves like $|x_1 - x_2|^{-p} \ln^q |x_1 - x_2|$ near $x_1 = x_2$ if

$$\lim_{x_1 \to x_2} |x_1 - x_2|^p \ln^{-q} |x_1 - x_2| V(x_1,x_2,\ldots,x_n) \tag{3.1}$$

exists and is nonzero in $\tilde{S}_{n-1}$. In general, since (3.1) is a direction dependent limit, we adopt a definition which averages over the possible directions in (3.1).

Definition: The n-point kernel $V \in \tilde{S}_n$ has a singularity of type (p,q) (where $p,q \geqslant 0$) at $x_1 = x_2$ if

$$\int_{S^{d-1}} d\eta \lim_{\rho \to 0} \rho^p \ln^{-q} \rho \; V(\rho\eta + x_2, x_2, \ldots, x_n) \tag{3.2}$$

exists and is nonzero in $\tilde{S}_{n-1}$. The term p is called the power index and q the logarithmic index of the singularity.

Thus, if V behaves like $|x_1 - x_2|^{-p} \ln|x_1 - x_2|^q$ for $|x_1 - x_2|$ small, the singularity has type (p,q). We shall calculate p and q by use of the following alternative characterization.

Lemma 2: The singularity at $x_1 = x_2$ of $V \in \tilde{S}_n$ has type (p,q) if and only if

$$\lim_{H\to\infty} H^{-q} M^{(d+r-p)H} V_{H,r}$$

exists and is nonzero in $\tilde{S}_{n-1}$, for any $r > p - d$, where

$$V_{H,r}(x_2,\ldots,x_n) \equiv \int d^d x_1 |x_1 - x_2|^r \times [\exp - M^H |x_1 - x_2|]\, V(x_1,x_2,\ldots,x_n)\,. \tag{3.3}$$

Proof: For any set of test functions $f_i \in C_0^\infty(R^d)$, $i = 2,\ldots,n$ and $r > p - d$,

$$\begin{aligned}
&\int \left[\prod_{i=2}^n dx_i f_i(x_i)\right] \lim_{H\to\infty} H^{-q} M^{(d+r-p)H} V_{H,r}(x_2,\ldots,x_n) \\
&\quad \equiv \lim_{H\to\infty} H^{-q} M^{(d+r-p)H} \int \left[\prod_{i=2}^n dx_i f_i(x_i)\right] V_{H,r}(x_2,\ldots,x_n) \\
&\quad = \lim_{H\to\infty} H^{-q} M^{(d+r-p)H} \int_0^\infty d\rho\, \rho^{d-1+r} e^{-M^H\rho} \int_{S^{d-1}} d\eta \int \left[\prod_{i=2}^n dx_i f_i(x_i)\right] V(\rho\eta + x_2, x_2,\ldots,x_n) \\
&\quad = \lim_{H\to\infty} \int_0^\infty d\tilde{\rho}\, \tilde{\rho}^{(d-1+r-p)} e^{-\tilde{\rho}} \int_{S^{d-1}} d\eta \left[\frac{H}{\ln \tilde{\rho} M^{-H}}\right]^{-q} [\ln \tilde{\rho} M^{-H}]^{-q} [\tilde{\rho} M^{-H}]^p \int \left[\prod_{i=2}^n dx_i f_i(x_i)\right] \\
&\qquad \times V(\tilde{\rho} M^{-H}\eta + x_2, x_2,\ldots,x_n) \\
&\quad = [-\ln M]^q \left[\int_0^\infty d\tilde{\rho}\, \tilde{\rho}^{(d-1+r-p)} e^{-\tilde{\rho}}\right] \int_{S^{d-1}} d\eta \lim_{\zeta\to 0} \left\{\int \left[\prod_{i=2}^n f_i(x_i) dx_i\right] \zeta^p \ln \zeta^{-q} V(\zeta\eta + x_2, x_2,\ldots,x_n)\right\}.
\end{aligned}$$

The last equality follows from the dominated convergence theorem. □

Now consider a kernel $V(G,\sigma,\vec{\rho})$ which contributes to the Feynman amplitude (2.5). Let $f_0 < f_1 < \cdots < f_I$ be the maximal chain of forks for which $G_i \equiv G_{f_i}$ contains both x_1 and x_2. For each i, $0 \leqslant i \leqslant I$, let

$$\delta_i \equiv \delta_{f_i} = d - \sum_{\lambda \in \Lambda(G_{f_i})} \delta_\lambda$$

and

$$\kappa_i \equiv \#\{f\colon f > f_i,\ f \not\geqslant f_{i+1} \text{ and } \rho_f = C_0\} \tag{3.4}$$

denote the number of dimension zero counterterms contained in the subgraph G_i but not in G_{i+1}.

Theorem 3: The singularity of $V(G,\sigma,\vec{\rho})$ at $x_1 = x_2$ has type (p,q) with

$$p = \max\{0, \delta_0 + d, \delta_1 + d, \ldots, \delta_I + d\}, \tag{3.5a}$$

$$q = \sum_{i=J}^{I} \kappa_i + \max\{N - 1, 0\}, \tag{3.5b}$$

where $N \equiv \#\{i\colon \delta_i = p\}$ and $J \equiv \min\{i\colon \delta_i = p\}$.

Remarks: (i) Since we prove only upper bounds on all quantities, Theorem 3 gives in fact only an upper bound on (p,q). There is some possibility, for example when fermions are present, that cancellations can occur that reduce the indices p and q. Such cancellations cannot easily be seen using the tree expansion method.

(ii) In the general case when interaction vertices with degree $\delta_v > 0$ occur, the formulas for p and q are tiresome to write down. For graphs without C subgraphs, however,

$$p = \max\{0, D_0 + d, D_1 + d, \ldots, D_I + d\}, \tag{3.5c}$$

$$q = \max\{N - 1, 0\}, \tag{3.5d}$$

where

$$D_f \equiv \sum_{l \in L(G_f)} d_l - d(v(G_f) - 1)$$

and $N \equiv \#\{i\colon D_i = p\}$. Note that

$$D_f = \delta_f - \sum_{v \in V(G_f)} \delta_v \leqslant \delta_f\,.$$

(iii) To determine the leading singularity of the full Feynman amplitude $V(G)$ all contributions $V(G,\sigma,\vec{\rho})$ must be checked, and the maximum p and q taken.

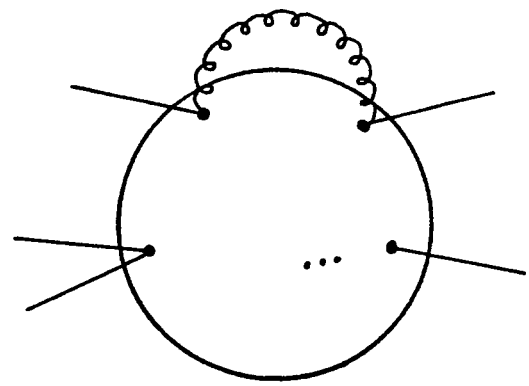
IV. PROOF OF THEOREM 3

By the lemma, it is sufficient to determine the asymptotic behavior for large H of

$$\|V_{H,r}\| = \left|\left| \int d^d x_1 |x_1 - x_2|^r \times [\exp - M^H |x_1 - x_2|]\, V(x_1,x_2,\ldots,x_n) \right|\right|$$

for any sufficiently large value of r [where V denotes the kernel of $V(G,\sigma,\vec{\rho})$]. Note that $V_{H,r}$ is associated to a graph $\tilde{G}$ obtained from G by the addition of one line:

$$|x_1 - x_2|^r \exp - M^H |x_1 - x_2|$$



,

where the extra line is similar in form to the estimates (2.2) for the lines of G.

Choose $r > \max\{\delta_i\colon 0 \leqslant i \leqslant I\}$. Then there are enough extra powers of $|x_1 - x_2|$ in the integrand of $\tilde{G}$ to annihilate the delta function dependence that comes if a localization operator acts on a subgraph G_i. Thus $V_{H,r}$ vanishes if one of the forks f_i has a C label. For the same reason, the renormalization operator acts trivially on the forks f_i.

The power counting for $\|V_{H,r}\|$ differs from (2.9) by a factor coming from the extra line in the subgraph G_I, and by the omission of renormalization factors for the forks f_i. When $h_I \geqslant H$, the factor $\exp - M^H|x_1 - x_2|$ is bounded by 1, while the factor $|x_1 - x_2|^r$ can be seen by a scaling argument to supply an additional factor M^{-rh_I}. When $h_I < H$, use of the bound

$$|x_1 - x_2|^r \exp(-M^H|x_1 - x_2|) < O(1)M^{-rH}\exp(-1/2M^H|x_1 - x_2|)$$

yields the factor M^{-rH}, and in addition the factor $M^{-d(H-h_I)}$ obtained by replacing one decay factor M^{-dh_I} by M^{-dH}. The basic bound is

$$\|V_{H,r}\| \leqslant O(1) \sum_{h_0,\dots,h_I} \left\{ \left[\prod_{i=0}^{I} M^{(h_i - h_{i-1})\delta_i} \right] \times [M^{-r(h_I,H)_> - d(H-h_I)_+}] \times \left[\sum_{h_f} \prod_f M^{(h_f - h_{\pi(f)})\delta(G_f)} \right] \right\}, \tag{4.1}$$

where

$$(x)_+ \equiv \begin{cases} x, & \text{if } x \geqslant 0, \\ 0, & \text{if } x < 0, \end{cases} \qquad (x,y)_> \equiv \begin{cases} x, & \text{if } x \geqslant y, \\ y, & \text{if } y \geqslant x. \end{cases}$$

The last factor of (4.1) is the contribution of those forks not in the chain $\{f_0,\dots,f_I\}$.

Except at C forks with $\delta_f = 0$, the scale sums for f are each bounded by $O(1)$ because of the exponential dependence on $(h_f - h_{\pi(f)})$. Marginal C forks introduce polynomial scale dependence:

$$\sum_{h_f=0}^{h_{\pi(f)}} (h_f + 1)^{\kappa} \leqslant O(1)(h_{\pi(f)} + 1)^{\kappa+1}, \quad \text{for } \kappa = 0,1,\dots . \tag{4.2}$$

Having done all scale sums except those for $f_0,\dots,f_I$, one finds

$$\|V_{H,r}\| \leqslant O(1) \sum_{h_0,\dots,h_I} \left[\prod_{i=0}^{I} (h_i + 1)^{\kappa_i} M^{(h_i - h_{i-1})\delta_i} \right] \times [M^{-d(H-h_I)_+ - r(h_I,H)_>}]. \tag{4.3}$$

For each j, $0 \leqslant j \leqslant I$, let

$$U_j(h_{j-1}) \equiv \sum_{h_j, h_{j+1},\dots,h_I} \left[\prod_{i=j}^{I} (h_i + 1)^{\kappa_i} M^{(h_i - h_{i-1})\delta_i} \right] \times [M^{-d(H-h_I)_+ - r(h_I,H)_>}]. \tag{4.4}$$

Then with $U_{I+1}(h_I) \equiv M^{-d(H-h_I)_+ - r(h_I,H)_>}$,

$$U_j(h_{j-1}) = \sum_{h_j = h_{j-1}+1}^{\infty} U_{j+1}(h_j)(h_j + 1)^{\kappa_j} M^{(h_j - h_{j-1})\delta_j}, \tag{4.5}$$

for $0 \leqslant j \leqslant I$.

Lemma 4: Let $\rho_{I+1} = -d$, $\alpha_{I+1} = \beta_{I+1} = \gamma_{I+1} = 0$, and for $0 \leqslant j \leqslant I$ let

$$\rho_j = \max\{-d, \delta_j, \delta_{j+1},\dots,\delta_I\},$$

$$\alpha_j = \sum_{i=J_j}^{I} \kappa_i, \quad \beta_j = \sum_{i=j}^{J_j - 1} \kappa_i, \quad \gamma_j = (N_{j-1} - 1)_+,$$

where $J_j = \min\{i\colon i \geqslant j \text{ and } \delta_i = \rho_j\}$ and $N_j = \#\{i\colon i \geqslant j \text{ and } \delta_i = \rho_j\}$.

(a) If $h_{j-1} < H$, then

$$U_j(h_{j-1}) \leqslant O(1) H^{\alpha_j}(h_{j-1} + 1)^{\beta_j}(H - h_{j-1})^{\gamma_j} \times M^{\rho_j(H - h_{j-1}) - rH}. \tag{4.6a}$$

(b) If $h_{j-1} \geqslant H$, then

$$U_j(h_{j-1}) \leqslant O(1)(h_{j-1} + 1)^{\alpha_j + \beta_j} M^{-rh_{j-1}}. \tag{4.6b}$$

Theorem 3 follows immediately from Lemma 4 and Lemma 2:

$$\|V_{H,r}\| \leqslant U_0(0) \leqslant O(1) H^{\alpha_0 + \gamma_0} M^{(\rho_0 - r)H},$$

which yields $p = \rho_0 + d$ and $q = \alpha_0 + \gamma_0$.

Proof of Lemma 4: Clearly (4.6) is true for $j = I + 1$. Inductively, (4.6) is shown for U_j using (4.5) under the assumption that (4.6) holds for U_{j+1}. Suppose $h_{j-1} \geqslant H$. Then

$$U_j(h_{j-1}) \leqslant O(1) \sum_{h_j = h_{j-1}+1}^{\infty} (h_j + 1)^{\alpha_{j+1} + \beta_{j+1} + \kappa_j} \times M^{-rh_j + (h_j - h_{j-1})\delta_j} \leqslant O(1)(h_{j-1} + 1)^{\alpha_j + \beta_j} M^{-rh_{j-1}},$$

since $r > \delta_j$ for all j. Now suppose $h_{j-1} < H$. Then

$$U_j(h_{j-1}) \leqslant O(1) \sum_{h_j = h_{j-1}+1}^{H-1} H^{\alpha_{j+1}}(h_j + 1)^{\beta_{j+1} + \kappa_j}(H - h_j)^{\gamma_{j+1}} M^{(h_j - h_{j-1})\delta_j + \rho_{j+1}(H - h_j) - rH} + O(1) \sum_{h_j = H}^{\infty} (h_j + 1)^{\alpha_{j+1} + \beta_{j+1} + \kappa_j} M^{-rh_j + (h_j - h_{j-1})\delta_j}. \tag{4.7}$$

By use of the estimates

$$\sum_{i=j+1}^{k-1} (i + 1)^{\alpha}(k - i)^{\beta} M^{\gamma i} \leqslant \begin{cases} O(1)(j + 1)^{\alpha}(k - j)^{\beta} M^{\gamma j}, & \text{if } \gamma < 0, \\ O(1) k^{\alpha}(k - j)^{\beta + 1}, & \text{if } \gamma = 0, \\ O(1) k^{\alpha} M^{\gamma k}, & \text{if } \gamma > 0, \end{cases} \tag{4.8}$$

which hold with $O(1)$ numbers which depend on α, β, and γ but not on j and k, the first term of (4.7) is found to be bounded by

$$O(1)H^{\alpha_{j+1}}(h_{j-1}+1)^{\beta_{j+1}+\kappa_j}(H-h_{j-1})^{\gamma_{j+1}}M^{\rho_{j+1}(H-h_{j-1})-rH}, \quad \text{if } \delta_j<\rho_{j+1}, \tag{4.9a}$$

$$O(1)H^{\alpha_{j+1}+\beta_{j+1}+\kappa_j}(H-h_{j-1})^{\gamma_{j+1}+1}M^{\rho_{j+1}(H-h_{j-1})-rH}, \quad \text{if } \delta_j=\rho_{j+1}, \tag{4.9b}$$

$$O(1)H^{\alpha_{j+1}+\beta_{j+1}+\kappa_j}M^{\delta_j(H-h_{j-1})-rH}, \quad \text{if } \delta_j>\rho_{j+1}\,. \tag{4.9c}$$

The second term is bounded by $O(1)H^{\alpha_{j+1}+\beta_{j+1}+\kappa_j}M^{\delta_j(H-h_{j-1})-rH}$, which is always less than $O(1)$ times the first term. Since the quantities α, β, γ, and ρ satisfy the recursion relations

$$\alpha_j=\alpha_{j+1},\quad \beta_j=\beta_{j+1}+\kappa_j,\quad \gamma_j=\gamma_{j+1},\quad \rho_j=\rho_{j+1},\quad \text{if } \delta_j<\rho_{j+1},$$

$$\alpha_j=\alpha_{j+1}+\beta_{j+1}+\kappa_j,\quad \beta_j=0,\quad \gamma_j=\gamma_{j+1}+1,\quad \rho_j=\rho_{j+1},\quad \text{if } \delta_j=\rho_{j+1},$$

$$\alpha_j=\alpha_{j+1}+\beta_{j+1}+\kappa_j,\quad \beta_j=0,\quad \gamma_j=0,\quad \rho_j=\delta_j,\quad \text{if } \delta_j>\rho_{j+1},$$

the Lemma follows from (4.9). □

ACKNOWLEDGMENTS

The author is grateful to Professor Joel Feldman for his helpful comments during the preparation of this paper.

