Beltrami algebra and symmetry of the Beltrami equation on Riemann surfaces

Han-Ying Guo
CCAST (World Laboratory), Beijing 100080, People's Republic of China and Institute of Theoretical Physics, Academia Sinica, Beijing 100080, People's Republic of China

Jian-Min Shen
Institute of Theoretical Physics, Academia Sinica, Beijing 100080, People's Republic of China

Shi-Kun Wang
CCAST (World Laboratory), Beijing, 100080, People's Republic of China and Institute of Applied Mathematics, Academia Sinica, Beijing 100080, People's Republic of China

Kai-Wen Xu
Department of Physics, Zhejiang University, Hangzhou, People's Republic of China

(Received 28 March 1989; accepted for publication 27 June 1990)

It is shown that the Beltrami equation has an infinite-dimensional symmetry, namely the Beltrami algebra, on its solution spaces. The Beltrami algebra with central extension and its supersymmetric version are explicitly found.

I. INTRODUCTION

Recently, much attention has been paid to study the deformation of complex structure of Riemann surfaces, since that subject is relevant to the problem of constructing operator formulations of strings[1,2] and the conformal field theory on higher genus Riemann surfaces,[3,4] as well as $2d$ quantum gravity.[5] The ones usually concentrated on are either the Virasoro algebra and the Krichever–Novikov algebra, some generators which act on the moduli space of Riemann surfaces, or the semidirect summation of Virasoro algebra and $U(1)$ Kac–Moody algebra occurring in the induced $2d$ gravity.

It is well known, however, that the quasiconformal transformations characterized by the Beltrami differentials and the Beltrami equation deform the complex structures on Riemann surfaces as well. Therefore, it is of interest to see whether there exist some particular symmetries relevant to the quasiconformal transformations and whether these symmetries play a certain role in the deformations of the complex structures on Riemann surfaces. In this paper, we concentrate upon the first problem. We show that there does exist certain infinite-dimensional symmetries in the solution space of the Beltrami equation, named the Beltrami algebra, and explore the properties of the algebra in detail for some simplest cases.

In Sec. II, we prove the existence of the Beltrami algebra and give the explicit construction of the algebra and its central extension form as well as the relevant Kac–Moody algebra for the simplest case with the coefficient of the Beltrami differential being constant. In Sec. III, we give a supersymmetry extension of the Beltrami algebra. Finally, in Sec. IV we give some brief discussions.

II. BELTRAMI EQUATION AND SYMMETRY OF ITS SOLUTIONS

As is well known, a smooth quasiconformal transformation on a Riemann surface is a homeomorphism of surface M satisfying the following condition,[6,7]

$$\bar{\partial} f - \mu\, \partial f = 0, \tag{1}$$

where μ is a Beltrami differential of the $(-1,1)$ type. In terms of a local coordinate patch μ can be expressed in the form:

$$\mu = \mu(z)\frac{d\bar{z}}{dz} = \mu(w)\frac{d\bar{w}}{dw}, \qquad \mu(z) = \mu(w)\frac{\partial \bar{w}}{\partial \bar{z}}\frac{\partial z}{\partial w}, \tag{2}$$

and there is a positive constant K such that

$$|\mu(z)| \leqslant (K-1)/(K+1).$$

Introduce the Beltrami operator $\Delta = \bar{\partial} - \mu\partial$ or

$$\Delta_{\bar{z}} = \partial_{\bar{z}} - \mu(z)\partial_z \tag{3}$$

in terms of the z coordinate and

$$\Delta_{\bar{w}} = \partial_{\bar{w}} - \mu(w)\partial_w \tag{4}$$

in terms of the w coordinate. We see that

$$\Delta_{\bar{z}} = \frac{\partial \bar{w}}{\partial \bar{z}}\Delta_{\bar{w}}. \tag{5}$$

Then the Beltrami equation (1) is simply rewritten as

$$\Delta f = 0 \tag{6}$$

or

$$\Delta_{\bar{z}} f = \frac{\partial \bar{w}}{\partial \bar{z}}\Delta_{\bar{w}} f = 0. \tag{7}$$

Now we want to find a set of symmetric operators $\{L\}$ that map the solutions into solutions of the Beltrami equation.

Let the field L be an operator acting on the solution space of the Beltrami equation. If the operator L satisfies

$$[L,\Delta] = R\Delta, \tag{8}$$

then the operator L is called a symmetric operator of the Beltrami equation (1). Clearly, the symmetric operators map the solutions into solutions of Eq. (1). The local expression of the relation (8) can be read in z and w coordinate patches, respectively,

 0022-2488/90/112543-05$03.00

$$[L_z,\Delta_{\bar z}] = R(z)\Delta_{\bar z},$$
$$[L_w,\Delta_{\bar w}] = R(w)\Delta_{\bar w}, \tag{9}$$

where

$$\begin{aligned} L_z &= X(z)\partial_z \oplus Y(z)\partial_{\bar z} \oplus M(z) \\ &= X(w)\partial_w \oplus Y(w)\partial_{\bar w} \oplus M(z(w)) \end{aligned} \tag{10}$$

are direct sums of the vector fields and scale field to be determined and $X(z)$ and $Y(z)$ are local coefficients of the vector fields in the z coordinate patch. Under coordinate transformation, $R(z)$ is related to $R(w)$ by the consistent condition

$$R(z) = R(w) - Y(w)\frac{\partial}{\partial \bar z}\left(\frac{\partial \bar z}{\partial \bar w}\right). \tag{11}$$

After some algebraic calculations, a set of equations can be derived from (9) and (10) as follows:

$$\begin{aligned} &\Delta_{\bar z} Y(z) = -R(z), \\ &\Delta_{\bar z} X(z) = -(X(z)\partial_z + Y(z)\partial_{\bar z} - R(z))\mu(z), \\ &\Delta_{\bar z} M(z) = 0. \end{aligned} \tag{12}$$

To prove that the symmetric operators L form an infinite-dimensional Lie algebra, we have the following theorem.

Theorem: If $L^{(i)}$, $i = 1,2$ are two symmetric operators then $L = [L^{(1)},L^{(2)}]$ is also a symmetric operator of Beltrami equation.

Proof: Let $L^{(i)}$, $i = 1,2$, be expressed as

$$\begin{aligned} L_z^{(i)} &= H_i(z) \oplus M_i(z) \\ &= X_i(z)\frac{\partial}{\partial z} \oplus Y_i(z)\frac{\partial}{\partial \bar z} \oplus M_i(z) \\ &= X_i(w)\frac{\partial}{\partial w} \oplus Y_i(w)\frac{\partial}{\partial \bar w} \oplus M_i(z(w)) \end{aligned} \tag{13}$$

and

$$[L_z^{(1)},\Delta_{\bar z}] = R_1(z)\Delta_{\bar z},$$
$$[L_z^{(2)},\Delta_{\bar z}] = R_2(z)\Delta_{\bar z}.$$

Then

$$\begin{aligned} L_z &= [L_z^{(1)},L_z^{(2)}] \\ &= [H_1(z),H_2(z)] \oplus \{L_z^{(1)}M_2(z) - L_z^{(2)}M_1(z)\} \\ &= X(z)\frac{\partial}{\partial z} \oplus Y(z)\frac{\partial}{\partial \bar z} \oplus M(z), \end{aligned} \tag{14}$$

where

$$\begin{aligned} X(z) &= X_1(z)\frac{\partial}{\partial z}(X_2(z)) - X_2(z)\frac{\partial}{\partial z}(X_1(z)) \\ &\quad + Y_1(z)\frac{\partial}{\partial \bar z}(X_2(z)) - Y_2(z)\frac{\partial}{\partial \bar z}(X_1(z)), \\ Y(z) &= Y_1(z)\frac{\partial}{\partial \bar z}(Y_2(z)) - Y_2(z)\frac{\partial}{\partial \bar z}(Y_1(z)) \\ &\quad + X_1(z)\frac{\partial}{\partial z}(Y_2(z)) - X_2(z)\frac{\partial}{\partial z}(Y_1(z)), \end{aligned}$$

$$M(z) = L_z^{(1)}M_2(z) - L_z^{(2)}M_1(z).$$

Since $L^{(i)}$ $(i = 1,2)$ map solutions as solutions of the Beltrami equation, we have

$$\Delta_{\bar z} M(z) = 0.$$

Therefore,

$$\begin{aligned} [L_z,\Delta_{\bar z}] &= [[H_1(z),H_2(z)],\Delta_{\bar z}] \oplus [M(z),\Delta_{\bar z}] \\ &= \left\{[H_1(z),[H_2(z),\Delta_{\bar z}]] - [H_2(z),[H_1(z),\Delta_{\bar z}]]\right\} \oplus \Delta_{\bar z} M(z) \\ &= [H_1(z),R_2(z)\Delta_{\bar z}] - [H_2(z),R_1(z)\Delta_{\bar z}] \\ &= (H_1(z)R_2(z))\Delta_{\bar z} + R_2(z)[H_1(z),\Delta_{\bar z}] \\ &\quad - (H_2(z)R_1(z))\Delta_{\bar z} - R_1(z)[H_2(z),\Delta_{\bar z}] \\ &= (H_1(z)R_2(z) - H_2(z)R_1(z))\Delta_{\bar z}. \end{aligned}$$

That is,

$$[L_z,\Delta_{\bar z}] = R(z)\Delta_{\bar z}, \quad R(z) = H_1(z)R_2(z) - H_2(z)R_1(z),$$

and

$$\begin{aligned} R(z) &= H_1(w)\left\{R_2(w) - Y_2(w)\frac{\partial}{\partial \bar z}\left(\frac{\partial \bar z}{\partial \bar w}\right)\right\} - H_2(w)\left\{R_1(w) - Y_1(w)\frac{\partial}{\partial \bar z}\left(\frac{\partial \bar z}{\partial \bar w}\right)\right\} \\ &= H_1(w)R_2(w) - H_2(w)R_1(w) + H_2(w)\left\{Y_1(w)\frac{\partial}{\partial \bar z}\left(\frac{\partial \bar z}{\partial \bar w}\right)\right\} - H_1(w)\left\{Y_2(w)\frac{\partial}{\partial \bar z}\left(\frac{\partial \bar z}{\partial \bar w}\right)\right\} \\ &= R(w) + X_2(w)\frac{\partial}{\partial w}\left[Y_1(w)\frac{\partial}{\partial \bar z}\left(\frac{\partial \bar z}{\partial \bar w}\right)\right] + Y_2(w)\frac{\partial}{\partial \bar w}\left[Y_1(w)\frac{\partial}{\partial \bar z}\left(\frac{\partial \bar z}{\partial \bar w}\right)\right] \end{aligned}$$

$$-X_1(w)\frac{\partial}{\partial w}\left[Y_2(w)\frac{\partial}{\partial z}\left(\frac{\partial z}{\partial w}\right)\right]-Y_1(w)\frac{\partial}{\partial \bar{w}}\left[Y_2(w)\frac{\partial}{\partial z}\left(\frac{\partial z}{\partial \bar{w}}\right)\right]$$
$$=R(w)-\left[Y_1(w)\frac{\partial}{\partial \bar{w}}(Y_2(w))-Y_2(w)\frac{\partial}{\partial \bar{w}}(Y_1(w))+X_1(w)\frac{\partial}{\partial w}(Y_2(w))-X_2(w)\frac{\partial}{\partial w}(Y_1(w))\right]\frac{\partial}{\partial z}\left(\frac{\partial z}{\partial \bar{w}}\right).$$

This implies that

$$R(z)=R(w)-Y(w)\frac{\partial}{\partial z}\left(\frac{\partial z}{\partial \bar{w}}\right)$$

satisfies (11), the consistent relation. The proof is now completed.

This theorem indicates that the symmetric operators L do possess a Lie algebra structure. Call it Beltrami algebra. Now, let us consider some simple cases.

A. The symmetry of $\partial_{\bar{z}} f=0$

Let $\mu=0$, then the Beltrami equation is reduced to the Cauchy–Riemann equation. Equations (12) become

$$\partial_{\bar{z}}X=0,\qquad \partial_{\bar{z}}Y=-R,\quad \partial_{\bar{z}}M=0. \tag{15}$$

For the simplest case, we consider the Cauchy–Riemann equation with its domain on $S^2\backslash\{0,\infty\}$. We may take $L_m=z^m(\partial/\partial z)$, a subset of all symmetry operators. The subset forms the so-called Virasoro algebra without central extension. We may also get two systems of solutions of (15) such that one of them is $M=Y=0, X=z^mI$, and another is $M=\lambda^a z^m, X=Y=0$, where I is the unit matrix and λ^a are the generators of some finite-dimensional Lie algebra. Then we find that the symmetry of solutions space of the Cauchy–Riemann equation defined on $S^2\backslash\{0,\infty\}$ is just the semidirect summation of the Virasoro algebra and the Kac–Moody algebra without central extension, i.e.,

$$[L_m,L_n]=(m-n)L_{m+n},\qquad [I^a_m,I^b_n]=C^{ab}_c I^c_{m+n},\qquad [L_m,I^a_n]=nI^a_{m+n}, \tag{16}$$

where C^{ab}_c is the structure constant of the Lie algebra $[\lambda^a,\lambda^b]=C^{ab}_c\lambda^c$, if we set $L_m=z^{m+1}\partial_z$, $I^a_m=\lambda^a z^m$.

The symmetry of global solutions of the Cauchy–Riemann equation on punctured higher genus Riemann surface, such as K–N algebra,[2] can also be constructed by means of this systematic method.

B. The symmetry of Beltrami equation with $\mu=$const

For this case of torus,, the coefficients of the symmetric operator L satisfy

$$(\partial_{\bar{z}}-\mu\partial_z)X=R\mu,\qquad (\partial_{\bar{z}}-\mu\partial_z)Y=-R,\qquad (\partial_{\bar{z}}-\mu\partial_z)M=0. \tag{17}$$

Following the previous procedure we may obtain two subsets $\{L^1_n\}$ and $\{L^2_n\}$ of symmetric operators that act on the solution space of Beltrami equation

$$\partial_{\bar{z}}f-\mu\partial_z f=0 \text{ with } \mu=\text{const}. \tag{18}$$

They have the following forms:

$$L^1_n=-(z+\mu\bar{z})^{n+1}\partial_z,\qquad L^2_n=-(z+\mu\bar{z})^{n+1}\mu^{-1}\partial_{\bar{z}} \tag{19}$$

and form an infinite-dimensional subalgebra of the Beltrami algebra. We denote it by $\mathscr{B}$ and still call it the Beltrami algebra.[8] The algebraic relations are

$$[L^1_n,L^1_m]=(m-n)L^1_{m+n},\qquad [L^2_n,L^2_m]=(m-n)L^2_{m+n},\qquad [L^1_n,L^2_m]=(m+1)L^2_{m+n}-(n+1)L^1_{m+n}. \tag{20}$$

It is easy to find that this algebra contains two Virasoro algebras as its subalgebras. We may consider the Beltrami as a semidirect summation of these two Virasoro algebras in the sense of $[L^1_n,L^2_m]\neq 0$ $(n,m\neq -1)$.