This work was supported by the Natural Sciences and Engineering Research Council.

[1]S. Weinberg, Phys. Rev. **118**, 838 (1960).

[2]J. P. Fink, J. Math. Phys. **9**, 1389 (1968).

[3]M. C. Bergère and Y. M. P. Lam, Commun. Math. Phys. **39**, 1 (1974); "Asymptotic expansion of Feynman amplitudes, Part II. The divergent case," Preprint, Freie Universität Berlin HEP May 74/9 (unpublished).

[4]M. C. Bergère, C. de Calan, and A. P. C. Malbouisson, Commun. Math. Phys. **62**, 137 (1978).

[5]G. Gallavotti and F. Nicolò, Commun. Math. Phys. **100**, 545 (1985); **101**, 247 (1985).

[6]J. Feldman, T. R. Hurd, L. Rosen, and J. D. Wright, *QED: A Proof of Renormalizability*, to appear in Springer *Lecture Notes in Physics*.

[7]C. Itzykson and J-B. Zuber, *Quantum Field Theory* (McGraw-Hill, New York, 1980).

Massive fields at null infinity

J. Winicour
Department of Physics and Astronomy, University of Pittsburgh, Pittsburgh, Pennsylvania 15260

(Received 30 March 1988; accepted for publication 11 May 1988)

The question of whether massive fields are consistent with the asymptotics of massless fields at null infinity is addressed. As a first step, it is shown at the level of linearized theory that massive fields have $O(1/R^{\infty})$ asymptotic behavior.

I. INTRODUCTION

A conformal space-time boundary $\mathscr{I}$ at null infinity has proved to be an exceedingly useful construction for describing the boundary conditions and asymptotic properties of the radiation fields associated with massless wave equations.[1] Here, we address the issue of whether this construction remains useful or even legitimate in the presence of fields with nonzero rest mass.

The term $\mathscr{I}$ was originally introduced in the context of general relativity to distinguish between the limits of going to infinity in timelike, spacelike, or null directions.[2] The asymptotic properties at $\mathscr{I}$ were found to be drastically different from those at spatial infinity or temporal infinity, as one might expect from physical considerations. For an oscillating, compact matter source, there should be no asymptotic falloff of the gravitational field strength in timelike directions, there should be a $1/r$ radiative falloff in null directions, and there should be a stronger falloff in spacelike directions. This is formalized by relating the physical manifold M with metric $g_{\mu\nu}$ conformally to a manifold with boundary $\hat{M}$ with metric $\hat{g}_{\mu\nu} = \Omega^2 g_{\mu\nu}$. Future (past) null infinity $\mathscr{I}^+$ ($\mathscr{I}^-$) is described by points $\Omega = 0$ and has topology $S^2 \times R^1$, with the S^2 describing the angular dependence and the R^1 the retarded (advanced) time dependence of the radiation. For Minkowski space-time, this construction of $\mathscr{I}^+$ can be accomplished by setting $\Omega = 1/r$. Then in the chart $\hat{x}^{\alpha} = (u,\hat{r},\theta,\phi)$, with $u = t - r$ and $\hat{r} = 1/r$, the metric $\hat{g}_{\mu\nu}$ is given by

$$d\hat{s}^2 = \hat{r}^2\,du^2 - 2\,du\,d\hat{r} - d\theta^2 - \sin^2\theta\,d\phi^2\,, \qquad (1.1)$$

which is smooth and nonsingular at $\mathscr{I}^+$, where $\hat{r} = 0$. In terms of this formalism, smoothness or differentiability of fields at $\mathscr{I}^+$ corresponds to the existence of a $1/r$ series expansion. Spatial infinity has a natural one-point compactification analogous to the situation for the Poisson equation and, as a result, has different smoothness properties than $\mathscr{I}^+$ (Refs. 3 and 4).

A C^{∞} order of differentiability at $\mathscr{I}$ is not essential in general relativity, but the assumption of some degree of smoothness is necessary in order to obtain a consistent treatment of gravitational radiation. The justification for this assumption has been based entirely on the properties of massless fields. In the case of special relativity, the conformal invariance of the zero rest mass field equations leads directly to C^{∞} behavior at $\mathscr{I}$, provided the sources and initial data are C^{∞} with suitable global properties. This smoothness embodies the generic $1/r$ falloff of radiation fields. Einstein's equation, while not conformally invariant, is conformal to a regular symmetric hyperbolic system of equations at $\mathscr{I}$. This again leads to a C^{∞} structure at $\mathscr{I}$, if not in general, at least under strong enough boundary conditions.[5] Furthermore, this result is preserved when gravity is coupled to massless fields that are smooth at $\mathscr{I}$, essentially because their energy momentum tensor falls off as $1/r^2$.

From a particle viewpoint, $\mathscr{I}$ is a natural structure in the zero rest mass case where it represents the end points of particle trajectories. But in the massive case, particles move along timelike geodesics and do not get to $\mathscr{I}$, except in unphysical situations, such as the hyperbolic space-time trajectories resulting from uniform acceleration. The particle picture implies that in the zero wavelength limit massive fields do not get to $\mathscr{I}$ and it suggests that, under reasonable conditions, this result might extend to more general massive fields.

The asymptotic behavior of special relativistic massive fields at timelike infinity agrees with the particle interpretation.[6,7] However, this is far from obvious at null infinity. In Minkowski space-time, consider the masive Klein–Gordon equation with source ρ,

$$\Box\Phi + m^2\Phi = \rho\,. \qquad (1.2)$$

Taking ρ to be a four-dimensional point source at p, with coordinates x^{α}, the physically appropriate solution at the point P in Minkowski space-time, with coordinates X^{α}, is given by the retarded Green's function

$$G(X^{\alpha};x^{\alpha}) = (1/2\pi)\theta(X^0 - x^0) \times\{\delta(\lambda) - \tfrac{1}{2}m\lambda^{-1/2}\theta(\lambda)\,J_1(m\lambda^{1/2})\}\,, \qquad (1.3)$$

where $\lambda = (X^{\infty} - x^{\alpha})(X_{\alpha} - x_{\alpha})$. Taking $x^{\alpha} = 0$, in the neighborhood of $\mathscr{I}^+$ to the strict future of P, this has the asymptotic behavior

$$G \sim m^{1/2}(8\pi^2 RU)^{-3/4}\cos\{mU^{1/2}(U + 2R)^{1/2} + \pi/4\}\,, \qquad (1.4)$$

in spherical null coordinates. Surprisingly this is a weaker falloff than the retarded Green's function for a massless field that goes as $1/R$ on the future null cone of p and that vanishes precisely in the strict future of P. When coupled to gravity, the energy momentum tensor associated with (1.4) would be inconsistent with the asymptotic flatness required to construct $\mathscr{I}$.

Then, do these massive fields play the role of bullies that spoil $\mathscr{I}$ for the massless ones? In this regard, the above example is inconclusive because of the infinite self-energy of the field of a source localized to a single space-time point. For the field (1.3), in addition to the problem in the neigh-

 0022-2488/88/092117-05$02.50

borhood of the point source, there is an infinite radiative power at $\mathscr{I}^+$. To pose the question properly we should smear out the point source. In fact, if the four-dimensional point source is extended to a straight timelike world line then it is well known that the resulting static field has the Yukawa form $e^{-mr}/r = \hat{r}\exp(-m/\hat{r})$. This is in fact a C^∞ field at $\mathscr{I}$, whose $\hat{r}$ derivatives all vanish. This example indicates that the rapidly varying asymptotic behavior (1.4) of the Green's function can lead to phase cancellations that eliminate the presence of the scalar field at $\mathscr{I}$.

Then, does this result for the Yukawa potential extend to the fields of well behaved time-dependent sources? It is difficult to attack this question in a direct manner because the massive Klein–Gordon equation is itself singular at $\mathscr{I}$. Setting $\Phi = \Omega\hat{\Phi}$, (1.2) takes the form

$$\{\hat{\Box} - (\hat{R}/6) + (m^2/\Omega^2)\}\hat{\Phi} = \rho/\Omega^3, \tag{1.5}$$

in terms of the wave operator $\hat{\Box}$ and curvature scalar $\hat{R}$ for $\hat{g}_{\mu\nu}$. The mass term acts like a singular potential at $\mathscr{I}$ (where $\Omega = 0$). When $m = 0$, the equation is regular at $\mathscr{I}$ (for smooth ρ of compact support) and standard arguments can be used to establish smoothness of the solutions at $\mathscr{I}$. For $m \neq 0$, no way is known to cast (1.5) into a regular form. However, if $\hat{\Phi}$ were *assumed* to be C^∞ at $\mathscr{I}$, then it would be easy to show that it must vanish there to all orders, in the same sense as the Yukawa potential. Setting $\hat{\Phi} = \Sigma a_n(u,\theta,\phi)\hat{r}^n$ and $\Omega = \hat{r}$, for which the transformed scalar curvature $\hat{R} = 0$, and using the metric (1.1), the wave equation gives

$$m^2 a_n - 2(n-1)\partial_u a_n - \{(n-1)(n-2) - L^2\}a_{n-2} = 0,$$

where L^2 is the angular momentum operator [with eigenvalues $l(l+1)$]. If the leading term in the expansion is finite (non-negative n), then this forces each a_n to vanish. Thus massive fields cannot get to $\mathscr{I}$ smoothly. But it remains to resolve whether they exhibit nonsmooth asymptotic behavior, such as the fractional powers of $1/r$ in (1.4).

In order to put this issue in a tractable and clear context we shall consider massive Klein–Gordon fields under the following assumptions.

(A1) The background geometry is Minkowski space-time.

(A2) The source ρ is C^∞ and of compact space-time support.

(A3) Initial data for the Klein–Gordon field are C^∞ and of compact support on some Cauchy hypersurface.

In the absence of any general approach, due to the singularity in (1.5), (A1) allows construction of explicit solutions, for arbitrary source and data, using the Green's function. Some weaker version of (A2) would be desirable since the Yukawa solution has a source of compact spatial support, but with infinite time extension. However, it is technically complicated to formulate and to deal with the conditions appropriate to a source that is asymptotically nonradiative in the infinite past and future. In (A3), we allow for the possibility of incoming scalar fields that are initially consistent with the smoothness of $\mathscr{I}$. This assumption rules out data that would blatantly force the scalar field to $\mathscr{I}$. While it would be preferable to pose this with a more subtle asymptotic falloff, (A3) suffices to investigate whether nonsmooth asymptotic behavior arises from the evolution of homogeneous fields.

The linearity of the Klein–Gordon equation lets us treat the homogeneous and inhomogeneous cases separately. In Sec. II, we show that the retarded Klein–Gordon fields of a source satisfying (A1) and (A2) do not get to $\mathscr{I}$. In Sec. III, we extend this result to source-free fields that evolve from data satisfying (A3). The key underlying mechanism for these results are certain iterative formulas for integrals of Bessel functions, expressed in (2.14) and (3.8).

Although only spinless fields are explicitly treated here, the results extend directly to massive special relativistic fields with spin, since in that case each field component satisfies the Klein–Gordon equation.

II. FIELD OF A COMPACT SOURCE

We continue to represent a field point P by coordinates X^α and a source point p by coordinates x^α. The retarded solution of the Klein–Gordon equation (1.2) due to a source ρ is

$$\Phi(X) = \int \rho(x)G(X;x)dv\,dt, \tag{2.1}$$

where $G(X;x)$ is given by (1.3) and where $dv = dx\,dy\,dz$. The solution vanishes for vanishing ρ so that it satisfies the boundary condition that there be no incoming fields in the asymptotic past. In terms of null coordinates $U = T - R$ and $R = [X^2 + Y^2 + Z^2)^{1/2}$, we are interested in the asymptotic behavior of Φ as $R \to \infty$ holding U constant, i.e., at $\mathscr{I}^+$. We will establish the following theorem.

Source Theorem: Under assumptions (A1) and (A2), the field (2.1) satifies $R^n\Phi \to 0$ at null infinity, for all n.

To begin, note that the Bessel function relationships[8] $J_1(\zeta) = -\partial_\zeta J_0(\zeta)$ and $J_0(0) = 1$ allow (1.3) to be reexpressed in the form

$$G(X;x) = (1/2\pi)\theta(T-t)\partial_\lambda\{\theta(\lambda)J_0(m\lambda^{1/2})\}. \tag{2.2}$$

Since we will eventually take the limit $T \to \infty$ and $R \to \infty$ (holding $U = T - R$ constant), we consider only those points P that lie outside the support of ρ and for which $T - t > 0$ when p lies inside the support of ρ. This is permissible according to assumption (A2). For these P, $\theta(T-t) = 1$ throughout the support of ρ, so that (2.1) and (2.2) combine to give

$$2\pi\Phi(X) = \int \rho(x)\partial_\lambda(\theta(\lambda)J_0(m\lambda^{1/2}))dv\,dt, \tag{2.3}$$

where the integration is still over all x^α. We now integrate (2.1) by parts in the sense of distributions. For convenience we choose coordinates X^α such that P lies along the positive Z axis with $X^\alpha = (T,0,0,Z)$, $R = Z$, and $U = T - Z$. Then setting $\epsilon^2 = 1/(2Z)$, we have $\lambda = h/\epsilon^2$, where

$$h = U - t + z + \epsilon^2\{(U-t)^2 - r^2\}. \tag{2.4}$$

Also, for any distribution $K(\lambda)$, $\partial_t K(\lambda) = (\partial_t\lambda)\partial_\lambda K(\lambda) = (h_t/\epsilon^2)\partial_\lambda K(\lambda)$, where

$$h_t = \partial_t h = -1 - 2\epsilon^2(U-t). \tag{2.5}$$

Here $h_t \to -1$ as $P \to \mathscr{I}^+$ and h_t is bounded away from zero for t in the support of ρ so that $1/h_t$ is smooth in the support of the integrand in (2.3). This allows integration by parts, with respect to t, to recast (2.3) into the form

$$2\pi\Phi = \epsilon^2 \int (D_t\rho)\theta(\lambda)J_0(m\lambda^{1/2})dv\,dt, \tag{2.6}$$

where, in an operator sense,

$$D_t = -(\partial_t)(1/h_t),$$

and where the space-time compactness of ρ has been used to drop the boundary terms. From (2.6), it follows that $\Phi = O(1/R)$ in the $\mathscr{I}^+$ limit.

Next we establish an iterative scheme based upon the Bessel function relationships[8]

$$J_n(0) = 0, \tag{2.7}$$

$$J_{n-1}(\zeta) = \zeta^{-n}\partial_\zeta(\zeta^n J_n(\zeta)), \tag{2.8}$$

valid for $n \geqslant 1$. For our purposes, we use (2.8) in the form

$$\begin{aligned} J_{n-1}(m\lambda^{1/2}) &= (2/m)\lambda^{-(n-1)/2}\partial_\lambda(\lambda^{n/2}J_n(m\lambda^{1/2})) \\ &= (2\epsilon^2/mh_t)\lambda^{-(n-1)/2}\partial_t(\lambda^{n/2}J_n(m\lambda^{1/2})). \end{aligned} \tag{2.9}$$

For $n = 1$, (2.9) gives

$$\int (D_t\rho)\theta(\lambda)J_0(m\lambda^{1/2})dv\,dt = \frac{2\epsilon^2}{m}\int \frac{D_t\rho}{h_t}\theta(\lambda)\partial_t(\lambda^{1/2}J_1(m\lambda^{1/2}))dv\,dt. \tag{2.10}$$

Integration of (2.10) by parts, with respect to t, gives rise to a boundary term that vanishes due to the compactness of ρ, a $\delta(\lambda)$ term that vanishes because of (2.7), and the surviving term

$$\int (D_t\rho)\theta(\lambda)J_0(m\lambda^{1/2})dv\,dt = \frac{2\epsilon^2}{m}\int (D_t^2\rho)\theta(\lambda)\lambda^{1/2}J_1(m\lambda^{1/2})dv\,dt. \tag{2.11}$$

Turning the case $n > 1$, (2.9) gives

$$\int (D_t^n\rho)\theta(\lambda)\lambda^{(n-1)/2}J_{n-1}(m\lambda^{1/2})dv\,dt = \frac{2\epsilon^2}{m}\int (D_t^n\rho)\theta(\lambda)\frac{1}{h_t}\partial_t(\lambda^{n/2}J_n(m\lambda^{1/2}))dv\,dt. \tag{2.12}$$

By the same argument that led from (2.10) to (2.11), (2.12) leads to

$$\int (D_t^n\rho)\theta(\lambda)\lambda^{(n-1)/2}J_{n-1}(m\lambda^{1/2})dv\,dt = \frac{2\epsilon^2}{m}\int (D_t^{n+1}\rho)\theta(\lambda)\lambda^{n/2}J_n(m\lambda^{1/2})dv\,dt. \tag{2.13}$$

Now, (2.11) and (2.12) iterate to give

$$\int (D_t\rho)\theta(\lambda)J_0(m\lambda^{1/2})dv\,dt = \left(\frac{2\epsilon^2}{m}\right)^n \int (D_t^{n+1}\rho)\theta(\lambda)\lambda^{n/2}J_n(m\lambda^{1/2})dv\,dt, \tag{2.14}$$

for arbitrary $n \geqslant 0$.

Finally, by substituting (2.14) into (2.6), we obtain

$$2\pi\Phi = \left(\frac{2}{m}\right)^n \epsilon^{2(n+1)} \int (D_t^{n+1}\rho) \times \theta(\lambda)\lambda^{n/2}J_n(m\lambda^{1/2})dv\,dt,$$

or, recalling $\lambda = h/\epsilon^2$,

$$2\pi\Phi = \left(\frac{2}{m}\right)^n \epsilon^{n+2} \int (D_t^{n+1}\rho)\theta(\lambda) \times h^{n/2}J_n(m\lambda^{1/2})dv\,dt. \tag{2.15}$$

The integrand in (2.15) has a finite limit as $\epsilon \to 0$, holding U constant so that $\Phi = O(\epsilon^n)$, for arbitrary n. But, by our choice of coordinates, $\epsilon^2 = 2/R$, so that this establishes the above Source Theorem.

III. HOMOGENEOUS FIELDS

Consider now the Cauchy hypersurface $t = 0$. The unconstrained Cauchy data for a Klein–Gordon field consists of $F(x^i) = \Phi(0,x^i)$ and $\dot{F}(x^i) = \partial_t\Phi(0,x^i)$. By Green's theorem, evolution of this data into the source-free region $T > 0$ leads to a field given by the integral at $t = 0$

$$\Phi(X) = \int (\dot{F}(x)G(X;x) - F(x)\partial_t G(X;x))dv, \tag{3.1}$$

where $G(X;x)$ is given by (1.3). We now establish the following theorem.

Cauchy Data Theorem: Under assumptions (A1) and (A3), the field (3.1) satisfies $R^n\Phi \to 0$ at null infinity, for all n.

We begin the proof by substituting the form of the Green's function (2.2), with $\theta(T-t) = 1$, into (3.1), which gives

$$2\pi\Phi(X) = \int \{\dot{F}(x)\partial_\lambda(\theta(\lambda)J_0(m\lambda^{1/2})) - F(x)\partial_t\partial_\lambda(\theta(\lambda)J_0(m\lambda^{1/2}))\}dv. \tag{3.2}$$

As in Sec. II, we choose coordinates such that $X^\alpha = (T,0,0,Z)$, with (2.4) and (2.5) and their associated notation unchanged. In addition, it is useful to introduce

$$h_z = \partial_z h = 1 - 2\epsilon^2 z \tag{3.3}$$

and the operator

$$D_z = -(\partial_z)(h_z). \tag{3.4}$$

For any distribution $K(\lambda)$, $\partial_\lambda K(\lambda) = (\epsilon^2/h_z)\partial_z K(\lambda)$ and $\partial_t K(\lambda) = (h_t/h_z)\partial_z K(\lambda)$. Using these relationships, we integrate the $F(x)$ term in (3.2) by parts twice, in the z direction, to obtain

$$2\pi\Phi(X) = \int f(x)\partial_\lambda(\theta(\lambda)J_0(m\lambda^{1/2}))dv, \tag{3.5}$$

where the boundary terms have been set to zero in accord with (A3) and where

$$f(x) = \dot{F}(x) - D_z(h_t F(x)). \tag{3.6}$$

Here $f(x)$ inherits the requirements of (A3) so that it now suffices to establish the Cauchy Data Theorem for (3.5). But (3.5) has the same form as the source integral

(2.3), except the t integration in (2.3) is replaced by $t=0$ evaluation in (3.5). As a result, the completion of the proof may be modeled upon the proof of the Source Theorem, with the integrations by parts now done with respect to z. We will only indicate the main steps.

The first parts integration of (3.5) gives

$$2\pi\Phi = \epsilon^2 \int (D_z f)\theta(\lambda) J_0(m\lambda^{1/2}) dv. \tag{3.7}$$

Next, the Bessel function relationships (2.7) and (2.8) lead to the iterative formula

$$\int (D_z f)\theta(\lambda) J_0(m\lambda^{1/2}) dv = \left(\frac{2\epsilon^2}{m}\right)^n \int (D_z{}^{n+1} f)\theta(\lambda)\lambda^{n/2} J_n(m\lambda^{1/2}) dv, \tag{3.8}$$

in analogy with (2.14). Then (3.7) and (3.8) combine to give

$$2\pi\Phi = \left(\frac{2}{m}\right)^n \epsilon^{n+2} \int (D_z{}^{n+1} f)\theta(\lambda) \times h^{n/2} J_n(m\lambda^{1/2}) dv\, dt, \tag{3.9}$$

for $n \geqslant 0$. This leads directly to the Cauchy Data Theorem.

Other initial value theorems can be based upon the Cauchy Data Theorem. For instance, consider the characteristic initial value problem for a Klein–Gordon field on a null cone N, for which the initial data consists solely of Φ itself. To be explicit, let N be the future null cone of the point $x^\alpha = 0$. Assign smooth, compact null cone data on N by the restriction of an arbitrary four dimensionally smooth, compact function to N. Then these data determine a unique Klein–Gordon field Φ to the future of N. Now, on a Cauchy hypersurface $t = t_1 > 0$ such that the support of the null cone data on N lies within $r = t_1$, assign Cauchy data as follows. Set the Cauchy data to zero on the part of t_1 exterior to the future of N and to the data induced by Φ on the remainder of t_1. These Cauchy data then satisfy assumption (A3) so that the Cauchy Data Theorem implies that the evolution of smooth, compact null cone data for a source-free massive field also satisfies $R^n\Phi \to 0$ at $\mathscr{I}^+$, for all n.

IV. SUMMARY

We have been motivated by the question: Are massive fields consistent with the asymptotics of massless fields at null infinity, when coupled according to general relativity? The scope of this paper has been to determine whether any conflicts might loom at the level of linearization off Minkowski space. There are two potential problems. One involves the global boundary conditions on the massive fields and their sources. The other stems from the intrinsic asymptotic features of massive fields, i.e., the bizarre asymptotic behavior of their retarded Green's function. We have skirted the first of these problems by considering massive fields whose sources and Cauchy data have compact space-time support. Our results are captured in the Source Theorem and the Cauchy Data Theorem. Together they imply that $R^n\Phi \to 0$ at $\mathscr{I}^+$, for all n, when Φ is the retarded field of a smooth source with compact space-time support superimposed with a source-free field having compact Cauchy data.

Clearly these theorems must also hold under weaker assumptions. For instance, in the case of a spatially compact, spherically symmetric source, the exterior field is the Yukawa potential, so that dropping temporal compactness does not necessarily lead to problems. By superposition of spatially translated versions of such spherical sources, the Source Theorem extends to spatially compact, smooth, static sources with arbitrary multipole dependence. Furthermore, by superposition of boosted versions of such multipole sources, it extends to a class of nonstatic sources, albeit quite restricted, whose spatial extent is unbounded in the infinite past and future. We would also expect extension to more general sources whose velocity is asymptotically constant in the past. On the other hand, there is a limit as to how fast the size of the source can grow asymptotically in time and still allow the Source Theorem to hold. This is indicated by the following example of a solution whose source and Cauchy data has compact support on every Cauchy hypersurface but for which $r\Phi$ does not have vanishing limit at $\mathscr{I}$. Construct this solution by setting $\Phi = (f(t-r) - f(t+r))/r$, where f is a smooth function of compact support. Then Φ is a smooth homogeneous solution of the massless wave equation but as a solution of the massive Klein–Gordon equation has source $\rho = m^2\Phi$. In this example, the massive field gets to $\mathscr{I}$ because its source moves at the speed of light.

This last example also indicates that a more general theorem would have to be posed in terms of an initial value problem based upon $\mathscr{I}^-$ as well as past timelike infinity. The problem here is that the massive wave equation is not regular at $\mathscr{I}^-$, as discussed earlier in conjunction with (1.5). Can somehow the singularity in (1.5) be renormalized? This would open up a new approach to the behavior of massive fields at $\mathscr{I}$ which, being free of any detailed knowledge of the Green's function, might extend to the fully general relativistic case. Under assumptions (A1)–(A3), all terms vanish in a $1/r$ power series expansion of a massive field about $\mathscr{I}^+$. (Or equivalently, in the compactified picture, all derivatives of the massive field vanish at $\mathscr{I}^+$.) Thus, under these assumptions, the massive field has the same asymptotic properties as the Yukawa potential. This suggests a renormalization $\Phi = \tilde{\Phi} e^{-mr}/r$, where $\tilde{\Phi}$ would satisfy a regular equation in a neighborhood of $\mathscr{I}$. But, taking $\Omega = 1/r = \hat{r}$ and using the compactified metric in coordinates described by (1.1), the Klein–Gordon equation implies

$$\tilde{\Box}\tilde{\Phi} - 2m(\partial_{\hat{r}} + \hat{r}^{-2}\partial_u)\tilde{\Phi} = 0, \tag{4.1}$$

in a neighborhood of $\mathscr{I}^+$ in which $\rho = 0$. In (4.1), the singular potential of (1.5) has been removed but the singularity has reappeared in the $\hat{r}^{-2}\,\partial_u\tilde{\Phi}$ term, so that this approach is not useful.

Thus while the results of this paper indicate compatibility of massive fields with the asymptotic properties of massless fields at $\mathscr{I}$ at the level of linearized theory, some more powerful approach is necessary to investigate the general relativistic case.

ACKNOWLEDGMENT

This research was supported by National Science Foundation Grant No. PHY-8403708.

[1]R. Penrose, Proc. R. Soc. London Ser. A **284**, 159 (1965).
[2]R. Penrose, Phys. Rev. Lett. **10**, 66 (1963).
[3]R.P. Geroch, in *Asymptotic Structure of Space-time*, edited by E. P. Esposito and L. Witten (Plenum, New York, 1977).
[4]A. Ashtekar, in *General Relativity and Gravitation*, edited by A. Held (Plenum, New York, 1980), Vol. 2.
[5]H. Friedrich, Commun. Math. Phys. **100**, 525 (1985).
[6]R. G. Cawley, Ann. Phys. (NY) **54**, 122 (1969).
[7]R. G. Cawley, J. Math. Phys. **11**, 761 (1970).
[8]G. N. Watson, *A Treatise on the Theory of Bessel Functions* (Cambridge U.P., Cambridge, 1944), 2nd ed.

A class of models of the universe filled with multidimensional objects

Piotr Chmielowski
Department of Physics, The Pennsylvania State University, 104 Davey Laboratory, University Park, Pennsylvania 16802

(Received 4 November 1987; accepted for publication 11 May 1988)

An extension of Friedmann–Robertson–Walker models to the case of universes filled with perfect dust of multidimensional objects is investigated in this paper. It is shown that the internal space scale factor can become very small due to cosmological evolution regardless of the dimension of this space. This result is similar to those of flat [W. Kopczynski, Phys. Rev. D **36**, 3589 (1987)] or vacuum [P. Turkowski and K. Maslanka, Gen. Relativ. Gravit. **19**, 611 (1987)] models.

I. INTRODUCTION

A description of the perfect fluid of multidimensional objects was recently developed (see Refs. 1–3). A model of the flat universe filled with the dust of such objects is considered in Ref. 3. In our paper an extension to a class of nonflat models is considered. A similar class of vacuum models is considered in Ref. 4.

In Sec. II a description of the models leading to field equations is presented. These equations are solved in Sec. III in the case of a universe filled with strings. Section IV contains a discussion of the solutions for models with three- or higher-dimensional objects considered as the sources of the field. Although general solutions are not known, their shape and most interesting features are determined.

It is stated that the evolution of the internal space to a very small size can be due to field equations, as well as in the flat and vacuum models. For a dust of three- or higher-dimensional objects the entire evolution time from singularity to singularity can be finite, even for open models.

II. COSMOLOGICAL MODELS AND THEIR FIELD EQUATIONS

We assume that the space-time is an m-dimensional manifold such that its spatial cross section consists of the $(k-1)$-dimensional internal space and $(m-k)$-dimensional external space. Our fundamental assumption is that the space-time is spatially homogeneous and isotropic both in internal and external spaces. Both internal and external spaces have spatially constant, but different, curvatures.

Then the metric has the form

$$ds^2 = dt^2 - a_I{}^2 \frac{\sum_{i=1}^{k-1}(dx^i)^2}{\left[1+(\alpha_I/4)\sum_{i=1}^{k-1}(x^i)^2\right]^2} - a_E^2 \frac{\sum_{i=k}^{m-1}(dx^i)^2}{\left[1+(\alpha_E/4)\sum_{i=k}^{m-1}(x^i)^2\right]^2}.$$

Constants α_I and α_E have the values ± 1 or 0. Two time-dependent functions a_I and a_E are the internal and external scale factors (radii) of the universe.

Therefore, the metric is the simplest extension of the Robertson–Walker metric distinguishing between internal and external spaces. We assume that the universe is filled with a perfect fluid of multidimensional objects described in Ref. 2. Therefore, the energy-momentum tensor is

$$t^i{}_j = (\varepsilon + p)q^i_j - p\delta^i_j,$$

where

$$q^i_j = \vartheta^{0i}\vartheta^0{}_j - \vartheta^{1i}\vartheta^1{}_j - \cdots - \vartheta^{k-1,i}\vartheta^{k-1}{}_j,$$

and the frame ϑ^i is defined as follows:

$$\vartheta^0 = dt,$$

$$\vartheta^i = \frac{a_I\,dx^i}{1+(\alpha_I/4)\sum_{i=1}^{k-1}(x^i)^2}, \quad i = 1,2,\ldots,k-1,$$

$$\vartheta^i = \frac{a_E\,dx^i}{1+(\alpha_E/4)\sum_{i=k}^{m-1}(x^i)^2}, \quad i = k,k+1,\ldots,m-k.$$

We use the notations from Ref. 3. Let

$$H_I = \dot{a}_I/a_I$$

and

$$H_E = \dot{a}_E/a_E,$$

where the dot denotes differentiation over time.

The evolution of the model can be determined from Einstein equations that have the following form:

$$\tfrac{1}{2}(k-1)(k-2)H_I{}^2 + \tfrac{1}{2}(m-k)(m-k-1)H_E{}^2 + (m-k)(k-1)H_I H_E + \tfrac{1}{2}(k-1)(k-2)(\alpha_I/a_I^2) + \tfrac{1}{2}(m-k)(m-k-1)(\alpha_E/a_E^2) = \varepsilon, \quad (1)$$

$$(k-2)\dot{H}_I + (m-k)\dot{H}_E + \tfrac{1}{2}(k-1)(k-2)H_I{}^2 + \tfrac{1}{2}(m-k)(m-k+1)H_E{}^2 + (m-k)(k-2)H_I H_E + \tfrac{1}{2}(k-2)(k-3)(\alpha_I/a_I^2) + \tfrac{1}{2}(m-k)(m-k-1)(\alpha_E/a_E^2) = \varepsilon, \quad (2)$$

$$(k-1)\dot{H}_I + (m-k-1)\dot{H}_E + \tfrac{1}{2}k(k-1)H_I{}^2 + \tfrac{1}{2}(m-k)(m-k-1)H_E{}^2 + (k-1)(m-k-1)H_I H_E + \tfrac{1}{2}(k-1)(k-2)(\alpha_I/a_I^2) + \tfrac{1}{2}(m-k-1)(m-k-2)(\alpha_E/a_E^2) = -p. \quad (3)$$

We consider the case of dust $p = 0$.

The energy-momentum conservation law can be written in the following form:

$$a_E{}^{m-k}\varepsilon = M(m-k), \quad (4)$$

 0022-2488/88/092122-05$02.50

where the right-hand side is a constant chosen following Ref. 3.

It can be proved directly (following Ref. 3) that only two of Eqs. (1)–(3) are independent, providing the matter density fulfills (4).

III. MODEL OF THE UNIVERSE FILLED WITH DUST OF STRINGS

Our considerations of the case of strings follow Ref. 3, the only difference being in the field equations. Subtracting (1) from (2) one gets

$$(k-2)\dot{H}_I + (m-k)\dot{H}_E + (m-k)H_E{}^2 - (m-k)H_I H_E = \alpha_I(k-2)/a_I{}^2 .$$

If $k=2$ this equation implies

$$\ddot{a}_E/a_E = \dot{a}_I/a_I$$

or

$$a_I = C|\dot{a}_E| ,$$

C being an integration constant.

Therefore, from Eq. (1) it follows that

$$\ddot{a}_E/a_E + \tfrac{1}{2}(m-3)(\dot{a}_E/a_E^2) + \tfrac{1}{2}(m-3)\alpha_E a_E{}^{-2} - M a_E{}^{2-m} = 0 .$$

This is an Euler–Lagrange equation for Lagrangian

$$L = \tfrac{1}{2}a_E{}^{m-3}\dot{a}_E{}^2 + Ma_E - \tfrac{1}{2}\alpha_E a_E{}^{m-3}$$

and can be easily integrated

$$\pm\sqrt{2}t = \int\left(Ea_E{}^{3-m} + Ma_E{}^{4-m} - \frac{1}{2}\alpha_E\right)^{-1/2} da_E .$$

The term E is an integration constant.