The Beltrami algebra also has a finite-dimensional subalgebra as follows:

$$[L^1_{-1},L^1_0]=-L^1_{-1},\qquad [L^1_{-1},L^2_0]=-L^2_{-1},\qquad [L^1_{-1},L^2_{-1}]=0,\qquad [L^2_{-1},L^2_0]=-L^2_{-1},\qquad [L^1_0,L^2_0]=-L^2_0+L^1_0,\qquad [L^1_0,L^2_{-1}]=L^1_{-1}. \tag{21}$$

The generators of this subalgebra may perform the simplest quasiconformal transformation, i.e., locally deforming a circle into an ellipse.

It should be mentioned here that a set of vector fields, which are the linear combination of L^1_n and L^2_n, maps all nonzero solutions of Eq. (17) into zero. Such a vector field has the form

$$T_n=L^1_n-L^2_n, \tag{22}$$

with $T_nF[(z+\mu\bar{z})]=0$, F analytically depending on $(z+\mu\bar{z})$. Hence, the Beltrami algebra $\mathscr{B}$ also contains an Abel subalgebra $\mathscr{H}$:

$$[T_n,T_m]=0. \tag{23}$$

Because for any $L^i_n\in\mathscr{B}$, $T_m\in\mathscr{H}$, there is

$$[L^i_n,T_m]=(m+1)T_{m+1}\in\mathscr{H},\quad i=1,2, \tag{24}$$

It follows that $\mathscr{H}$ is an ideal of $\mathscr{B}$. From Eq. (24) we also find that

$$[L^i_n,T_{-1}]=0,\quad i=1,2. \tag{25}$$

It is easy to obtain the corresponding semidirect summation of Beltrami algebra and Kac–Moody algebra for the case at hand, which has a local form

$$M^a_n=\lambda^a\otimes(z+\mu\bar{z})^n, \tag{26}$$

where λ^a are generators of a finite-dimensional Lie algebra

$$[\lambda^a,\lambda^b]=f_c^{ab}\lambda^c. \tag{27}$$

This algebra $\mathscr{K}$ reads

$$[M^a_m,M^b_n]=f_c^{ab}M^c_{m+n} \tag{28}$$

$$[L^i_n,M^a_m]=mM^a_{m+n},\quad i=1,2. \tag{29}$$

Finally, we discuss the central extensions of both the Beltrami algebra and the semidirect summation algebra $\mathscr{K}$.

Let $\hat{L}^i_n=L^i_n+C^i_\alpha L^{*\alpha}$, $1\leqslant\alpha\leqslant N$, be the generators carrying a central term L^* with

$$[L^{*\alpha},L^{*\beta}]=0,\quad [L^{*\alpha},L^i_m]=0. \tag{30}$$

Without loss of generality we may ignore the index α, then the central extensions of the Beltrami algebra are given by

$$\begin{aligned}
&[\hat{L}^i_n,\hat{L}^i_m]=(m-n)\hat{L}^i_{m+n}+C^{ii}(m\cdot n)L^*,\quad i=1,2,\\
&[\hat{L}^1_n,\hat{L}^2_m]=(m+1)\hat{L}^2_{m+n}\\
&\qquad -(n+1)\hat{L}^1_{m+n}+C^{12}(m\cdot n)L^*.
\end{aligned} \tag{31}$$

From the Jacobi identity we get the following algebraic equations of structure constants C^{ij}, $i,j=1,2$:

$$\begin{aligned}
&(n-m)C^{ii}(m-n,l)+(l-n)C^{ii}(n+l,m)\\
&\quad+(m-l)C^{ii}(l+m,n)=0,\quad i=1,2,\\
&(n+1)C^{12}(m+n,l)-(n+1)C^{12}(n+l,m)\\
&\quad+(m-l)C^{21}(l+m,n)-(m+1)C^{22}(m+n,l)\\
&\quad+(l+1)C^{22}(n+l,m)=0,\\
&(n+1)C^{21}(m+n,l)-(n+1)C^{21}(n+l,m)\\
&\quad+(m-l)C^{12}(l+m,n)-(m+1)C^{11}(m+n,l)\\
&\quad+(l+1)C^{11}(n+l,m)=0.
\end{aligned} \tag{32}$$

After some tedious calculations we find

$$\begin{aligned}
&[\hat{L}^i_n,\hat{L}^i_m]=(m-n)\hat{L}^i_{m+n}\\
&\qquad+c(m^3-m)\delta_{m+n,0}L^*,\quad i=1,2,\\
&[\hat{L}^1_n,\hat{L}^2_m]=(m+1)\hat{L}^2_{m+n}-(n+1)\hat{L}^1_{m+n}\\
&\qquad+c(m^3-m)\delta_{m+n,0}L^*.
\end{aligned} \tag{33}$$

Along the previous line we also get the semidirect summation algebra $\mathscr{K}$ with central extension:

$$\begin{aligned}
&[\hat{M}^a_m,\hat{M}^b_n]=f_c^{ab}\hat{M}^c_{m+n}+km\delta_{m+n,0},\\
&[\hat{L}^i_n,\hat{M}^a_m]=m\hat{M}^a_{m+n},\quad i=1,2.
\end{aligned} \tag{34}$$

III. SUPER-BELTRAMI ALGEBRA ON SUPER-RIEMANN SURFACES

Let $\hat{f}$ be a super-quasiconformal mapping $\hat{f}:\hat{S}\to\hat{S}'$. The local coordinate (z,θ) on the compact super-Riemann surface $\hat{S}$ is related to the local coordinate (w,ϕ) on $\hat{S}'$ by the super-Beltrami equation:[9]

$$\begin{aligned}
&\partial_{\bar{z}}w+\phi\partial_{\bar{z}}\phi)=\mu(\partial_z w+\phi\partial_z\phi),\\
&-\partial_\theta w+\phi\partial_\theta\phi=\nu(\partial_z w+\phi\partial_z\phi),
\end{aligned} \tag{35}$$

where $\mu(z,\bar{z},\theta)$ and $\nu(z,\bar{z},\theta)$ are the complex-valued super-Beltrami coefficients. In the case of $\mu=0$, $\nu=\theta$, Eqs. (35) are reduced to the super-Cauchy–Riemann equations.[10]

Using the way described in Sec. II one may discuss the symmetry of solution space of the super-Beltrami equation and get the super version of the Beltrami algebra. For that, expanding $w(z,\theta),\phi(z,\theta),\mu(z,\theta)$, and $\nu(z,\theta)$ with respect to θ we express Eqs. (35) in the form

$$\begin{aligned}
&\partial_{\bar{z}}w^0+\phi^1\partial_{\bar{z}}\phi^1=\mu^0(\partial_z w^0+\phi^1\partial_z\phi^1),\\
&\partial_{\bar{z}}w^1+\phi^0\partial_{\bar{z}}\phi^1-\phi^1\partial_{\bar{z}}\phi^0\\
&\quad=\mu^0(\partial_z w^1+\phi^0\partial_z\phi^1-\phi^1\partial_z\phi^0)+\mu^1(\partial_z w^0+\phi^1\partial_z\phi^1),\\
&-w^1+\phi^1\phi^0=\nu^1(\partial_z w^0+\phi^1\partial_z\phi^1),\\
&(\phi^0)^2=\nu^1(-\partial_z w^1+\phi^1\partial_z\phi^0-\phi^0\partial_z\phi^1)\\
&\qquad+\nu^0(\partial_z w^0+\phi^1\partial\phi^1),
\end{aligned} \tag{36}$$

where $w=w^0+\theta w^1$, $\phi=\phi^1+\theta\phi^0$, $\mu=\mu^0+\theta\mu^1$, and $\nu=\nu^1+\theta\nu^0$.