The result is a generalization of Friedmann universes. The main difference is that the (internal) metric is always singular when the external radius reaches its extreme values. The behavior of the solutions is similar to that of the flat case (see Ref. 3).

For $m=5$ and $k=2$ we can obtain exactly the Friedmann solutions, appropriately choosing the constants M and E. However, there exist other solutions, e.g., if $\alpha_E = 1$ then the solutions can have the shape not only of a cycloid but also of hypo(hyper)-cycloids.

IV. MODELS FILLED WITH THREE- OR HIGHER-DIMENSIONAL OBJECTS

A. General analysis

An investigation of solutions for $k \geqslant 3$ is presented in this section. Objects that fill the internal space are assumed to have more dimensons than 2 (sheets, etc.). Although the exact analytic form of solutions still remains unknown, their shape and most important features are determined.

The system of Eqs. (1) –(3) is equivalent to (1), (2), and (4). Therefore, we investigate the following equations:

$$\tfrac{1}{2}(k-1)(k-2)H_I{}^2 + \tfrac{1}{2}(m-k)(m-k-1)H_E{}^2 + (m-k)(k-1)H_I H_E + \tfrac{1}{2}(k-1)(k-2)(\alpha_I/a_I^2) + \tfrac{1}{2}(m-k)(m-k-1)(\alpha_E/a_E^2) = (m-k)M/a_E{}^{m-k} ,$$

and

$$(k-2)\dot{H}_I + (m-k)\dot{H}_E + (m-k)H_E{}^2 - (m-k)H_I H_E = \alpha_I(k-2)/a_I^2 .$$

Let us change the variables:

$$a_I = f[(m-k)\ln a_E] = f(s)$$

and

$$e^s = a_E{}^{m-k}, \quad s\in[-\infty, +\infty] .$$

Let the prime denote differentiation over the new variable s. Then it is easy to see that

$$H_I = \dot{a}_I/a_I = (f'/f)(m-k)H_E .$$

Therefore, the first of the considered equations yields

$$\tfrac{1}{2}(k-1)(k-2)(f'/f)^2(m-k)^2H_E{}^2 + \tfrac{1}{2}(m-k)(m-k-1)H_E{}^2 + (m-k)^2(k-1)(f'/f)H_E{}^2 + \tfrac{1}{2}(k-1)(k-2)(\alpha_I/f^2) + \tfrac{1}{2}(m-k)(m-k-1)\alpha_E e^{-2s/(m-k)} = M(m-k)e^{-s} ,$$

where

$$H_E = \dot{a}_E/a_E = [1/(m-k)]\dot{s} .$$

This equation is then equivalent to

$$\dot{s} = \pm\left[\frac{M(m-k)e^{-s} - \tfrac{1}{2}(k-1)(k-2)\alpha_I f^{-2} - \tfrac{1}{2}(m-k)(m-k-1)\alpha_E e^{-2s/(m-k)}}{W(f'/f)}\right]^{1/2} , \tag{5}$$

where

$$W(f'/f) = \tfrac{1}{2}(k-1)(k-2)(f'/f)^2 + (k-1)(f'/f) + \tfrac{1}{2}[(m-k-1)/(m-k)] .$$

Therefore the second of the considered equations is a second-order differential equation for the function $f(s)$. Using

$$\dot{H}_I = (f'/f)'\dot{s}^2 + (f'/f)\ddot{s} ,$$

and

$$\dot{H}_E = [1/(m-k)]\ddot{s} ,$$

we finally get

$$\left\{\left(\frac{f'}{f}\right)'[(m-k-1)(k-2)-(k-1)(m-k)]\right.$$
$$\times\left.\frac{M(m-k)e^{-s}-\frac{1}{2}(k-1)(k-2)\alpha_I f^{-2}-\frac{1}{2}(m-k)(m-k-1)\alpha_E e^{-2s/(m-k)}}{2(m-k)W^2(f'/f)}\right\}-\left\{\left[(k-2)\left(\frac{f'}{f}\right)+1\right]\right.$$
$$\times\left.\frac{M(m-k)e^{-s}-(k-1)(k-2)\alpha_I f^{-2}-(m-k-1)\alpha_E e^{-2s/(m-k)}}{W(f'/f)}\right\}+\left\{\left[\frac{1}{m-k}-\left(\frac{f'}{f}\right)\right]\right.$$
$$\times\left.\frac{M(m-k)e^{-s}-\frac{1}{2}(k-1)(k-2)\alpha_I f^{-2}-\frac{1}{2}(m-k)(m-k-1)\alpha_E e^{-2s/(m-k)}}{W(f'/f)}\right\}-\frac{\alpha_I(k-2)}{f^2}=0, \tag{6}$$

Equation (6) is a nonlinear second-order differential equation. Its solution determines the relation between internal and external scale factors. If one knows the solution of Eq. (6) one can plug it into (5) and obtain the time dependence of internal and external scale factors.

In the simplest nontrivial case $\alpha_I = 0, \alpha_E \neq 0$, Eq. (6) is an Abel equation of the first kind[5] for the function (f'/f). However, solutions of such equations are unknown.

Below we present some lemmas useful in the analysis of solutions of Eqs. (5) and (6) and then turn back to a general discussion for various topological cases. We do not consider the flat case $\alpha_I = \alpha_E = 0$; it is described in Ref. 3.

Lemma 1: If in a general solution of Eqs. (5) and (6) the function s diverges to infinity, then the function f tends to zero.

Lemma 2: If in general solution of Eqs. (5) and (6) the function s diverges to minus infinity, then the function f tends to infinity.

Lemma 3: The polynomial $W(f'/f)$ is finite for every finite time t.

Lemma 4: If $\alpha_I \neq 0$ and the function $s(t)$ tends to its extreme value, then this value is reached at a finite time. The function f is not equal to zero at that time.

Lemma 5: Whenever a general solution $s(t)$ diverges to plus or minus infinity and $\alpha_i \neq 0$, then the infinite limit is reached after a finite time. If $\alpha_I = 0$ only the minus infinity value of s is reached after a finite time.

For proof of the lemmas, see The Appendix.

B. Properties of general solutions

Below we present the most important properties of the solutions of Eqs. (5) and (6) determining the shape of functions $a_E(t)$ and $a_I(t)$ and thus the evolution of the cosmological model.

(1) Case $\alpha_I = 1, \alpha_E = 1$: There are two subcases.

First, $W(f'/f)$ is positive and from (5) s must be smaller then a sufficiently large constant. Then function $s(t)$ must start from minus infinity and reach its maximum value after a finite time. Then it decreases and in a finite time tends to minus infinity.

Function s can be greater than a sufficiently small constant. Then the radii of the universe have nonsingular oscillatory behavior. Solutions for case $k = 2$ indicate that the choice between the two above shapes depends on the integration constants.

Second, it is possible that $W(f'/f)$ is negative. Then the shape of the solutions $s(t)$ is roughly a reflection in the t axis of the solutions discussed above.

(2) Case $\alpha_I = 1$, $\alpha_E = 0$ and (3) case $\alpha_I = 1$, $\alpha_E = -1$: The shape of solutions is the same as discussed in case 1.

(4) Case $\alpha_I = 0$, $\alpha_E = 1$: Function $s(t)$ must have an extremum. If it has a maximum, then the total time of evolution is finite and s tends to minus infinity at the beginning and at the end of evolution.

If $s(t)$ has a minimum, then the total time of evolution from infinity through minimum back to infinity can be infinite. It follows from Eq. (5) that in this case function f tends to zero at very early and late time.

(5) Case $\alpha_I = 0$, $\alpha_E = 0$: For analytic solutions see Ref. 3.

(6) Case $\alpha_I = 0, \alpha_E = -1$: There are two possible behaviors, the choice between them depending on the sign in Eq. (5). Function $s(t)$ can start from minus infinity and increase to infinity in a finite time or it can start from infinity and decrease to minus infinity in a finite time.

(7) Case $\alpha_I = -1$, $\alpha_E = 1$: We cannot state whether the solutions are bounded or not. If the solutions are bounded their shape is similar to that described in case 1. If the solutions are not bounded their shape is described in case 6.

(8) Case $\alpha_I = -1$, $\alpha_E = 0$ and (9) case $\alpha_I = -1$, $\alpha_E = -1$: The shape of solutions is the same as in case 6.

In all the above cases the behavior of function $f(s(t))$ is determined by Lemma 5.

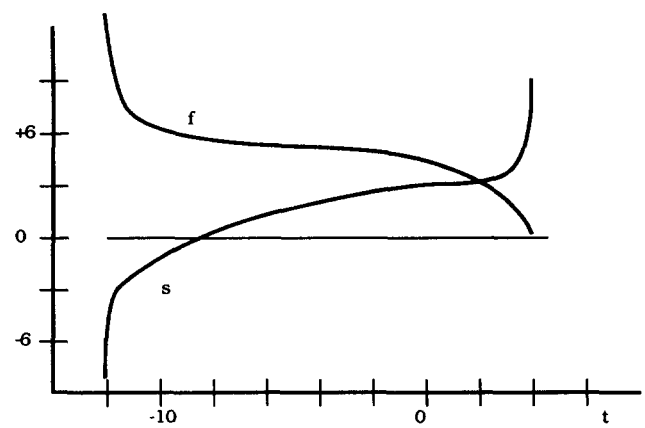


FIG. 1. Shape of solutions in cases 6–9.

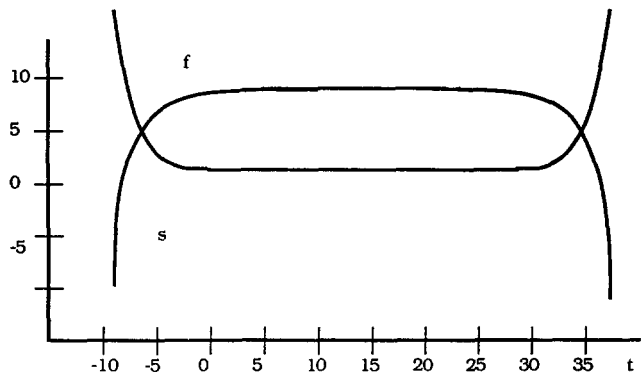


FIG. 2. Shape of solutions in cases 1–5.

C. Remarks

The above analysis cannot exclude the case in which during evolution $W(f'/f)$ becomes zero. In such a case the time derivative of $s(t)$ becomes infinite as well as the curvature scalar and the solution cannot be continued.

If solutions with $W(f'/f)$ not equal to zero at any point exist, they can have only the shape described in Sec. IV B. Some numeric solutions suggest that such solutions exist and that this is the only possible case.

Figure 1 shows numeric solutions derived in case 9. Solutions in cases 6–8 can have the same shape.

Figure 2 was derived in case 1 and it shows the possible shape of solutions for cases 2–5.

APPENDIX: PROOF OF THE LEMMAS

1. Proof of Lemma 1

Let us assume that $s \gg 1$ and that there exists such a $c \geqslant 0$ that

$$c/f^2 \simeq e^{-2s/(m-k)}$$

in the asymptotic regime. Then some of the terms in Eq. (6) can be neglected and this equation can be simplified to

$$\frac{1}{2}\left(\frac{f'}{f}\right)' \frac{[(k-2)(k-1)\alpha_I + c(m-k)(m-k-1)\alpha_E][(m-k-1)(k-2)-(k-1)(m-k)]}{(m-2)[c(m-k)(m-k-1)\alpha_E(f'/f)-(k-2)\alpha_I]W(f'/f)} = 1.$$

If s diverges to infinity then the integral

$$\int d\left(\frac{f'}{f}\right) \frac{[(k-2)(k-1)\alpha_I + c(m-k)(m-k-1)\alpha_E][(m-k-1)(k-2)-(k-1)(m-k)]}{(m-2)[c(m-k)(m-k-1)\alpha_E(f'/f)-(k-2)\alpha_I]W(f'/f)}$$

must also diverge to infinity. Therefore (f'/f) tends to one of the three values:

$$f'/f \to g_i, \quad i = 1, 2, \text{ or } 3,$$

such that

$$W(g_1) = W(g_2) = 0,$$

and

$$g_3 = (k-2)\alpha_I / c(m-k)(m-k-1)\alpha_E .$$

Values of g_1 and g_2 are negative. If (f'/f) tends to one of these values then c must be zero. Moreover,

$$f = D\exp(g_i s), \quad g_i < 0, \tag{A1}$$

where D is an integration constant.

If α_I and α_E have different signs and are not equal to zero, then g_3 is also negative and (A1) is the only possible asymptotic solution.

If α_E equals zero, then g_3 does not exist and one gets (A1) as the only solution.

If α_I and α_E have the same signs one could obtain

$$f = [(k-2)/(m-k-1)]e^{s/(m-k)} \tag{A2}$$

as another possible solution. However, Eq. (6) is a second-order differential equation and its general solution should depend on two arbitrary constants. The leading asymptotic term should depend on at least one arbitrary constant, as solution (A1) really does. Function (A2) does not depend on any arbitrary constant and thus cannot be considered as an asymptotic behavior of any general solution of Eq. (6).

One can also examine the limit $c \to \infty$ of the considered integral, but one does not get any new results.

There is another possibility—that the value of $f^2 e^{-2s/(m-k)}$ oscillates when s becomes large. We do not consider such a case because we found an asymptotic form of solution (A1) depending on an appropriate number of arbitrary constants. For every initial condition Eq. (6) should have a unique solution and thus its asymptotic form is (A1).

2. Proof of Lemma 2

We assume that

$$C/f^2 \simeq e^{-s}, \quad \text{for } s \ll -1.$$

Then Eq. (6) simplifies to

$$\frac{1}{2}\left(\frac{f'}{f}\right)'\left[CM(m-k) - \frac{(k-1)(k-2)}{2}\right]\left[\frac{(m-k-1)(k-2)}{m-k} - k + 1\right]$$
$$\times\left(W\left(\frac{f'}{f}\right)\left[CM(m-k)\left[\frac{1}{2}(k-2)+1\right]\left(\frac{f'}{f}\right) + CM\left[\frac{1}{2}(m-k)-1\right] + \frac{\alpha_I(k-2)(m-2)}{2(m-k)}\right\}\right)^{-1} = 1.$$

The same analysis as in Lemma 1 leads to the following asymptotic form of solution:

$$f = F\exp(g_i s), \quad \text{for each } g_i < 0, \tag{A3}$$

F being an integration constant.

Values of g_1 and g_2 are the same as in Lemma 1 and

$$g_3 = -\{CM[\tfrac{1}{2}(m-k)-1] + \alpha_I(k-2)(m-2)/(m-k)\}/CM(m-k)[\tfrac{1}{2}(k-2)+1].$$

If g_3 is non-negative one can get an additional asymptotic solution that does not become infinite for $s \to -\infty$. However, this solution does not depend on any arbitrary constant and thus does not originate from a general solution. Moreover, the same analysis as in Lemma 1 shows that (A3) is the only asymptotic solution of Eq. (6) for very small external scale factors.

3. Proof of Lemma 3

Let us consider the limit $W(f'/f) \to \pm\infty$ as $s \to s_0$ of Eq. (6). Then we get

$$1/f^2 \sim 1/(f'/f)^2$$

or

$$1/f \sim (s - s_0).$$

Then from Eq. (5) we get

$$\dot{s} \sim (s - s_0),$$

or finally

$$\ln(s - s_0) \sim (t - t_0).$$

Therefore, for any finite time t, s is not equal to s_0 and $W(f'/f)$ is finite.

4. Proof of Lemma 4

Let us assume that the numerator of the right-hand side of Eq. (5) tends to zero, whereas s tends to s_0 and $W(f'/f)$ does not tend to zero. Then, from Eq. (6) we find that (f'/f) tends to a finite value and

$$f = C\exp(s - s_0).$$

Equation (5) leads to

$$(s - s_0) \sim (t - t_0)^2.$$

The extreme value s_0 is reached in a finite time.

Equations (1)–(3) contain a second-order derivative of function s with respect to time. So this derivative should be continuous. If a maximum (minimum) value is reached the solution must be continued for $t > t_0$ in such a way that $t = t_0$ is actually a maximum (minimum).

5. Proof of Lemma 5

Now we assume $\alpha_I \neq 0$. Let us consider $s \gg 1$. Then in Eq. (5) some terms may be neglected and others substituted from (A1). Finally we get

$$\dot{s} \sim \exp(|g_i|s), \quad \text{for } i = 1, 2, \text{ or } 3,$$

or

$$\exp(-|g_i|s) \sim (t - t_0).$$

For $s \gg 1$ the left-hand side is limited, so the right-hand side must be limited, too.

If we assume $s \ll -1$ then from Eqs. (A3) and (5) we obtain

$$\dot{s} \sim e^{-s/2}$$

and

$$e^{s/2} \sim (t - t_0).$$

For $s \ll -1$ the left-hand side is limited and so is the right-hand side.

[1] J. Stachel, Phys. Rev. D **21**, 2171 (1980).
[2] W. Kopczynski, Phys. Rev. D **36**, 3582 (1987).
[3] W. Kopczynski, Phys. Rev. D **36**, 3589 (1987).
[4] P. Turkowski and K. Maslanka, Gen. Relativ. Gravit. **19**, 611 (1987).
[5] E. Kamke, *Differentialgleichungen, Losungsmethoden und Losungen* (Chelsea, New York, 1959).

The Dirac equation in a non-Riemannian manifold. I. An analysis using the complex algebra

S. Marques[a)]
NASA/Fermilab Astrophysics Center, Fermi National Accelerator Laboratory, M.S. 209, P.O. Box 500, Batavia, Illinois 60510

(Received 9 December 1987; accepted for publication 27 April 1988)

The Dirac wave equation is obtained in the non-Riemannian manifold of the Einstein–Schrödinger nonsymmetric theory. A new internal connection is determined in terms of complex vierbeins, which shows the coupling of the electromagnetic potential with gravity in the presence of a spin-$\frac{1}{2}$ field.

I. INTRODUCTION

The Einstein–Schrödinger (ES) nonsymmetric theory[1] was an attempt to geometrize, in an unitary way, the gravitational and the electromagnetic fields. However, structure problems in the theory have not permitted a coherent interpretation of the field equations. On the other hand, the technique of geometrizing fields introduced by Einstein continues to be a useful tool in classical field theory. Actually there has been some work[2] where the electromagnetic field is made explicit; this is accomplished by first taking the skew-symmetric part of the metric as being proportional to the electromagnetic tensor and, then, adding a sourcelike term to the Lagrangian of the theory. This permits one to reobtain the Einstein–Maxwell equations through a correspondence principle. When we face the problem of developing a Dirac theory using the space-time manifold of the ES nonsymmetric theory, we use then complex vierbeins. This is equivalent to introducing an internal **C** space in the manifold of the general relativity theory. However, we have the problem of how to obtain a coherent mechanism that fits the correct rules for the transformation group. We can see, for example, that the way the problem was developed by Marques and Oliveira,[4] in a study of the geometrical properties of **C**, **Q**, and **O** tangent spaces, does not permit the corresponding Dirac field equations to exist (we will consider the complex case presently). The reason is that the internal **C** connection was ignored when they introduced complex vierbeins. Instead, they generalize the tangent real connection (the one originating on the local real-tangent space) to a complex one. This induced them to generalize both the Lorentz group (to a pseudo-unitary group) and its representation $U(L)$, for the generalized Dirac field theory. However, in spite of it being possible to show that, on the local tangent space, the trace of the symmetric part of the tangent connection should correspond to the electromagnetic field, there is no way to obtain the desirable correspondence to an (actual **R**) Dirac field theory. The problem can be solved by considering besides the tangent connection, the connection corresponding to an internal **C** space. This forces us to maintain the Lorentz group as being that of the (local) space-time transformations on the (local) tangent space. On the other hand, we must also have an "internal" transformation, corresponding to the internal **C** space.

In the space-time of general relativity it is possible to generalize the Dirac field equation by doing the transition

$$\gamma^{\mu} \to \gamma^{\mu}(x)\,, \quad \psi_{,\mu} \to \psi_{\|\mu} = \psi_{,\mu} + \Delta_{\mu}\psi\,,$$

where $\psi(x)$ is now the electron wave function in the curved space-time, and Δ_{μ} is the geometrical connection with relation to the internal space generated by the constant γ matrices $\{\Gamma_i\}$. It is easy to show that Δ_{μ} is given by

$$\Delta_{\mu} = \tfrac{1}{8}([\gamma^{\nu},\gamma_{\nu,\mu}] - \{{}^{\rho}_{\mu\nu}\}[\gamma^{\nu},\gamma_{\rho}])\,, \tag{1.1}$$

where $\{{}^{\rho}_{\mu\nu}\}$ are the Christoffel symbols for the space-time connection. In terms of real vierbeins h^{ν}_{a} (and their inverses h^{a}_{ν}), (1.1) is written as

$$\Delta_{\mu} = (-1/4i)(h^{\nu b}h^{a}_{\nu,\mu} - \{{}^{\rho}_{\mu\nu}\}h^{a}_{\rho}h^{\nu b})\sigma_{ab}\,. \tag{1.2}$$

The function $\psi^{i}(x)$, $i = 1,\ldots,4$, above, satisfies a Dirac equation defined on the curved space-time manifold of general relativity, which now has the form

$$\gamma^{\mu}(x)(\psi_{,\mu} + \Delta_{\mu}\psi) - \mu\psi = 0\,, \tag{1.3}$$

where μ is a mass coefficient. Thus the gravitational field is present in this equation through the connection Δ_{μ}.

The nonsymmetric manifold of the ES theory with the locally defined complex vierbeins referred to above will be used in this work. These vierbeins define new Fock–Ivanenko coefficients which permit the construction of the corresponding Dirac equations related to the non-Riemannian manifold of the ES theory. In Sec. II we will present briefly the properties of the complex tangent space as well as the corresponding field equations obtained in the ES nonsymmetric theory. In Sec. III the generalized Fock–Ivanenko coefficients will be determined, as well as the new Dirac equation. In Sec. IV we will proceed to their analysis. Throughout, we will use the $\mu,\nu,\ldots$ indices as those on the non-Riemannian manifold; the $a,b,\ldots$ indices will be those on the complex tangent space.

II. THE COMPLEX TANGENT SPACE

According to the correspondence principle there exists, at each point of the curved space-time of general relativity, a local tangent space[5] with the structure of a flat space-time, with the metric given by the Minkowski tensor η_{ab}. Therefore, we must have the line element

[a)] On leave from UNESP, Campus de Ilha Solteira, São Paulo, Brazil.

 0022-2488/88/092127-05$02.50

$ds^2 = g_{\mu\nu}\,dx^\mu\,dx^\nu = \eta_{ab}\,dx^a\,dx^b$, locally, where, $g_{\mu\nu} = g_{\nu\mu}$.

In the ES nonsymmetric theory, the metric of curved space-time has the symmetry property $g^*_{\mu\nu} = g_{\nu\mu}$, $g_{\mu\nu} = g_{\mu\nu} + ik_{\mu\nu}$. Defining complex vierbeins e^a_μ (and their inverse e^μ_a), we have that[6]

$$g_{\mu\nu} = e^{*a}_{\ \nu} e^b_\mu \eta_{ab}, \tag{2.1}$$

$$g^{\mu\nu} = e^{*\mu}_{\ a} e^\nu_b \eta^{ab}, \tag{2.2}$$

where η_{ab} (and its inverse η^{ab}) is the metric of the tangent space which we take here as the Minkowski tensor. The metric $g_{\mu\nu}$ and its inverse $g^{\mu\nu}$ are such that $g^{\mu\nu}g_{\sigma\nu} = \delta^\mu_\sigma$, where the order of indices are significant. From there, we obtain the orthogonality conditions for the complex vierbeins:

$$e^{*a}_{\ \mu} e^\mu_b = e^a_\mu e^{*\mu}_{\ b} = \delta^a_b\,, \tag{2.3}$$

$$e^{*a}_{\ \mu} e^\nu_a = e^a_\mu e^{*\nu}_{\ a} = \delta^\nu_\mu. \tag{2.4}$$

As is well known, the transformation law for vectors in the complex tangent space, local to a curved space-time, is defined by

$$e'^a_\mu(x) = L^{\ a}_{\ b}(x) e^b_\mu(x)\,, \tag{2.5}$$

where $L^{\ a}_{\ b}$ are the Lorentzian rotation matrices, which have the property

$$L^{\ T}\eta L = \eta \tag{2.6}$$

and, as $e^a_\mu(x)$ is a complex function, $e^a_\mu = e^a_{\mu_R} + ie^a_{\mu_I}$. Then

$$\bar{e}^a_\mu(x) = e^a_{\mu_R} - ie^a_{\mu_I} \tag{2.7}$$

is the conjugate of e^a_μ. This means that we have attached to the Minkowskian tangent space, an "internal space," the **C** space. The "internal" transformation law of an object of the **C** space, K, is

$$K' = U(1)K\,, \tag{2.8}$$

where $U(1)$ stands for a unitary 1×1 (local) transformation matrix, $U(1) = e^{i\phi(x)}$, and

$$\bar{K}' = \bar{U}(1)\bar{K}\,, \tag{2.9}$$

where $\bar{U}(1) = U^{-1}(1) = e^{-i\phi(x)}$. A more general transformation law for the complex vierbeins now should be

$$e'^a_\mu(x) = U(1)L^{\ a}_{\ b}(x)e^b_\mu(x)\,. \tag{2.10}$$

The covariant derivative of the vierbeins e^a_μ and $e^{*a}_{\ \mu}$ on this tangent space are now given by

$$e^a_{\mu\|\nu} = e^a_{\mu,\nu} + \Lambda_\nu^{\ a}{}_b e^a_\mu + C_\nu e^b_\mu\,, \tag{2.11}$$

$$e^{*a}_{\mu\|\nu} = e^{*a}_{\mu,\nu} + \Lambda_\nu^{\ a}{}_b e^{*b}_\nu - C_\nu e^{*a}_{\ \mu}\,, \tag{2.12}$$

where $\Lambda_\nu^{\ a}{}_b$ is the tangent connection related to the Minkowskian space and C_ν is the "internal connection." Their transformation laws are, respectively,

$$\Lambda'_\nu = L\Lambda L^{-1} - L_{,\nu}L^{-1} \quad \text{(space-time transformations)}\,, \tag{2.13}$$

$$C'_\nu = U(1)C_\nu U^{-1}(1) - U_{,\nu}(1)U^{-1}(1) \quad \text{(internal transformations)}\,. \tag{2.14}$$

Here C_ν transforms as a vector under space-time transformations. Considering the particular case where we have only the internal transformations represented by the matrices $U(1) = 1 + i\phi$, the internal connection C_ν transforms in first order, as

$$C'_\nu = C_\nu + i\phi_{,\nu}\,, \tag{2.15}$$

which is the same as a gauge transformation law for an electromagnetic potential. We also have the relation

$$R^{\ \rho}_{\mu\nu\gamma} e^a_\rho - S_{\nu\gamma}^{\ \ a}{}_b e^b_\mu = 0\,, \tag{2.16}$$

where $R^{\ \rho}_{\mu\nu\gamma}$ is the curvature in the non-Riemannian space-time written in terms of the nonsymmetic affinity, and $S_{\nu\gamma}$ is the curvature over the complex tangent space,

$$S_{\nu\gamma} = \Lambda_{\nu,\gamma} - \Lambda_{\gamma,\nu} - [\Lambda_\nu, \Lambda_\gamma]\,, \tag{2.17}$$

which is skew symmetric with respect to the curved-space indices and anti-Hermitian in the tangent space indices. The internal curvature can be also obtained:

$$P_{\nu\gamma} = C_{\nu,\gamma} - C_{\gamma,\nu}\,. \tag{2.18}$$

Then, in the particular case of (2.15), the internal curvature can be considered to correspond to the Maxwell electromagnetic tensor.

One of the field equations of the ES nonsymmetric theory, obtained through a variational principle, is $g_{\mu\ (\underset{+}{\nu};\underset{-}{\alpha})} = 0$, where the symbol (;) means that the connection used in this equation is the Schrödinger connection,[1] $\theta^\rho_{\ \mu\nu}$, $\theta_\mu = \theta^\rho_{\ \mu\underset{\nu}{\rho}} = 0$. (The notation used in this work is the same used by Marques and Oliveira in Ref. 4. It has been kept the tranditional use of the " $+$ $-$ " covariant derivative used in the ES nonsymmetric theory, Ref. 1. See also, M. A. Tonellat, Ref. 9.) (In general, $\theta^\rho_{\ \mu\alpha}$ is a nonsymmetric connection such that $\theta^{*\rho}_{\ \ \mu\alpha} = \theta^\rho_{\ \alpha\mu}$.) This equation corresponds to the following vierbein equations:

$$e^a_{\mu|\underset{+}{\alpha}} = \left(e^{*a}_{\mu|\underset{-}{\alpha}}\right)^* = e^a_{\mu,\alpha} - \theta^\rho_{\ \mu\alpha}e^a_\rho + \Lambda_\alpha^{\ a}{}_b e^b_\mu + C_\alpha e^a_\mu = 0\,, \tag{2.19}$$

where (2.1) was used. Taking the inverse equation: $g^{\mu\ \nu}_{\ \ (\underset{+}{\ };\underset{-}{\ })\alpha} = 0$, and (2.2), we have the corresponding equations for the inverse vierbeins

$$e^{*\mu}_{\ a|\underset{+}{\alpha}} = \left(e^\mu_{a|\underset{-}{\alpha}}\right)^* = e^{*\mu}_{\ a,\alpha} + \theta^\mu_{\ \rho\alpha}e^{*\rho}_{\ a} - \Lambda_\alpha^{\ b}{}_a e^{*\mu}_{\ b} - C_\alpha e^{*\mu}_{\ b} = 0\,. \tag{2.20}$$

We can rewrite Eqs. (2.13) and (2.14) as

$$e^a_{\mu|\underset{+}{\alpha}} = e^a_{\mu,\alpha} - \theta^\mu_{\ \rho\alpha}e^\rho_a - \Lambda_\alpha^{\ b}{}_a e^\mu_b = 0\,, \tag{2.21}$$

$$e^{*\mu}_{\ a|\underset{+}{\alpha}} = e^{*\mu}_{\ a,\alpha} + \theta^\mu_{\ \rho\alpha}e^{*\rho}_{\ a} + \Lambda_\alpha^{\ b}{}_a e^{*\mu}_{\ b} = 0\,, \tag{2.22}$$

where

$$\Lambda_\alpha^{\ a}{}_b = \Lambda_\alpha^{\ a}{}_b + \delta^a_b C_\alpha\,, \quad \Lambda^*_\alpha{}^a{}_b = \Lambda_\alpha^{\ a}{}_b - \delta^a_b C_\alpha\,. \tag{2.23}$$

From (2.19) and (2.20) we obtain the relation

$$\Lambda_\alpha^{\ a}{}_b = e^a_\mu e^{*\mu}_{\ b;\underset{+}{\alpha}} = -e^a_{\mu;\underset{+}{\alpha}} e^{*\mu}_{\ b}\,, \tag{2.24}$$

and from (2.23),

$$\Lambda_\alpha = \mathrm{Re}\left[e^a_\mu e^{*\mu}_{\ b;\underset{+}{\alpha}}\right] = \mathrm{Re}\left[-e^a_{\mu;\underset{+}{\alpha}} e^{*\mu}_{\ b}\right]\,, \tag{2.25}$$

$$C_\alpha = i\left(\mathrm{Im}\left[e^a_\mu e^{*\mu}_{\ b;\underset{+}{\alpha}}\right]\right) = i\left(\mathrm{Im}\left[-e^a_{\mu;\underset{+}{\alpha}} e^{*\mu}_{\ b}\right]\right)\,. \tag{2.26}$$

Taking (2.25), we can expand it in terms of real and imaginary parts. We then obtain for Λ_α

$$\Lambda_\alpha = \mathrm{Re}\left[e^a_\mu e^{*\mu}_{\underset{+}{b};\alpha} \right]$$
$$= (e^a_{\alpha_R} e^\mu_{b_{R,\alpha}} + e^a_{\mu_I} e^\mu_{b_{I,\alpha}} + e^a_{\mu_R} \theta^\mu_{\underline{\rho\alpha}} e^\rho_{b_R} + e^a_{\mu_I} \theta^\mu_{\underline{\rho\alpha}} e^\rho_{b_I}$$
$$- e^a_{\mu_I} \theta^\mu_{\underset{\vee}{\rho\alpha}} e^\rho_{b_R} + e^a_{\mu_R} \theta^\mu_{\underset{\vee}{\rho\alpha}} e^\rho_{b_I}) \qquad (2.27)$$

or

$$\Lambda_\alpha = \mathrm{Re}\left[- e^a_{\underset{+}{\mu};\alpha} e^{*\mu}_b \right]$$
$$= (- e^a_{\mu_{R,\alpha}} e^\mu_{b_R} - e^a_{\mu_{I,\alpha}} e^\mu_{b_I} + \theta^\rho_{\underline{\mu\alpha}} e^a_{\rho_R} e^\mu_{b_R} + \theta^\rho_{\underline{\mu\alpha}} e^a_{\rho_I} e^\mu_{b_I}$$
$$+ \theta^\rho_{\underset{\vee}{\mu\alpha}} e^a_{\rho_R} e^\mu_{b_I} - \theta^\rho_{\underset{\vee}{\mu\alpha}} e^a_{\rho_I} e^\mu_{b_R}) \,. \qquad (2.28)$$