Take $\mu^1=0$, $\mu=\mu^0=\text{const}$ and appropriate value ν^1 and ν^0, we then find a set of special solutions of Eqs. (36):

$$\begin{aligned}
w^0&=\sum_n\epsilon_n(z+\mu\bar{z})^{n+1},\\
w^1&=\sum_r d_r\eta(z+\mu\bar{z})^{(r+1)/2},\\
\phi^0&=\sum_n\epsilon_n(z+\mu\bar{z})^n,\\
\phi^1&=\sum_r d_r\eta(z+\mu\bar{z})^{(r+1)/2},
\end{aligned} \tag{37}$$

where η is a Grassmann constant, $n,m\in\mathbb{Z}$, r is taken to be the half-integer in Neveu–Schwartz sector and the integer in Ramond sector.

Denote

$$\begin{aligned}
w^0_n&=(z+\mu\bar{z})^{n+1},\\
w^1_r&=\theta\eta(z+\mu\bar{z})^{(r+1)/2},\\
\phi^0_n&=(z+\mu\bar{z})^n\theta,\\
\phi^1_r&=\eta(z+\mu\bar{z})^{(r+1)/2}.
\end{aligned} \tag{38}$$

We find there exist two sets of vectors $\{L^i_n\}$, $i=1,2$ which map $\{w^0_n,\phi^0_n\}$, $\{w^1_n,\phi^1_r\}$ onto themselves, and another two sets of vectors $\{G^i_r\}$, $i=1,2$ which map $\{w^0_n,\phi^0_n\}$ to $\{w^1_r,\phi^1_r\}$ and vice versa. These vectors are given by

$$\begin{aligned}
L^1_n&=(z+\mu\bar{z})^{n+1}\frac{\partial}{\partial z}+\frac{1}{2}(n+1)(z+\mu\bar{z})^n\theta\frac{\partial}{\partial\theta},\\
L^2_n&=(z+\mu\bar{z})^{n+1}\frac{\partial}{\mu\partial\bar{z}}+\frac{1}{2}(n+1)(z+\mu\bar{z})^n\theta\frac{\partial}{\partial\theta},\\
G^1_r&=(z+\mu\bar{z})^{r+1/2}\left(\frac{\partial}{\partial\theta}-\theta\frac{\partial}{\partial z}\right),\\
G^2_r&=(z+\mu\bar{z})^{r+1/2}\left(\frac{\partial}{\partial\theta}-\theta\frac{\partial}{\mu\partial\bar{z}}\right).
\end{aligned} \tag{39}$$

We find

$$\begin{aligned}
&[L^i_n,L^j_m]=(m+1)L^i_{m+n}-(n+1)L^i_{m+n},\\
&[L^i_n,G^j_r]=(r+n/2+1)G^j_{n+r}-(n+1)G^i_{n+r},\\
&[G^i_r,G^j_s]=-L^i_{r+s}-L^j_{r+s},\quad i,j=1,2.
\end{aligned} \tag{40}$$

This is what we called super-Beltrami algebra which contains the Neveu–Schwartz algebra and the Ramond algebra as subalgebras. It is easy to see that (40) is the supersymme-

tric generalization of (20).

The central extensions of the algebra (40) can be uniquely obtained, from solving the cocycle condition of the central extension super-Beltrami algebra:

$$
\begin{aligned}
[\hat{L}^i_n,\hat{L}^j_m] &= (m+1)\hat{L}^j_{m+n} - (n+1)\hat{L}^i_{m+n} \\
&\quad + (c/8)(m^3-m)\delta_{m+n,0}, \\
[\hat{L}^i_n,\hat{G}^j_r] &= (r+n/2+1)\hat{G}^j(n+r) - (n+1)\hat{G}^i_{n+r}, \\
[\hat{G}^i_r,\hat{G}^j_s] &= -\hat{L}^i_{r+s} - \hat{L}^j_{r+s} - \frac{c}{2}(r^2-\tfrac{1}{4})\delta_{r+s,0},
\end{aligned}
\tag{41}
$$

where c is a constant.

IV. DISCUSSIONS

We have proved that the Beltrami equation has the infinite-dimensional symmetry on its solution space for Riemann surfaces and, in particular, have explicitly constructed the Beltrami algebra for the case of $\mu = \text{const}$, its central extension form, its semidirect summation with Kac–Moody algebra, as well as its supersymmetric versions. We see that the special homeomorphic solutions in the solution space of Beltrami equation describe quasiconformal transformations of Riemann surfaces. Therefore, on the one hand, the infinite-dimensional symmetry on the solution space implies that there may also exist various infinite-dimensional symmetries to characterize the generic quasiconformal transformations as well as their supersymmetric versions. On the other hand, it would be of interest to find out such symmetries relevant to quasiconformal transformations and to see their relations with the deformation of complex structures of Riemann surfaces. From that it may be possible to deeply discuss the symmetry of moduli space and the physical applications of the symmetries. In fact, from the result $c = 28$ in our BRST formalism of the Beltrami algebra,[11] we find the quasiconformal symmetry actually plays an important role in the quantum theory of $2d$ gravity. The subjects related to them will be explained elsewhere.[12]

ACKNOWLEDGMENT

This work was supported in part by The National Natural Science Foundation of China.

[1] C. Vafa, Phys. Lett. B **190**, 47 (1987); L. Alavarez-Gaume, C. Gomez, G. Moore, and C. Vafa, "Strings in the operator formalism," CERN preprint TH 4883/87.

[2] I. M. Krichever and S. P. Novikov, Funk. Anal. Pril. **21**(2), 46 (1987); **21**(4), 47 (1987).

[3] D. Friedan, in *Proceedings of the Santa Barbara Workshop on Unified String Theories*, edited by M. Green and D. Gross (World Scientific, Singapore, 1986).

[4] T. Eguchi and H. Oguri, Nucl. Phys. B **282**, 308 (1987).

[5] A. M. Polyakov, Mod. Phys. Lett. A **11**, 893 (1987); V. G. Knizhnik, A. M. Polyakov, and A. B. Zamolodchikov, *ibid.* A **3**, 819 (1988).

[6] S. L. Krushkal, *Quasiconformal Mappings and Riemann Surfaces* (Winston, Washington, DC, 1979).

[7] Li Zong, *Quasiconformal Mapping and its Applications in the Theory of Riemann Surfaces* (Science, Beijing, 1988).

[8] H. Y. Guo, J. M. Shen, S. K. Wang, and K. W. Xu, Chinese Phys. Lett. **2**, 53 (1989).