Analogously, from (2.26), we obtain for C_α

$$C_\alpha = i(\mathrm{Im}\left[e^a_\mu e^{*\mu}_{\underset{+}{a};\alpha} \right])$$
$$= i(e^a_{\mu_I} e^\mu_{b_{R,\alpha}} - e^a_{\mu_R} e^\mu_{b_{I,\alpha}} + e^a_{\mu_R} \theta^\mu_{\underset{\vee}{\rho\alpha}} e^\rho_{b_R} + e^a_{\mu_I} \theta^\mu_{\underline{\rho\alpha}} e^\rho_{b_I}$$
$$+ e^a_{\mu_I} \theta^\mu_{\underline{\rho\alpha}} e^\rho_{b_R} - e^a_{\mu_R} \theta^\mu_{\underline{\rho\alpha}} e^\rho_{b_I}) \qquad (2.29)$$

or

$$C_\alpha = i(\mathrm{Im}\left[- e^a_{\underset{+}{\mu};\alpha} e^{*\mu}_b \right])$$
$$= i(e^a_{\mu_{R,\alpha}} e^\mu_{b_I} - e^a_{\mu_{I,\alpha}} e^\mu_{b_R} - \theta^\rho_{\underline{\mu\alpha}} e^a_{\rho_R} e^\mu_{b_I} + \theta^\rho_{\underline{\mu\alpha}} e^a_{\rho_I} e^\mu_{b_R}$$
$$+ \theta^\rho_{\underset{\vee}{\mu\alpha}} e^a_{\rho_R} e^\mu_{b_R} + \theta^\rho_{\underset{\vee}{\mu\alpha}} e^a_{\rho_I} e^\mu_{b_I}) \,. \qquad (2.30)$$

Suppose we have a theory where the antisymmetrical part of the space-time connection is zero, $\theta^\rho_{\underset{\vee}{\mu\nu}} = 0$, but still with complex vierbeins, i.e., a theory where we have a complex, antiysmmetrical part for the metric. We now obtain for the tangent and internal connections,

$$\Lambda_\nu = (e^a_{\mu_R} e^\mu_{b_{R,\nu}} + e^a_{\mu_I} e^\mu_{b_{I,\nu}} + e^a_{\mu_R} \Gamma^\mu_{\rho\nu} e^\rho_{b_R} + e^a_{\mu_I} \Gamma^\mu_{\rho\nu} e^\rho_{b_I}) \,.$$
$$= (- e^a_{\mu_{R,\nu}} e^\mu_{b_R} - e^a_{\mu_{I,\nu}} e^\mu_{b_I} + e^a_{\rho_R} \Gamma^\rho_{\mu\nu} e^\mu_{b_R} + e^a_{\rho_I} \Gamma^\rho_{\mu\nu} e^\mu_{b_I}) \qquad (2.31)$$

and

$$C_\nu = i(e^a_{\mu_I} e^\mu_{b_{R,\nu}} - e^a_{\mu_R} e^\mu_{b_{I,\nu}} + e^a_{\mu_I} \Gamma^\mu_{\rho\nu} e^\rho_{b_R} - e^a_{\mu_R} \Gamma^\mu_{\rho\nu} e^\rho_{b_I})$$
$$= i(e^a_{\mu_{R,\nu}} e^\mu_{b_I} - e^a_{\mu_{I,\nu}} e^\mu_{b_R} - e^a_{\rho_R} \Gamma^\rho_{\mu\nu} e^\mu_{b_I} + e^a_{\rho_I} \Gamma^\rho_{\mu\nu} e^\mu_{b_R}) \,, \qquad (2.32)$$

where we used the notation $\Gamma^\rho_{\mu\nu}$ for the symmetrical connection.

We can see that the relation of the (complex) metric, with the new complex vierbeins, adds new extra terms to the tangent connection. Also, the internal connection has a relation with the vierbeins, which would not exist if the vierbeins are real. It is noticeable, from (2.31) and (2.32), that the same happens in a "complex theory" without a complex torsion term. It easy to conclude, as in Ref. 2, that the Einstein–Maxwell theory is reached in a convenient limit such as to eliminate the complex part of the metric and therefore the corresponding complex ones for the vierbeins. However, some years ago, this fact was criticized by theoretical analysis,[3] which does not change the power of a geometrical analysis. Thinking from a geometrical point of view, we will go forward, obtaining the (Dirac) field equations and see what we can get in this "complex theory."

III. THE GENERALIZATION OF THE FOCK–IVANENKO COEFFICIENTS

The Dirac constant γ matrices satisfy the anticommutation relations

$$\{\gamma_a, \gamma_b\} = 2\eta_{ab} \mathbf{1}_4 \,, \qquad (3.1)$$
$$\{\gamma^a, \gamma^b\} = 2\eta^{ab} \mathbf{1}_4 \,, \qquad (3.2)$$

where η_{ab} (and its inverse η^{ab}) is the Minkowski tensor with signature $+2$, and to the relation $\gamma^a{}_{,b} = 0$. The set formed with combinations of γ matrices,

$$\{\Gamma_i\} = \{\mathbf{1}_4, \gamma_a, \sigma_{ab} = (i/2)[\gamma_a, \gamma_b],$$
$$\gamma_5 = \gamma_0\gamma_1\gamma_2\gamma_3 \,, \gamma_5\gamma_a\} \,,$$

composes a linearly independent set in the internal space of the Dirac wavefunctions ψ.

Now, multiplying (3.1) by e^{*a}_ν and e^b_μ, and using (2.1), we obtain

$$\{\gamma_\mu, \dot{\gamma}_\nu\} = 2g_{\mu\nu} \mathbf{1}_4 \,, \qquad (3.3)$$

where $g_{\mu\nu}$ is now the ES nonsymmetric metric. In (3.4) we have defined γ_μ and $\dot{\gamma}_\mu$ by

$$e^a_\mu \gamma_a = \gamma_\mu \,, \quad e^{*a}_\mu \gamma_a = \dot{\gamma}_\mu \,. \qquad (3.4)$$

Analogously, multiplying (3.2) by $e^{*\mu}_a$ and e^ν_b, we obtain

$$\{\dot{\gamma}^\mu, \gamma^\nu\} = 2g^{\mu\nu} \mathbf{1}_4 \,, \qquad (3.5)$$

where

$$e^\mu_a \gamma^a = \gamma^\mu \,, \quad e^{*\mu}_a \gamma^a = \dot{\gamma}^\mu \,, \qquad (3.6)$$

and the relation (2.2) was used. The covariant derivative of $\gamma^\mu(x)$ over the non-Riemannian manifold of nonsymmetric theory, is given by

$$\gamma_{\underset{+}{\mu}|\nu} = \gamma_{\mu,\nu} - \Omega^\rho_{\mu\nu}\gamma_\rho + [\Delta_\nu, \gamma_\mu] \,, \qquad (3.7)$$

where Δ_μ is the internal connection, corresponding to the space of generalized γ matrices (or, also, of Dirac wave functions space), and $\Omega^\rho_{\mu\nu}$ is a more general space-time affinity (that, at least in principle, includes the internal connection C_μ). Taking the identity (3.5) and the Eq. (2.19), we have that

$$\gamma_{\underset{+}{\mu}|\nu} = (e^a_{\underset{+}{\mu}} \gamma_a)_{|\nu} = (e^a_{\underset{+}{\mu}|\nu})\gamma_a = 0 \,, \qquad (3.8)$$

since γ_a is a constant matrix. In the same way, we obtain

$$\dot{\gamma}_{\underset{-}{\mu}|\nu} = (e^{*a}_{\underset{-}{\mu}|\nu})\gamma_a = 0 \,. \qquad (3.9)$$

Expanding (3.8) and (3.9) we have

$$\gamma_{\underset{+}{\mu}|\nu} = \gamma_{\mu,\nu} - \theta^\rho{}_{\mu\nu}\gamma_\rho + [\Delta_\nu, \gamma_\mu] + C_\nu\gamma_\mu = 0 \,, \qquad (3.10)$$

$$\dot{\gamma}_{\underset{-}{\mu}|\nu} = \dot{\gamma}_{\mu,\nu} - \theta^\rho{}_{\nu\mu}\dot{\gamma}_\rho + [\Delta_\nu, \dot{\gamma}_\mu] - C_\nu\dot{\gamma}_\mu = 0 \,. \qquad (3.11)$$

We can observe then, from (3.10) and (3.11), that we obtain a relation similar of that of general relativity, i.e.,

$$\Delta_\nu = (1/4i)\Lambda^{\underset{\vee}{ab}}_\nu \sigma_{ab} = (1/4i)\mathrm{Re}\left[e^a_\mu e^{*\mu b}_{\underset{+}{\quad};\nu} \right]\sigma_{ab} \,, \qquad (3.12)$$

or

$$\Delta_\nu = (1/4i)\mathrm{Re}\left[- e^a_{\underset{+}{\mu|\nu}} e^{*\mu b} \right]\sigma_{ab} \,, \qquad (3.13)$$

where it was used (2.21) for Λ_ν.

If we now consider $\psi(x)$ as the wave function of a spin-$\frac{1}{2}$ particle of mass m, placed in a non-Riemmanian manifold of ES nonsymmetric theory, $\bar{\psi}(x) = \psi^\dagger\gamma_0$ will be the wave

function of its antiparticle, and the corresponding Dirac wave equations are, respectively,

$$\gamma^{\mu}(\vec{\partial}_{\mu}+\Delta_{\mu}+C_{\mu})\psi-\mu\psi=0\,, \tag{3.14}$$

$$-\bar{\psi}(\overleftarrow{\partial}_{\mu}+\Delta_{\mu}-C_{\mu})\dot{\gamma}^{\mu}-\mu\bar{\psi}=0\,, \tag{3.15}$$

where $\mu=mc/\hbar$.

The new operator $(\partial_{,\mu}+\Delta_{\mu}+C_{\mu})$ comes from the covariant derivation of the function $\psi(x)$, which, besides being an object that transforms, locally, under the representation of Lorentz group [U(L)], also transforms under the (internal) U(1) group. Equations (3.14) and (3.15) describe particles placed in a curved non-Riemannian space-time of the ES nonsymmetric theory, since the connections Δ_{μ} and C_{μ} are now related to complex vierbeins, as well as the complex space-time connection.

Another way to obtain Eqs. (3.14) and (3.15) is through a minimal action principle. In this case the action is

$$A=\int\mathscr{L}\,d^4x\,,$$

where the Lagrangian is given by

$$\mathscr{L}=\sqrt{-g}\,[\bar{\psi}\gamma^{\mu}(\psi_{,\mu}+\Delta_{\mu}\psi+C_{\mu}\psi)+(\bar{\psi}_{,\mu}+\bar{\psi}\Delta_{\mu}-\bar{\psi}C_{\mu})\dot{\gamma}^{\mu}\psi-\mu\bar{\psi}\psi]\,. \tag{3.16}$$

From (3.15), the wave equation for the charge conjugate function, ψ^c, is

$$\dot{\gamma}^{\mu}(\psi^c_{,\mu}+\Delta_{\mu}\psi^c-C_{\mu}\psi^c)-\mu\psi^c=0\,, \tag{3.17}$$

where $\psi^c=C\bar{\psi}^T$, and C is the charge conjugate matrix. Therefore, if the wave equation of a particle is constructed with the set γ^{μ} and $(\Delta_{\mu}+C_{\mu})$, the wave equation for its "charge conjugate" will be constructed with the set $\dot{\gamma}^{\mu}$ and $(\Delta_{\mu}-C_{\mu})$.

Let us write now the internal connection C_{μ} as

$$C_{\mu}=ileA_{\mu}(x)\,. \tag{3.18}$$

Then, after (2.15), we can interpret e as the electric charge for the electron, $A_{\mu}(x)$ as the electromagnetic potential, and l will be a constant such that it balances units. Equations (3.14) and (3.17) can be written now as

$$\gamma^{\mu}(\partial_{\mu}+\Delta_{\mu}+ielA_{\mu})\psi-\mu\psi=0\,, \tag{3.19}$$

$$\dot{\gamma}^{\mu}(\partial_{\mu}+\Delta_{\mu}-ielA_{\mu})\psi^c-\mu\psi^c=0\,. \tag{3.20}$$

IV. CONCLUSION

We have learned that complexifying the space-time manifold of general relativity is equivalent to attaching it an internal C space. The new metric is no longer symmetric and its antisymmetric part should be proportional to the electromagnetic tensor [see Einstein, Ref. 1, Eqs. (11)–(17) in Sec. III]. Through complex vierbeins, it is possible to obtain a (complex) tangent space local to the nonsymmetric curved space of the ES type. Using these concepts we obtained here Dirac field equations for a spin-$\frac{1}{2}$ particle. The internal complex connection corresponds to the electromagnetic potential.

We observe that it is possible to define the complex vierbeins as

$$e^a_{\mu}=e^a_{\mu_R}+i\kappa\lambda n^a_{\mu}\,, \tag{4.1}$$

where κ is considered now as a parameter, and λ is a constant. We use here, as in Ref. 2,

$$\lambda\propto\frac{e}{L^2}\sim\frac{c^3e}{\hbar G}=1.82\times10^{56}\,\frac{\text{statvolt}}{\text{cm}}\,.$$

In a limit where the parameter $\kappa\to0$, we obtain from (2.27) and (2.28) the real connection Λ_{ν} from general relativity. Using the above expression for the vierbeins, we can display the interesting behavior of the new Dirac equations that appears when we split up $\gamma^{\mu}(x)$ in terms of its real and imaginary parts, and also suppose a complex mass term: $\mu=\mu_R+i\mu_I$, where we again can take $\mu_I=\kappa\lambda m$. Then, from (3.19),

$$[e^{\mu}_{a_R}\gamma^a(\partial_{\mu}+\Delta_{\mu}+ielA_{\mu})\psi+\nu\mu_R\psi]+ik\lambda\,[n^{\mu}_a\gamma^a(\partial_{\mu}+\Delta_{\mu}+ielA_{\mu})\psi+m\psi]=0\,. \tag{4.2}$$

In the limit of the parameter $\kappa\to0$, we should get the normal Dirac equations in the presence of gravitational and electromagnetic fields. Therefore, it means that we can get another identical set of Dirac equations if we take $n^{\mu}_a\equiv e^{\mu}_{a_R}\sim h^{\mu}_a$, and $m\equiv\mu_R$, where h^{μ}_a and μ_R will be vierbeins and the mass term of general relativity theory.

ACKNOWLEDGMENTS

The author thanks Mark Rubin, Richard Holman, Marcelo Gleiser, and Tom R. Taylor for helpful discussions. Also, the author is very grateful to Patricio S. Letelier and Robert Geroch, who read and critized the manuscript.

This work was supported by National Research Council (CNP_q) at Universidade Federal do Mato Grosso, Brazil. This work was also supported in part by the DOE and by the NASA at Fermilab, and by CAPES-MEC Brazil.

APPENDIX: COMMENTS ON A MORE GENERAL TRANSFORMATION LAW IN A TANGENT SPACE ASSOCIATED TO A COMPLEX INTERNAL SPACE

Let us consider, instead of (2.10), a more general transformation law for objects in the complex tangent space. Considering, for instance, the vectors e^a_{μ}, it can be defined as

$$e'^a_{\mu}(x)=L^a{}_b(x)e^b_{\mu}(x)\,, \tag{A1}$$

$$e'^{*a}_{\mu}(x)=L^{*a}{}_b e^{*b}_{\mu}(x)\,. \tag{A2}$$

The complex matrix $\mathbf{L}^a{}_b$ now is a kind of pseudo-Lorentz matrix that follows the relation

$$\mathbf{L}^{\dagger}\eta\mathbf{L}=\eta\,. \tag{A3}$$

The covariant derivative of e^a_{μ} on this complex tangent space is then defined as

$$e^a_{\mu|\nu}=e^a_{\mu,\nu}+\Lambda_{\nu}{}^a{}_b e^b_{\mu}\,, \tag{A4}$$

$$e^{*a}_{\mu|\nu}=e^{*a}_{\mu,\nu}+\Lambda^*{}_{\nu}{}^a{}_b e^{*b}_{\mu}\,, \tag{A5}$$

where the affinity is complex. Its transformation law is

$$\Lambda'_{\mu}=\mathbf{L}\Lambda_{\mu}\mathbf{L}^{-1}-\mathbf{L}_{,\mu}\mathbf{L}^{-1},$$

$$\Lambda'^*_{\mu}=L^*\Lambda^*\mathbf{L}^{*-1}-\mathbf{L}^*_{,\mu}\mathbf{L}^{*-1}\,. \tag{A6}$$

It is directly shown (see Ref. 4 that, through the Einstein field equations for the nonsymmetric theory (a complex theory), $g_{\mu\,(\nu\alpha)}=0$, we obtain the same corresponding field equations for the vierbeins described in Eqs. (2.21) and

(2.22). However, we must also have $\eta_{a\ \ b|\mu}^{\ \ +\ -} = 0$, where the "minus" sign corresponds to the complex conjugate of the affinity Λ_μ:

$$\eta_{a\ \ b|\mu}^{\ \ +\ -} = \eta_{ab,\mu} - \Lambda_\mu{}^c{}_a \eta_{cb} - \Lambda^*{}_\mu{}^c{}_b \eta_{ac} = 0 . \tag{A7}$$

As η_{ab} lowers indices, we have that Λ_μ is anti-Hermitian with respect to the index of the tangent space. Then, we have

$$\Lambda_{\mu ab} = \Lambda_{\mu \underset{\vee}{ab}} + i\Lambda_{\mu \underline{ab}} . \tag{A8}$$

The expansion of L in first order is, from (A.1)–(A.3),

$$\mathbf{L} \cong \mathbf{1} + \epsilon + i\mu, \quad \mathbf{L}^{-1} \cong \mathbf{1} - \epsilon - i\mu , \tag{A9}$$

where $\epsilon = \epsilon(x)$ are infinitesimal rotation matrices as before and $\mu = \mu(x)$ are symmetric infinitesimal matrices. We can write the latter as

$$\mu_{ab} = (a + \mathbf{1}\tfrac{1}{4} \operatorname{Tr} \mu)_{ab} , \tag{A10}$$

where a is a symmetric trace-free matrix. Considering then, a particular transformation such that

$$\mathbf{L} \cong \mathbf{1} + \tfrac{1}{4}K , \quad K = \operatorname{Tr} \mu , \tag{A11}$$

the affinity Λ_α of this complex theory transforms as

$$\Lambda'_\alpha = \Lambda_\alpha - (i/4)K_{,\alpha} , \quad \operatorname{Tr} \Lambda'_\alpha = \operatorname{Tr} \Lambda_\alpha - iK_\alpha , \tag{A12}$$

which is similar to the gauge transformation of an electromagnetic potential. (In the same way, we can show that the complex part of a nonsymmetric tangent curvature obtained with the above Λ_α will be related to the Maxwell electromagnetic tensor.)

Now, from (A7), (3.8), and (3.9), we can easily obtain a relation between Λ_α and the connection Δ_α:

$$\begin{aligned} \Lambda_\alpha{}^{ab} e_{\mu b} \gamma_a &= \eta^{ac} e_{\mu c} [\Delta_\alpha^{(1)}, \gamma_a] , \\ \Lambda^*{}_\alpha{}^{ab} e^*_{\mu b} \gamma_a &= \eta^{ab} e^*_{\mu c} [\Delta_\alpha^{(2)}, \gamma_a] , \end{aligned} \tag{A13}$$

where $\Delta_\alpha^{(1)}$ and $\Delta_\alpha^{(2)}$ are now (general) Dirac connections corresponding to the Fock–Ivanenko coefficients. However, expanding $\Delta_\alpha^{(1)}$ and $\Delta_\alpha^{(2)}$ in terms of the set $\{\Gamma_i\}$, we can see that the real part of Λ_α is of the same form as in general relativity, but that there is no way to relate the complex symmetrical part of Λ_α in terms of that expansion, since the only symmetrical term there, which is proportional to the unit element of the set (the unit 4×4 matrix), is eliminated through the commutator in (A13). This shows that this is not the correct choice for the transformation matrix **L**. As we saw the correct one is the product expressed in (2.10).

[1] A. Einstein, Rev. Mod. Phys. **20**, 35 (1948); *The Meaning of Relativity* (Princeton U.P., Princeton, NJ, 1955), Appendix 2; E. Schroedinger, *Space-Time Structures* (Cambridge U.P., London, 1954).

[2] J. W. Moffat and D. H. Boal, Phys. Rev. D **11**, 1375 (1975); W. B. Bonnor, Proc. R. Soc. London Ser. A **226**, 336 (1954).

[3] R. B. Mann and J. W. Moffat, Phys. Rev. D **26**, 1858 (1982); G. Kunstater, P. Leivo, and P. Savaria, Quantum Gravit. **1**, 7 (1984); J. W. Moffat, Phys. Rev. D **19**, 1554, 3562 (1979); J. Math. Phys. **21**, 1798 (1980); G. Kunstater, J. W. Moffat, and P. Savaria, Phys. Rev. D **19**, 3559 (1979).

[4] It is in preparation an extension of this theory using quaternion and octonion algebra. See also S. Marques and C. G. Oliveira, J. Math. Phys. **26**, 3131 (1985); Phys. Rev. D **36**, 1716 (1987).

[5] D. R. Brill and J. A. Wheeler, Rev. Mod. Phys. **29**, 465 (1957); D. R. Brill, and J. M. Cohen, J. Math. Phys. **7**, 238 (1966).

[6] D. W. Sciama, J. Math. Phys. **2**, 472 (1961); P. D. P. Smith, Nuovo Cimento B **24**, 255 (1974).

[7] P. A. M. Dirac, Proc. R. Soc. London Ser. A **117**, 610 (1928).

[8] V. Fock and D. Yvanenko, C. R. Acad. Sci. **188**, 1470 (1928); V. Fock, Z. Phys. **57**, 261 (1929); see also J. L. Anderson, *Principles of Relativity* (Academic, New York, 1967), p. 360.

[9] M. A. Tonellat, *Les Théories Unitaires de l'Électromagnétisme e de la Gravitation* (Gauthier Villars, Paris, 1965), p. 301; see, also E. Schroedinger, Ref. 1, p. 110.

Stochasticity in Yang–Mills theory

Javier Villarroel
Departamento de Fisica Teórica, Facultad de Ciencias, Universidad de Salamanca, 37008-Salamanca, Spain

(Received 19 October 1987; accepted for publication 4 May 1988)

Some stochastic properties of a subsystem of Yang–Mills classical mechanics are studied which do not fulfill automatically Gauss' law, a possibility not considered before. However, significant differences with other subsystems already studied are not found for this system.

I. INTRODUCTION

Consider the SU(2) Yang–Mills classical field theory when we make the assumption that the potentials depend only on time, i.e., $A_i^a = A_i^a(t)$ $(i = 1,2,3,\ldots, a = 1,2,3)$ and take the gauge $A_0^a = 0$. The motion equations are then

$$\ddot{A}_i^a + g^2(A_j^b A_i^a A_j^b - A_j^a A_j^b A_i^b) = 0, \tag{1}$$

and we have to add three constraint equations besides,

$$g^2 \varepsilon^{abc} A_i^b \dot{A}_i^c = 0 \tag{2}$$

(where, of course, the dots denote temporal derivatives, while g is the coupling constant).

Although the former is really a greatly simplifying ansatz, the resulting system we have obtained, also termed Yang–Mills classical mechanics[1] (YMCM), is still complex enough in order to solve it, so one must consider[1–4] a still more restrictive ansatz, namely, to suppose

$$A_3^a = 0, \tag{3}$$

$$A_i^a = \frac{O_i^a f^{(a)}(t)}{g} \quad (i = 1,2 \text{ and no summation over } a),$$

where O_i^a is a constant orthogonal matrix, i.e.,

$$O_i^a O_i^b = \delta^{ab}, \tag{4}$$

and then studying the resulting two-dimensional mechanical system. One of the pleasant features of this ansatz is that it automatically fulfills the constraint equations (2) so we do not have to take care of them anymore. It only remains then to solve the nontrivial equations

$$\ddot{f}^{(1)} + f^{(1)}(f^{(2)})^2 = \ddot{f}^{(2)} + f^{(2)}(f^{(1)})^2 = 0. \tag{5}$$

The physical interest of this system was motivated by studying[1] the nature of the vacuum fluctuations and the asymptotic states of non-Abelian gauge fields. But the information we could obtain from it concerning the integrability properties of the theory is perhaps more important. In fact, the results about the subsystem described by Eq. (5) of the above-mentioned references point out that it is probably a strongly stochastic system, in the sense that it does not have enough first integrals so that, according to Liouville's theorem,[5] it can be integrated. If that were the case, it would rule out the expectatives concerning the integrability of the Yang–Mills equations that the work in the self-dual sector had raised. We deem that this question is so important that it demands further study, so in this work we have tried to extend the above-mentioned analysis to a different subsystem. The fact that makes our system significantly different from the other already considered is that it does not fulfill Gauss' law [Eq. (2)] from the beginning. We are interested in such a situation because, as has been shown,[6] Gauss' law is a nonholonomous constraint (in the sense that this expression is usually understood in classical mechanics) and we think that it is interesting to see what physical modifications bring this fact into the picture. We shall concentrate then on the dynamical system described by Eqs. (1) and (2); we will not make the further ansatz (3) and (4), but instead will make a different one, to be described later.

II. LAGRANGE MULTIPLIERS

Equations (1) and (2) describe a dynamical system with nine degrees of freedom subject to some nonholonomous constraints. The usual way to work in classical mechanics with such constraints is to add them to the motion equations by means of some Lagrange multipliers. In our case, as we are working in the gauge $A_0^a = 0$ we hope that, if we carry out this modification, we will find at the end that the Lagrange multipliers are identically zero. But it could be an interesting exercise trying to check it. If we do such a thing we obtain that Eq. (1) now gets modified in the following form:

$$\ddot{A}_i^a = g^2(A_j^b A_i^a A_j^b - A_j^a A_j^b A_i^b) = \sum_{c=1}^{3} g^2 \varepsilon^{cba} A_i^b \lambda^c, \quad i = 1,2,3, \quad a = 1,2,3, \tag{6}$$

where the λ's are unknown Lagrange multipliers. To this set we also have to add Eq. (2). Now our problem is to obtain the value of the λ's. This is an easy problem but as the calculations are long we shall only sketch the way to obtain the result. To do it, solve, using Eq. (2), for three of the $\dot{A}_i^a$ as a function of the remaining A_j^b, $\dot{A}_j^b$. Then, substitute them into (2) and demand consistency. This way, we obtain a linear system of equations for the λ, which turn out to be homogeneous, so its determinant must be identically zero. But this determinant Δ is

$$\Delta \equiv \left(\sum_{i=1}^{3}\sum_{a=1}^{3}(A_i^a)^2\right)\left(\sum_{i<j}\sum_{a<b}(A_i^a A_j^b - A_j^a A_i^b)^2\right) = 0. \tag{7}$$

The second term can be shown to be the potential that appears when we formulate YMCM in the Hamiltonian formalism. Clearly none of these terms vanishes, unless we

 0022-2488/88/092132-05$02.50

compel our theory to be trivial, so we conclude that all the multipliers are zero, as we hoped.

III. GEODESIC FLOW

We now proceed to define the subsystem we are going to study. Together with Eqs. (1) and (2) we shall suppose that

$$A_i^3 = A_3^a = 0 \quad (i = 1,2,3,\ a = 1,2,3), \tag{8}$$

so we have a dynamical system with four degrees of freedom which does not verify automatically Gauss' law. In fact, Eq. (1) under the former restrictions are now

$$\begin{aligned} &\ddot{x} + g^2 w(xw - yz) = 0, \\ &\ddot{y} - g^2 z(xw - yz) = 0, \\ &\ddot{z} - g^2 y(xw - yz) = 0, \\ &\ddot{w} + g^2 x(xw - yz) = 0, \end{aligned} \tag{9}$$

and Gauss' law [Eq. (2)] stands as

$$g^2(y\dot{x} - x\dot{y} + w\dot{z} - z\dot{w}) = 0, \tag{10}$$

where we are using the notation $x = A_1^1$, $y = A_1^2$,

$$z = A_2^1, \quad w = A_2^2,$$

and we can derive these equations from the Hamiltonian

$$H = \tfrac{1}{2}(\dot{x}^2 + \dot{y}^2 + \dot{z}^2 + \dot{w}^2) + (g^2/2)(xw - yz)^2 \equiv E. \tag{11}$$

For this system, we shall study its properties from different points of view. First, in this section we shall consider it as a geodesic flow.

It is well known that any Hamiltonian flow can locally be viewed as the geodesic flow on a Riemannian manifold. Moreover, the Lobatchewsky–Hadamard theorem[7] suggests to us that if the scalar curvature of the manifold turns out to be negative, then the flow is unstable, with strongly stochastic properties, corresponding to a Kolmogorov or K system.[8]

There are two reasons why this criterion does not always let us decide if a given Hamiltonian system is integrable: first, the latter statement is only a necessary but not sufficient condition for the system to be integrable; and second, the association of our system with the geodesic flow on a Riemannian manifold is not valid at the boundary of the manifold, given by $E = V(x)$ (V being the potential), where the metric becomes singular so that association is valid only locally. But global properties are often the origin of chaos (remember, for example, Sinai's billiards), so that, although we have a system whose associated Riemannian manifold is of positive curvature, we can not conclude that it is integrable.

Anyway we shall carry out this test for our system. In our case the metric is

$$g_{ij} = (E - (g^2/2)(xw - yz)^2)\delta_{ij}. \tag{12}$$

After some calculation we obtain for the scalar curvature $\mathbf{R}$,

$$\mathbf{R} = \frac{12(x^2 + y^2 + z^2 + w^2)E}{(E - (g^2/2)(xw - yz)^2)^3} \geq 0, \tag{13}$$

so the necessary condition of nonintegrability is not fulfilled, but as remarked before, we cannot conclude from it that the system is integrable.

IV. PAINLEVÉ'S ANALYSIS

A useful criterion in order to decide about the integrability of a system of differential equations is the so-called Painlevé's analysis, which, roughly speaking, consists in the study of the singularities of the system. We shall suppose that the reader knows how this procedure works (see, for example, Ref. 9) and we shall apply it to our case. Then suppose

$$\begin{aligned} &x = x_0(t - t_0)^P, \quad z = z_0(t - t_0)^P, \\ &y = y_0(t - t_0)^P, \quad w = w_0(t - t_0)^P, \end{aligned} \tag{14}$$

where x_0, y_0, z_0, w_0, t_0, and P are constants to be determined and $\mathrm{Re}(p) < 0$. When substituting into Eqs. (9) we find $p = -1$ together with

$$z_0 - y_0, \quad x_0 = w_0,$$

or

$$z_0 = y_0, \quad x_0 = -w_0,$$

and

$$x_0^2 + y_0^2 + 2 = 0. \tag{15}$$

Since $P\ (= -1)$ is an integer, Painlevé's condition is fulfilled, and so we have to go on with the test. We now look for the resonances. Inserting into Eqs. (9) the ansatz

$$\begin{aligned} &x = x_0/(t - t_0) + h_1(t - t_0)^{r-1}, \\ &y = S/(t - t_0) + h_2(t - t_0)^{r-1}, \\ &z = -S/(t - t_0) + h_3(t - t_0)^{r-1}, \\ &w = x_0/(t - t_0) + h_4(t - t_0)^{r-1}, \end{aligned} \tag{16}$$

where $S \equiv i(2 + x_0^2)^{1/2}$, h_i are constants, and r is the value of the resonance to be determined. We obtain that the following equation must hold:

$$\begin{vmatrix} (r-1)(r-2) + x_0^2 & sx_0 & -sx_0 & x_0^2 - 2 \\ sx_0 & (r-1)(r-2) + s^2 & 2 - x_0^2 & sx_0 \\ -sx_0 & 2 - x_0^2 & (r-1)(r-2) + x_0^2 & -sx_0 \\ x_0^2 - 2 & sx_0 & -sx_0 & (r-1)(r-2) + x_0^2 \end{vmatrix} = 0. \tag{17}$$

This gives us an equation of eighth degree in r, from where we obtain its possible values. They are $r = 0$, $r = 1$ corresponding to the arbitrariness of x_0, to, respectively, $r = 4$ reflecting the fact that the Hamiltonian is a homogeneous function of fourth degree, $r = 3$, and finally $r = [3 \pm (7)^{1/2}i]/2$. The last two roots are not integers, but

instead they are nonrational numbers that indicate the presence of an algebraic branch point, a signal of nonintegrability. But as Painlevé analysis is, in general, coordinate dependent, we still have to check that these branch points cannot be removed, i.e., that our system is analytic at t_0. We discover this to be the case, so the nonrational roots correspond to essential branch points. This fact, according to Painlevé's conjecture, points out that our system is nonintegrable, confirming the analysis of other authors.[10]

V. KOLMOGOROV–ARNOLD–MOSER (KAM) THEOREM

A richer situation occurs when we study the proper ties of an SU(2)-gauge system with spontaneous symmetry breaking. Such a situation is described by the Hamiltonian (see Refs. 2 and 4)

$$H = H_{\mathrm{YM}} + \tfrac{1}{4}g^2K^2A_i^aA_i^a, \tag{18}$$

where H_{YM} is the Yang–Mills Hamiltonian and K is a constant.