[9] L. Crane and J. M. Rabin, Comm. Math. Phys. **113**, 601 (1988).

[10] D. Friedan, E. Martinec, and S. Shenker, Nucl. Phys. B **271**, 83 (1986).

[11] H. Y. Guo, J. M. Shen, S. K. Wang, and K. W. Xu, Commun. Theor. Phys. **14**, 123 (1990).

[12] H. Y. Guo, S. K. Wang, Z. H. Wang, and K. Wu, Commun. Theor. Phys. **14**, 99 (1990).

Three-dimensional quantum groups from contractions of $SU(2)_q$

Enrico Celeghini
Dipartimento di Fisica dell'Università and I.N.F.N. Sezione di Firenze, Largo E. Fermi, 2 I 50125 Firenze, Italy

Riccardo Giachetti
Dipartimento di Matematica, Università di Cagliari and I.N.F.N. Sezione di Firenze, Firenze, Italy

Emanuele Sorace and Marco Tarlini
I.N.F.N. Sezione di Firenze, Firenze, Italy

(Received 7 February 1990; accepted for publication 16 May 1990)

Contractions of Lie algebras and of their representations are generalized to define new quantum groups. An explicit and complete exposition is made for the one-dimensional Heisenberg $H(1)_q$ and the two-dimensional Euclidean quantum group $E(2)_q$ obtained by contracting $SU(2)_q$.

I. INTRODUCTION

In recent years the concept of a quantum group has extensively entered mathematical and physical literature. In theoretical physics, quantum groups have received huge interest due to their connection with the quantum Yang–Baxter equation[1] that plays a deep role in many problems of current research, as, for instance, exactly soluble problems of statistical mechanics,[2] integrable quantum field theories,[3] and factorized S matrices.[4] The mathematical nature of quantum groups has been progressively clarified and put in connection with the largely developed theory of Hopf algebras[5] and universal enveloping algebras of Lie algebras:[6] Indeed quantum groups can be regarded as deformations of the latter structures with suitable "classical limit" conditions. As a consequence, many results concerning the structure and the representations of Lie groups have been extended to the quantum case as well and the q deformation $U_q(g)$ of the universal enveloping algebra has been defined for any simple Lie algebra g.

Probably the most studied case of the quantum group is obviously $U_q(\mathrm{su}(2)) = \mathrm{SU}(2)_q$,—see, e.g., Refs. 5 and 7—, whose generators in the Chevalley basis are indicated by the usual notation $\{J_3, J_\pm\}$ and for which the commutation relation of the su(2) Lie algebra are deformed into

$$[J_3, J_\pm] = \pm J_\pm ,$$
$$[J_+, J_-] = [2J_3]_q \equiv \frac{\mathrm{sh}(zJ_3)}{\mathrm{sh}(z/2)}, \quad z = \log q. \tag{1.1}$$

This group has also been discussed using a generalization of Schwinger's method to the theory of angular momentum, thus introducing different q-deformed analogs of the raising and lowering operators of the harmonic oscillator.[8,9] On the other hand, this approach has been considered only from a technical point of view: an independent significance as a quantum group obtained by q deforming the (nonsemisimple) Heisenberg algebra $h(1)$ generated by the three elements $\{A, A^+, H\}$ with H central and $[A, A^+] = H$, has not been looked for. Since we think that nonsemisimple quantum groups deserve an appropriate consideration in themselves, we shall present another treatment of the problem, which, on the one side can be generalized to a large class of algebras, while on the other provides a mathematically handy formalism. Thus, instead of pursuing the Schwinger's approach to the angular momentum, we prefer to look for a generalization of the contraction technique. In order to be as concrete as possible, we shall consider the simplest cases, namely the three-dimensional ones.

The purpose of this paper is indeed to discuss the q deformation of other possible three-dimensional Lie algebras that exist besides su(2): one of them is the just mentioned $h(1)$, while the other is the Lie algebra $e(2)$ of the rigid motions of the plane $\mathbf{R}^2$. The discussion we are going to present is based on the widely known method of contractions of Lie algebras[10] and on its generalization for construction of the corresponding representations.[11] The central idea is that, since the two nonsemisimple algebras $h(1)$ and $e(2)$ are obtained by suitable contractions of su(2), quantum groups $H(1)_q$ and $E(2)_q$ can be coherently defined by an extension of the contraction procedure, which must also be effective on the parameter q determining the quantum deformation. We thus have to deal with a commutative diagram:

$$\begin{array}{ccc} & \overset{\varepsilon\to 0}{} & \\ \mathrm{SQG} & \to & \mathrm{CQG} \\ q\to 1 \quad \downarrow & & \downarrow \quad q\to 1, \\ \mathrm{SG} & \underset{\varepsilon\to 0}{\to} & \mathrm{CG} \end{array}$$

where SQG and CQG, SG and CG, respectively, mean simple and contracted quantum group, simple and contracted group.

II. THE HEISENBERG QUANTUM GROUP AND ITS REPRESENTATIONS

According to the scheme previously proposed, we define the following transformation on the generators and on the parameter $z = \log q$ whose vanishing limit reproduces the classical su(2):

$$'(A, A^+, H, w) = \mathrm{diag}\{\varepsilon^{1/2}, \varepsilon^{1/2}, 2\varepsilon, \varepsilon^{-1}\}'(J_+, J_-, J_3, z). \tag{2.1}$$

Taking the limit $\varepsilon \to 0$ we get

 0022-2488/90/112548-04$03.00

$$[H,A]=0,\quad [H,A^+]=0,\quad [A,A^+]=\frac{\mathrm{sh}(wH/2)}{w/2}. \tag{2.2}$$

We remark that the scaling $w=\varepsilon^{-1}z$ is determined by the requirement that the contraction procedure makes sense and does not directly give the Lie algebra $h(1)$. Indeed, by the above choice, the hyperbolic sine structure is preserved for the commutator of A and A^+: obviously $h(1)$ is again recovered in the limit $w\to 0$. We now show that the relations (2.2) are actually consistent.

Proposition A: Equations (2.2) define a quantum group that will be called the Heisenberg quantum group $\mathrm{H}(1)_q$.

Proof: We use the contraction procedure to determine coproduct, counit, and antipode for (2.2) and we show that the Hopf algebra requirements are satisfied.[12] From the definition of z, from the coproduct of $\mathrm{SU}(2)_q$ (Ref. 5):

$$\Delta J_+ = e^{-zJ_3/2}\otimes J_+ + J_+\otimes e^{zJ_3/2}$$

and using the transformation (2.1) we find

$$\Delta A = e^{-wH/4}\otimes A + A\otimes e^{wH/4},$$

$$\Delta A^+ = e^{-wH/4}\otimes A^+ + A^+\otimes e^{wH/4}.$$

For H we have obviously

$$\Delta H = 1\otimes H + H\otimes 1,$$

and a straightforward calculation shows that

$$[\Delta A,\Delta A^+]=\frac{\mathrm{sh}(w\,\Delta H/2)}{w/2}.$$

For the counit the situation is even simpler, since

$$\epsilon(A)=\epsilon(\epsilon^{1/2}J_+)=0$$

and similarly

$$\epsilon(A^+)=\epsilon(H)=0.$$

Now, it is immediate to verify that

$$(\epsilon\otimes\mathrm{id})\Delta A=(\mathrm{id}\otimes\epsilon)\Delta A=A. \tag{2.3}$$

Moreover,

$$(\mathrm{id}\otimes\Delta)\Delta A = e^{-wH/4}\otimes e^{-wH/4}\otimes A + e^{-wH/4}\otimes A\otimes e^{wH/4} + A\otimes e^{wH/4}\otimes e^{wH/4},$$

and from the analogous computation we find that

$$(\mathrm{id}\otimes\Delta)\Delta A=(\Delta\otimes\mathrm{id})\Delta A. \tag{2.4}$$

Finally, for the antipode γ, we see that $\gamma(J_3)=-J_3$ implies $\gamma(H)=-H$ while

$$\gamma(A)=\epsilon^{1/2}\gamma(J_+)=-\epsilon^{1/2}e^{zJ_3/2}J_+e^{-zJ_3/2} \xrightarrow[\varepsilon\to 0]{} -e^{wH/4}Ae^{-wH/4}=-A,$$

and analogously $\gamma(A^+)=-A^+$. Denoting by m the product in the enveloping algebra a final calculation shows that

$$m(\mathrm{id}\otimes\gamma)\Delta A=m(\gamma\otimes\mathrm{id})\Delta A=\epsilon(A)1, \tag{2.5}$$

each of the three terms being zero. As the relations (2.3)–(2.5) also hold for the generators A^+ and H, the proof is complete.

Let us now extend to the representations of $\mathrm{SU}(2)_q$ the contraction procedure. We have:[8]

$$J_\pm|j,m\rangle=([j\pm m]_q[j\pm m+1]_q)^{1/2}|j,m\pm 1\rangle, \tag{2.6}$$

$$J_3|j,m\rangle=m|j,m\rangle,$$

where, as usual, $[\alpha]_q=\mathrm{sh}(z\alpha/2)/\mathrm{sh}(z/2)$. Performing the change of variables

$$\begin{pmatrix}p\\n\\w\end{pmatrix}=\begin{pmatrix}2\varepsilon & 0 & 0\\ 1 & -1 & 0\\ 0 & 0 & \varepsilon^{-1}\end{pmatrix}\begin{pmatrix}j\\m\\z\end{pmatrix},$$

Eqs. (2.6), with $s=\exp\{\varepsilon w\}$, can be written in the form

$$J_-|\varepsilon^{-1}p,n\rangle=([n]_s[\varepsilon^{-1}p-n+1]_s)^{1/2}|\varepsilon^{-1}p,n-1\rangle,$$

$$J_+|\varepsilon^{-1}p,n\rangle=([n+1]_s[\varepsilon^{-1}p-n]_s)^{1/2}|\varepsilon^{-1}p,n+1\rangle,$$

$$J_3|\varepsilon^{-1}p,n\rangle=(\varepsilon^{-1}p/2-n)|\varepsilon^{-1},n\rangle.$$

Observing that, for $\varepsilon\to 0$,

$$[n]_s=\frac{\mathrm{sh}(\varepsilon wn/2)}{\mathrm{sh}(\varepsilon w/2)}\to n$$

while

$$[\varepsilon^{-1}p-n]_s=\frac{\mathrm{sh}(w(p-\varepsilon n)/2)}{\mathrm{sh}(\varepsilon w/2)}\to\varepsilon^{-1}\frac{\mathrm{sh}(wp/2)}{w/2}$$

and following step by step the procedure of Ref. 11, we finally find

$$A|p,n\rangle=\left(n\frac{\mathrm{sh}(wp/2)}{w/2}\right)^{1/2}|p,n-1\rangle,$$

$$A^+|p,n\rangle=\left((n+1)\frac{\mathrm{sh}(wp/2)}{w/2}\right)^{1/2}|p,n+1\rangle, \tag{2.7}$$

$$H|p,n\rangle=p|p,n\rangle.$$

In the same spirit we find that the Casimir reads

$$C=(4/w^2)\mathrm{sh}^2(wH/2).$$

The limit $z\to 0$ of (2.6) reproduces obviously the representations of su(2) as well as the limit $w\to 0$ of (2.7) gives the corresponding representations of $h(1)$.

III. THE EUCLIDEAN QUANTUM GROUP

The same method developed in Sec. II will be now used to define a quantum deformation $\mathrm{E}(2)_q$ of the Euclidean group of the plane. We find it more convenient to perform a change of basis and substitute the generators $\{J_+,J_-,J_3\}$ by $\{J_1,J_2,J_3\}$, where

$$J_1=\tfrac{1}{2}(J_++J_-),\quad J_2=-(i/2)(J_+-J_-).$$

It is then straightforward to see that, in this basis, Eqs. (1.1) take the form

$$[J_1,J_2]=i\frac{\mathrm{sh}(zJ_3)}{2\,\mathrm{sh}(z/2)},\quad [J_2,J_3]=iJ_1,\quad [J_3,J_1]=iJ_2. \tag{3.1}$$

As in the case of the Heisenberg group, here again we

have one possible contraction that completely eliminates the quantum structure. Indeed, if the generators J_k are scaled according to

$$J_i \to \varepsilon^{-1} J_i, \quad i = 1,2, \quad J_3 \to J_3,$$

then the contracted relations are simply

$$[J_1, J_2] = 0, \quad [J_2, J_3] = iJ_1, \quad [J_3, J_1] = iJ_2,$$

namely those of the Lie algebra $e(2)$. Let us, instead, consider the scaling

$$^{t}(P_y, J, P_x, w) = \mathrm{diag}\{\varepsilon, 1, \varepsilon, \varepsilon^{-1}\}\,{}^{t}(J_1, J_2, J_3, z).$$

Taking now the limit $\epsilon \to 0$, a nontrivial quantum behavior is maintained and the contracted relations read

$$[P_x, P_y] = 0, \quad [J, P_x] = iP_y,$$
$$[J, P_y] = -(i/w)\mathrm{sh}(wP_x). \tag{3.2}$$

We are again in position to prove the analog of Proposition A.

Proposition B: Equations (3.2) define a quantum group that will be called the Euclidean quantum group in dimension two $E(2)_q$.

Proof: Since again we have to show that the Hopf algebra properties are satisfied, we simply give the form of coproduct, counit, and antipode and omit the straightforward calculations. These are

$$\Delta J = e^{-wP_x/2} \otimes J + J \otimes e^{wP_x/2}, \quad \Delta P_x = 1 \otimes P_x + P_x \otimes 1,$$
$$\Delta P_y = e^{-wP_x/2} \otimes P_y + P_y \otimes e^{wP_x/2},$$

for the coproduct;

$$\epsilon(J) = \epsilon(P_x) = \epsilon(P_y) = 0,$$

for the counit;

$$\gamma(J) = -J + (i/2)wP_y,$$
$$\gamma(P_x) = -P_x, \quad \gamma(P_y) = -P_y,$$

for the antipode.