For this system we shall try to extend the results of Refs. 2 and 4 to our subsystem, i.e., that described by conditions (8). Our Hamiltonian is then

$$\begin{aligned} H = {} & \tfrac{1}{2}(P_x^2 + P_y^2 + P_z^2 + P_w^2) \\ & + (g^2/2)(xw - yz)^2 \\ & + g^2(K^2/4)(x^2 + y^2 + z^2 + w^2). \end{aligned} \tag{19}$$

This system is characterized by one parameter $q = g^2K^4/E$ (E being the energy) in such a way that when q goes to zero we recover our original Hamiltonian, which looks to be nonintegrable, while when q grows the dominant term is the last one, which is just an harmonic oscillator potential that is integrable. Then that parameter controls the integrability of the system. We shall then suppose that we are in the sector $q \gg$, so we have an integrable harmonic oscillator for the potential, for which action-angle variables do exist, perturbed by the Yang–Mills potential. These action-angle variables $J_i(t)$, $\phi_i(t)$, $i = 1,2,3,4$, are defined by

$$\begin{aligned} x &= (2\sqrt{2}\,J_1/gk)^{1/2}\sin\phi_1, & P_x &= (\sqrt{2}\,gkJ_1)^{1/2}\cos\phi_1, \\ y &= (2\sqrt{2}\,J_2/gk)^{1/2}\sin\phi_2, & P_y &= (\sqrt{2}\,gkJ_2)^{1/2}\cos\phi_2, \\ z &= (2\sqrt{2}\,J_3/gk)^{1/2}\sin\phi_3, & P_z &= (\sqrt{2}\,gkJ_3)^{1/2}\cos\phi_3, \\ w &= (2\sqrt{2}\,J_4/gk)^{1/2}\sin\phi_4, & P_w &= (\sqrt{2}\,gkJ_4)^{1/2}\cos\phi_4, \end{aligned} \tag{20}$$

in terms of which the resulting Hamiltonian is

$$\begin{aligned} H &= H_0 + H_1, \\ H_0 &= (gk/\sqrt{2})(J_1 + J_2 + J_3 + J_4), \\ H_1 &= (4/k^2)((J_1J_4)^{1/2}\sin\phi_1\sin\phi_4 \\ &\quad - (J_2J_3)^{1/2}\sin\phi_2\sin\phi_3)^2. \end{aligned} \tag{21}$$

At $q \to \infty$ the Hamiltonian is clearly integrable, so phase space is foliated by invariant tori, but the interesting problem here is what will happen when introducing the perturbation H_1. The main result in this respect is the[11–13] KAM theorem, which asserts that if the Hamiltonian H_0 is nondegenerate, i.e., Det $(\partial^2H_0/\partial J_\alpha\,\partial J_\beta) \neq 0$, then most of the invariant tori (the nonresonant ones) will survive, provided the perturbation H_1 is small enough. In our case the situation is slightly different; because H_0 is degenerate, we have to use perturbation theory for degenerate systems, as was stated by Arnold.[14,15] Using the (new) canonical transformation

$$\begin{aligned} I_1 &= J_1 + J_2 + J_3 + J_4, & \psi_1 &= \phi_1, \\ I_2 &= J_2, & \psi_2 &= \phi_2 - \phi_1, \\ I_3 &= J_3, & \psi_3 &= \phi_3 - \phi_1, \\ I_4 &= J_4, & \psi_4 &= \phi_4 - \phi_1, \end{aligned} \tag{22}$$

we obtain

$$\begin{aligned} H = {} & (gk/\sqrt{2})I_1 + (4/k^2)(I_4(I_1 - I_2 - I_3 - I_4)^{1/2} \\ & \times\sin\psi_1\sin(\psi_1 + \psi_4) \\ & - (I_2I_3)^{1/2}\sin(\psi_1 + \psi_2)\sin(\psi_3 + \psi_4))^{1/2}. \end{aligned} \tag{23}$$

It is now clear that I_1, ψ_1 are the fast variables, while the remaining variables are the slow ones. What happens physically is that our system is carrying out quick spins around a "guide center," say. Arnold's idea is to eliminate these rotations by means of temporal average, in order to see only the essential motion. Accordingly, we shall take

$$\begin{aligned} \overline{H}_1 \equiv {} & \int_0^{2\pi} H_1\,d\psi_1 \\ = {} & \frac{1}{k^2}\Big(\Big(1 + \frac{1}{2}\cos 2\psi_4\Big)(I_1 - I_2 - I_3 - I_4)I_4 \\ & + I_2I_3\Big(1 + \frac{1}{2}\cos 2(\psi_2 - \psi_3) \\ & + (I_2I_3I_4(I_1 - I_2 - I_3 - I_4))\Big)^{1/2} \\ & \times(\cos(\psi_2 - \psi_4)\cos\psi_3 + \cos(\psi_2 - \psi_3)\cos\psi_4 \\ & + \cos(\psi_3 - \psi_4)\cos\psi_2)\Big) \end{aligned} \tag{24}$$

as our basic Hamiltonian. Clearly we now have a constant of motion, which is I_1. The reason for studying the solutions of this Hamiltonian is that, according to Arnold's results in perturbation theory for degenerate systems, most of the trajectories for the perturbed Hamiltonian $H_0 + H_1$ will be close to those solutions provided $\overline{H}_1$ is small enough. Among these that are particularly interesting are those corresponding to

$$\dot{I}_i = \dot{\psi}_i = 0, \quad i = 2,3,4,$$

which describe periodic solutions and are of great importance to describe the tori's destruction. These equations read

$$\begin{aligned} & I_2I_3\sin 2(\psi_2 - \psi_3) + (J_1I_2I_3I_4)^{1/2}(\sin(\psi_2 - \psi_4)\cos\psi_3 \\ & \quad + \sin(\psi_2 - \psi_3)\cos\psi_4 + \sin(\psi_3 - \psi_4)\sin\psi_2) = 0, \\ & I_2I_3\sin 2(\psi_2 - \psi_3) + (J_1I_2I_3I_4)^{1/2}(-\cos(\psi_2 - \psi_4)\sin\psi_3 \\ & \quad + \sin(\psi_2 - \psi_4)\cos\psi_4 - \sin(\psi_3 - \psi_4)\cos\psi_2) = 0, \\ & J_1I_4\sin 2\psi_4 - (J_1I_2I_3I_4)^{1/2}(\sin(\psi_2 - \psi_4)\cos\psi_3 \\ & \quad - \cos(\psi_2 - \psi_3)\sin\psi_4 + \sin(\psi_3 - \psi_4)\cos\psi_2) = 0, \\ & I_3(1 + \tfrac{1}{2}\cos 2\psi_3) - I_4(1 + \tfrac{1}{2}\cos 2\psi_4) \\ & \quad + (I_3I_4/J_1I_2)^{1/2}(\cos(\psi_2 - \psi_4)\cos\psi_3 \\ & \quad + \cos(\psi_2 - \psi_3)\cos\psi_4 \\ & \quad + \cos(\psi_3 - \psi_4)\cos\psi_2)(J_1 - I_2) = 0, \end{aligned} \tag{25}$$

$$I_2(1+\tfrac{1}{2}\cos 2(\psi_2-\psi_3))-I_4(1+\tfrac{1}{2}\cos 2\psi_4)$$
$$+(I_2I_4/J_1I_3)^{1/2}(\cos(\psi_2-\psi_4)\cos\psi_3$$
$$+\cos(\psi_2-\psi_3)\cos\psi_4$$
$$+\cos(\psi_3-\psi_4)\cos\psi_2)(J_1-I_3)=0,$$

where J_1 should be understood to be $J_1=I_1-I_2-I_3-I_4$. To these equations we have to add also the nontrivial constraint (10), which in these coordinates reads

$$(J_1I_2)^{1/2}\sin\psi_2+(I_3I_4)^{1/2}\sin(\psi_4-\psi_3)=0, \tag{26}$$

and we easily obtain from these equations the relation

$$\psi_2+\psi_3-\psi_4=n\pi, \tag{27}$$

with n being an integer number. Hereafter the calculation turns out to be slightly tedious so we shall not describe it, but only give the results. We also obtain

$$I_2=I_3=I_4=I_1/4. \tag{28}$$

In order to describe the possible solutions, we have to give, besides the general relations (27) and (28), the values of ψ_2, ψ_3, and ψ_4. They are

(a) $n=2p+1$, and

$-a_1$: $\psi_2=(m+h)\pi/2$, $\psi_3=(m-h)\pi/2$,

$-a_2$: $\psi_2=m\pi$, ψ_3 arbitrary,

$-a_3$: ψ_2 arbitrary, $\psi_3=m\pi$.

(b) $n=2p$, and

$-b_1$: $2=m\pi$, $3=h\pi$,

$-b_2$: $2=2m\pi$, $3=(2h+1)(\pi/2)$,

with p, m, and h being arbitrary integers.

There are then five possible periodic solutions for our system. The way to obtain them is, first, solve the periodic equations and then, impose Gauss' law, discarding those solutions that do not fulfill it. Moreover, we have not considered the case when two of the I_i, say I_2 and I_3, are set equal to zero, but we shall have more to say about it.

A. Stability of the solutions

By studying the Hessian at the solutions we have already obtained, we can discern their stability properties. Then we obtain the following.

(a) *a_l with m,h integers with the same parity:* Then the solutions are stable and $\overline{H}_1=0$. In the diagram of Poincaré's map it corresponds to an elliptic point with index 1, that is (more graphically), to a point in the center of an "island," and therefore stable. (Actually, the study of the stability of this particular case needs consideration on the quartic form d^4H_1, with d standing for differentiation.)

a_l with h,m integers with different parity: The solution turns out now to be unstable, and, $\overline{H}_1=4I^2$. It corresponds to an hyperbolic point in Poincaré's map, that is to say, to a point of self-intersection with the separatrix.

These considerations would also apply to the remaining cases for which we obtain the following.

a_2 (a_3): The solution is stable independent of the value of ψ_3 (ψ_2, respectively), corresponding to a minimum of the potential $\overline{H}_1=0$.

(b) *b_1:* It corresponds to an unstable solution, which in fact is a maximum of the potential as $\overline{H}_1=6I^2$.

b_2: It is a local minimum, with $\overline{H}_1=2I^2$ and so it is stable.

B. The case $J_2=J_3$. The temporal evolution

When doing $J_2=J_3=0$, the averaged Hamiltonian

$$H\equiv E=\frac{gk}{\sqrt{2}}(J_1+J_4)+\frac{J_1J_4}{k^2}\left(1+\frac{1}{2}\cos 2(\phi_1-\phi_4)\right) \tag{29}$$

can be completely integrated so we shall consider this case with special interest. Introducing new coordinates, defined by the canonical transformation

$$I_1=J_1+J_4,\quad \psi_1=\phi_1,$$
$$I_4=J_4,\quad \psi_4=\phi_4-\phi_1, \tag{30}$$

it is clear that besides the energy E, we have another constant of motion, namely $I_1\equiv C$. The motion equations are then

$$\dot{I}_4=[I_4(C-I_4)/k^2]\sin 2\psi_4,$$
$$\dot{\psi}_4=[(C-2I_4)/k^2](1+\tfrac{1}{2}\cos 2\psi_4), \tag{31}$$

and we obtain that the motion is governed by the following elliptic integral:

$$\int dI_4\left(I_4^4+2CI_4^3-\left(\frac{3C^2+8d}{3}\right)I_4^2+\frac{8dCI_4}{3}-\frac{4d^2}{3}\right)^{-1/2}=\frac{(3)^{1/2}}{k^2}(t-t_0), \tag{32}$$

where $d\equiv(E-cgk/2^{1/2})K^2$.

The roots of the polynomial that appear in the denominator are

$$r_1=\frac{C+(C^2-8d/3)^{1/2}}{2},\quad r_2=\frac{C+(C^2-8d)^{1/2}}{2},$$
$$r_3=\frac{C-(C^2-8d)^{1/2}}{2},\quad r_4=\frac{C-(C^2-8d/3)^{1/2}}{2}. \tag{33}$$

The properties of the motion strongly depend on how many complex and real roots we have, so our first task is to determine their respective number.

To do it, it is necessary to study the possible values that the constant d can take. Using that I_4 is restricted to be between $0\le I_4\le C$, and translating this condition (together with $0\le\psi_4\le 2\pi$), we obtain that, in terms of d, ψ_4, the accessible regions of the phase space are

(a) $0\le d\le C^2/8$ (and also $0\le\psi\le 2\pi$).

In this case all the roots are real and the motion is limited to the zone $r_4\le I_4\le r_3$. We have clearly bound states and the solution is given by

$$I_4(t) = \frac{(C - r_4)m\,\mathrm{sn}^2((3(r_4 - r_3))^{1/2}((t - {}^t r_4)/2k^2), m) - r_4}{m\,\mathrm{sn}^2(3(r_4 - r_3)^{1/2}(1/2k^2)(t - tr_4), m) - 1}, \tag{34}$$

where

$$m \equiv \frac{r_1 - r_4 + r_3 - r_2}{r_1 - r_4 + r_2 - r_3}, \quad t_{r_4} \equiv t(I_4 = r_4) \tag{35}$$

and sn being an elliptic Jacobi function.

(b) The other possibility is that $C^2/8 \le d \le 3C^2/8$. Moreover, in this case ψ_4 is restricted to the range

$$8d/C^2 - 2 \le \cos 2\psi_4 \le 1.$$

Then clearly r_1, r_4 are real roots, while r_2, r_3 are complex. The motion takes place between $r_4 \le I_4 \le C$ and the explicit solution is now given by

$$I_4(t) = \frac{(H(t) - 1)r_1 B - (H(t) + 1)r_4 A}{(H(t) - 1)B - (H(t) + 1)A}, \tag{36}$$

where

$$\begin{aligned} A^2 &\equiv C^2 + 16d/3 - 2C(C^2 - 8d/3)^{1/2}/4, \\ B^2 &\equiv C^2 + 16d/3 + 2c(C^2 - 8d/3)^{1/2}/4, \\ H(t) &\equiv \mathrm{cn}((3AB)^{1/2}/k^2)(t - t_{r_4}), m), \\ m &\equiv \frac{1}{2} + \frac{C^2 - 32d/3}{8AB}. \end{aligned} \tag{37}$$

These expressions together with Eq. (29) completely determine the motion.

VI. DISCUSSION

We have generalized the study of Yang–Mills classical mechanics to a case where Gauss' law is not automatically fulfilled so it has to be added by means of a Lagrange multiplier and retained up the the end. Although this fact clearly differentiates the considered system, the results obtained do not differ significantly from others of previous works.

Again the curvature of the associated Riemannian manifold does not let us draw the conclusion that our system is not integrable, but the subsequent Painlevé analysis points it out to be the case.

We then make use of the KAM theory for degenerate systems, obtaining the temporal evolution in a somewhat simplified case, and only the periodic solutions in the general one, besides their stability properties. Some questions arise now. What will happen when the perturbation grows? The KAM theorem does not help us to answer this question.

Perhaps an even more interesting question is, if stochasticity would disappear if we only restrict the study to the manifold of solutions that go to zero when t goes to infinity, that is to say, to fields fulfilling $A_i^a(t) \underset{t\to\infty}{\to} 0$, which is the physically relevant case. In the case of self-dual equations we know them to be integrable, even in the general case $A_i^a(x)$, $i = 0,1,2,3$.[16,17]

Finally, we can ask ourselves about the influence of the classical stochastic properties in the subsequent quantized theory. It is not clear at the moment if chaos is a property also shared by quantum theories. To answer this question some authors are trying to develop semiclassical quantization methods or using other techniques which could characterize quantum chaos. It has been pointed out that chaos would be manifested at the quantum level in some irregular spacing of the energy levels.[18–20] On the other hand, the Schrödinger equations being linear, it would be surprising that its solutions could manifest stochastic properties. Finally, if they do, it remains to be known if any important physical property is due to this cause.

[1] S. G. Matinyan, G. K. Savvidi, and Ter-Arutyunyan-Savvidi, Sov. Phys. JETP **53**, 421 (1981).
[2] S. G. Matinyan, G. K. Savvidi, and Ter-Arutyunyan-Savvidi, Pis'ma Zh. Eksp. Teor. Fiz. **34**, 613 (1981).
[3] H. M. Asatryan and G. K. Savvidy, Phys. Lett. A **99**, 290 (1983).
[4] G. K. Savvidy, Nucl. Phys. B **246**, 302 (1984).
[5] V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, Berlin 1978).
[6] J. Villarroel, Phys. Lett. B **181**, 321 (1986).
[7] V. I. Arnold and A. Avez, *Ergodic Problems of Classical Mechanics* (Benjamin, New York, 1968).
[8] P. Halmos, *Lectures on Ergodic Theory* (Chelsea, New York, 1958).
[9] M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715 (1980).
[10] W. Steeb, C. Villet, and A. Kunick, J. Phys. A: Math. Gen. **18**, 3269 (1985).
[11] V. Arnold, Russian Math. Surveys **18**, 85 (1963).
[12] A. Kolmogorov, Dokl. Akad. Nauk. USSR **98**, 527 (1954).
[13] J. Moser, Nachr. Akad. Wiss, Gottingen Math. Phys. Kl.a **1962**.
[14] V. Arnold, Dokl. Akad. Nauk. USSR **138**, 13 (1961).
[15] V. Arnold, Dokl. Akad. Nauk. USSR **145**, 487 (1962).
[16] R. S. Ward, Phys. Lett. A **61**, 81 (1977).
[17] M. R. Atiyah and R. S. Ward, Commun. Math. Phys. **55**, 117 (1977).
[18] M. Berry and M. Tabor, Proc. R. Soc. London Ser. A **356**, 375 (1977).
[19] M. Berry, "Semiclassical mechanics of regular and irregular motion," in *Les Houches Lectures 1981* (North-Holland, Amsterdam, 1983).
[20] I. Percival, Adv. Chem. Phys. **36**, 1 (1977).