It is interesting to determine the Casimir of $\mathrm{E}(2)_q$. We recall that the quantum group $\mathrm{SU}(2)_q$ has two different expressions for the Casimir according to the fact that the representation has integer or half-odd j (Ref. 8), that are, respectively,

$$J_- J_+ + [J_3]_q [J_3 + 1]_q$$

and

$$J_- J_+ + ([J_3 + \tfrac{1}{2}]_q)^2.$$

It is easy to be seen that the contraction of both of them gives the same result and we shall explicitly work out the integer j case. We have

$$J_- J_+ + [J_3]_q [J_3 + 1]_q$$
$$= J_1^2 + J_2^2 - \frac{\mathrm{sh}(zJ_3)}{2\,\mathrm{sh}(z/2)} + \frac{\mathrm{sh}(zJ_3/2)\mathrm{sh}(z(J_3+1)/2)}{\mathrm{sh}^2(z/2)}$$
$$= \varepsilon^{-2} P_y^2 + J^2 - \frac{\mathrm{sh}(wP_x)}{2\,\mathrm{sh}(\varepsilon w/2)} + \frac{\mathrm{sh}(wP_x/2)\mathrm{sh}(w(P_x+\varepsilon)/2)}{\mathrm{sh}^2(\varepsilon w/2)}.$$

Renormalizing and taking the limit $\epsilon \to 0$ we get the Casimir operator for $\mathrm{E}(2)_q$,

$$C = P_y^2 + (4/w^2)\mathrm{sh}^2(wP_x/2),$$

whose limit for vanishing w reproduces the classical $P_x^2 + P_y^2$.

Again following Ref. 11, we can obtain the representations of $\mathrm{E}(2)_q$, that, in the momentum basis, are very similar to the usual ones because here also we diagonalize P_x and P_y. Of course the Casimir is conserved, which implies that we are not performing a rotation on a circle but on a less symmetrical curve of equation $C = \mathrm{const}$. This is reflected in the expression for J that reads

$$J = iP_y \frac{\partial}{\partial P_x} - \frac{i}{w}\mathrm{sh}(wP_x)\frac{\partial}{\partial P_y}.$$

IV. CONCLUSIONS

We want to conclude this paper by making some observations both on the procedure and on the meaning of the concepts that have been introduced.

From a technical point of view we want again to stress the fact that, as in the standard case, there is not a unique way to perform a contraction. In the case of quantum groups, the q parameter must be included in the scaling of the variables and care has to be taken in order not to fall directly into the classical situation. In the envisaged examples, the scaling of the q parameter was, in fact, determined by this requirement. We do not expect this is still holding in more complicated situations.

More general considerations are suggested by the $\mathrm{E}(2)_q$ case. Indeed we see that situations with apparently lower symmetry, as is the case of the curve $C = \mathrm{const}$, can be treated by concepts and tools analogous to completely symmetrical cases. In our knowledge this is an almost unique technique where it is possible to dismiss with continuity part of the symmetry while keeping all the other features unchanged. In this perspective we are working to investigate the quantum group $\mathrm{E}(3,1)_q$ which, in our opinion, is the right instrument to describe the (classical!) first corrections to the flat metric due to the presence of a gravitational field.

[1] M. Jimbo, Int. J. Mod. Phys. A **4**, 3759 (1989).
[2] R. J. Baxter, *Exactly Solved Models in Statistical Mechanics* (Academic, New York, 1982).
[3] P. P. Kulish and E. K. Sklyanin, Zap. Nauchn. Sem. LOMI **95**, 129 (1980).
[4] A. B. Zamolodchikov and Al. B. Zamolodchikov, Ann. Phys. (NY) **120**, 253 (1979).
[5] V. G. Drinfeld, *Proceedings of The International Congress of Mathematicians Berkeley, CA*, edited by A. M. Gleason (AMS, Providence, RI, 1986), p. 798.
[6] Dixmier *Algèbres Enveloppantes* (Gauthier-Villars, Paris, 1974).
[7] S. Woronowicz, Publ. RIMS **23**, 117 (1987).
[8] L. C. Biedenharn, J. Phys. A: Math. Gen. **22**, L873 (1989); "A q-boson

realization of the quantum group $SU_q(2)$ and the theory of q-tensor operators," invited paper at Clausthal Summer Workshop on Mathematical Physics (1989).
[9] A. J. Macfarlane, J. Phys. A: Math. Gen. **22**, 4581 (1989).
[10] R. Gilmore, *Lie Groups, Lie Algebras and Some of Their Applications* (Wiley, New York, 1974).
[11] E. Celeghini and M. Tarlini, Nuovo Cimento B **61**, 265 (1981); B **65**, 172 (1981); B **68**, 133 (1982).
[12] E. Abe, *Hopf Algebras*, Cambridge Tracts in Math. No. 74 (Cambridge U. P., Cambridge, 1980).

Classification of all star irreps of gl($m|n$)

M. D. Gould and R. B. Zhang
Department of Mathematics, University of Queensland, Brisbane, Australia

(Received 11 October 1989; accepted for publication 23 May 1990)

All the finite-dimensional star irreps of gl($m|n$) are classified in terms of their highest weights, thereby completing the classification of all finite-dimensional star irreps of the basic classical Lie superalgebras. The lowest weights of such irreps are determined explicitly and it is shown that the contravariant and covariant tensor irreps of gl($m|n$) are star irreps of type (1) and (2), respectively. This explains the applicability of the Young diagram method to these two classes of representations. However, it is shown that one-parameter families of typical star irreps occur that are intrinsically different from the tensorial irreps.

I. INTRODUCTION

In the first two papers of this series, we classified the star and grade star irreps of gl($m|1$) (Ref. 1) and $C(n)$ (Ref. 2). In this paper, we complete the classification of all finite-dimensional star irreps for the type I basic classical Lie superalgebras by obtaining the star irreps of gl($m|n$). In view of the fact that type II basic classical Lie superalgebras[3] do not admit star irreps,[4,5] this completes the classification of the star irreps of all the basic classical Lie superalgebras.

Star and grade star irreps are the natural generalizations of Hermitian representations of ordinary Lie algebras to the case of Lie superalgebras. They were first introduced by Scheunert *et al.*,[4] who demonstrated that all simple basic classical Lie superalgebras admit at most two types of star and grade star irreps. The corresponding modules of such irreps possess a positive definite metric, thus are most likely to have physical applications, where unitarity is a basic requirement.

Quite apart from their possible physical applications, these irreps are mathematically interesting as well, especially the star ones. As pointed out in Ref. 4, the tensor product of two star irreps belonging to one type reduces completely into star irreps of the same type, and a star irrep branches only into star irreps of a given (canonical) subsuperalgebra. These properties, certainly not shared by other irreps in general, make the star irreps behave very much like irreducible representations of ordinary Lie algebras. Thus it is plausible that the classical techniques used in studying the representation theory of Lie algebras will be applicable to this class of irreps as well. For example, the Young diagram technique[6] proves to be a very powerful tool in dealing with the so-called contravariant and covariant tensor irreps.[6-9] Its applicability for the type I basic classical Lie superalgebras is actually based on the fact that the contravariant and covariant tensor irreps are star irreps of the two different types.

Although some star and grade star irreps of certain superalgebras were studied individually in the past,[5,10] a systematic classification of such irreps is still lacking. Therefore, our aim in this series of papers is to carry out the classification of all the finite dimensional star and grade star irreps of the basic classical Lie superalgebras. The present article, together with the previous papers,[1,2] accomplishes the classification of the star irreps. In the next paper, the grade star irreps will be completely classified.

The main results of this paper are summarized in Theorems 1–4, which specify the necessary and sufficient conditions on the highest weight of a given finite-dimensional irreducible gl($m|n$) module in order for it to be star. The lowest weights of the atypical star irreps are determined in Propositions 5 and 6, and in Proposition 7 we show that the contravariant and covariant tensor irreps discussed in Refs. 6–9 are star irreps of type (1) and type (2), respectively, explaining why the Young diagram method and the related method using Schur functions are so successful in dealing with these irreps. However, it is demonstrated that there exists an additional large class of typical star irreps.

The structure of the paper is as follows. Section II briefly specifies our notation and conventions. Sections III and IV provide a complete classification of the typical and atypical star irreps, respectively. In Sec. V we demonstrate that the contravariant and covariant tensor irreps are star irreps, and finally, in Sec. VI we conclude the paper with a summary of the main results and some remarks on further research.

II. NOTATIONS AND CONVENTIONS

We will use the same notations as in Refs. 1 and 2, but for convenience we briefly explain them here.

The gl($m|n$) generators E^a_b satisfy the graded commutation relations

$$[E^a_b, E^c_d] = \delta^c_b E^a_d - (-1)^{[(a)+(b)][(c)+(d)]}\delta^a_d E^c_b, \tag{1}$$

where a (and b, c, etc.) collectively represent even indices i and odd indices μ with $1 \leqslant i \leqslant m$, $1 \leqslant \mu \leqslant n$. The parity factor (a) is defined by

$$(a) = \begin{cases} 0, & a = i, \\ 1, & a = \mu, \end{cases}$$

and we say a generator E^a_b is even if $(a) + (b) \equiv 0 (\mathrm{mod}\ 2)$ and odd if $(a) + (b) \equiv 1 (\mathrm{mod}\ 2)$. It should be pointed out that the bracket on the lhs of (1) refers to the normal commutator except in the case when both generators E^a_b and E^c_d are odd, in which case it refers to the anticommutator.

As a basis for the Cartan subalgebra of gl($m|n$) we choose E^i_i, $1 \leqslant i \leqslant m$, and E^μ_μ, $1 \leqslant \mu \leqslant n$, whose eigenvalues serve to label the weights of the representations, and we de-

 0022-2488/90/112552-08$03.00

note an arbitrary gl$(m|n)$ weight Λ by

$$\Lambda = (\lambda_1,\lambda_2,...,\lambda_m|\omega_1,\omega_2,...,\omega_n)$$

$$= \sum_{i=1}^{m} \lambda_i \epsilon_i + \sum_{\mu=1}^{n} \omega_\mu \delta_\mu.$$

With this convention, the sets of even and odd positive roots are given by

$$\Phi_0^+ = \{\epsilon_i - \epsilon_j | 1 \leqslant i < j \leqslant m\} \cup \{\delta_\mu - \delta_\nu | 1 \leqslant \mu < \nu \leqslant n\},$$

$$\Phi_1^+ = \{\epsilon_i - \delta_\mu | 1 \leqslant i \leqslant m, 1 \leqslant \mu \leqslant n\}.$$

The weights ϵ_i and δ_μ form a basis for the weight space, which inherits a nondegenerate bilinear form from the structure of gl$(m|n)$ such that

$$(\epsilon_i,\epsilon_j) = \delta_{ij}, (\delta_\mu,\delta_\nu) = -\delta_{\mu\nu}, (\epsilon_i,\delta_\mu) = 0.$$

We denote the half-sums of the even and odd positive roots by ρ_0 and ρ_1, respectively, and set $\rho = \rho_0 - \rho_1$. The ρ's can be worked out easily from Φ_0^+ and Φ_1^+ and explicit results are given in Ref. 1. For a given subset θ of Φ_1^+, we also define

$$\rho_1(\theta) = \frac{1}{2} \sum_{\alpha \in \theta} \alpha, \quad \theta \subseteq \Phi_1^+. \tag{2}$$

Every finite-dimensional irreducible highest weight gl$(m|n)$ module $V(\Lambda)$ admits a natural Z gradation

$$V(\Lambda) = \bigoplus_{k=0}^{d} V_k(\Lambda),$$

where $V_k(\Lambda), 0 \leqslant k \leqslant d$, constitute modules of the even subalgebra of gl$(m|n)$, i.e., gl$(m) \oplus$ gl(n). We refer to $V_d(\Lambda)$, assumed nonzero, as the minimal Z-graded component of $V(\Lambda)$ and we say that $V(\Lambda)$ has $d+1$ levels.

The above gradation induces the following partitioning of the weights in $V(\Lambda)$

$$\Pi(\Lambda) = \bigcup_{k=0}^{d} \Pi_k(\Lambda),$$

where $\Pi_k(\Lambda)$ is the set of distinct weights in the subspace $V_k(\Lambda)$, which is stable under the Weyl group of gl$(m) \oplus$ gl(n) [also referred to as the Weyl group of gl$(m|n)$].

Let us denote the set of real dominant weights of gl$(m|n)$ by D_+. Following Ref. 1, on each irreducible highest weight gl$(m|n)$ module $V(\Lambda)$ with $\Lambda \in D_+$, there exists a well defined nondegenerate sesquilinear form $\langle | \rangle$, which has among others, the property that for all v and ω belonging to $V(\Lambda)$

$$\langle E^a_b v | w \rangle = (-)^{[\eta(v) + (\epsilon+\eta)(a) + \epsilon(b)] \cdot [(a)+(b)]} \langle v | E^b_a w \rangle,$$

where $(v) = k$ if $v \in V_k(\Lambda)$. Here, η and ϵ are grading parameters that may take values 0 and 1, leading to four inequivalent forms.[1]

Following Scheunert *et al.*[4] gl$(m|n)$ admits two types of star and two types of grade star modules. We say that $V(\Lambda)$ is an irreducible star module of type (1) [resp. (2)] if the sesquilinear form $\langle | \rangle$ with $\eta = 0$ and $\epsilon = 0$ (resp. $\epsilon = 1$) is positive definite equipping $V(\Lambda)$ with an inner product, and the corresponding irrep is called type (1) [resp. (2)] star. Similarly, $V(\Lambda)$ is called a grade star module of type (1) [resp. (2)] if the form $\langle | \rangle$ with $\eta = 1$ and $\epsilon = 0$ (resp. $\epsilon = 1$) is positive definite. In Ref. 1 we proved that there is a one-to-one correspondence between the two types of star irreps via duality. More precisely, we have the following proposition.

Proposition 1: The dual of a type (1) star irrep is a type (2) star irrep and vice versa. ■

This proposition will be extensively applied later in the paper when we classify type (2) star irreps. For the two types of grade star irreps, this one-to-one correspondence breaks down; actually we have the following.

Proposition 2: The dual of a type (1) grade star module with an even number of levels is a type (2) grade star module and vice versa, while the dual of a grade star module with an odd number of levels is grade star of the same type. ■

In the remainder of the paper, we will always use $V(\Lambda)$ to denote a finite-dimensional irreducible gl$(m|n)$ module with a real highest weight $\Lambda \in D_+$. It is also assumed that its maximal Z-graded component $V_0(\Lambda)$ is a Hermitian gl$(m) \oplus$ gl(n) module, or in other words, we choose a compact real form for the even subalgebra. The highest weight vector of $V(\Lambda)$ will be denoted by v^Λ.

III. TYPICAL STAR MODULES OF gl$(m|n)$

In this section we study the typical star irreps of gl$(m|n)$. The main results are Theorems 1 and 2 that classify the typical type (1) and type (2) star irreps, respectively, according to their highest weights.

Let us start with the typical type (1) star irreps. Following the strategy of Refs. 1 and 2, we define the gl$(m) \oplus$ gl(n) invariant:

$$\Gamma = \sum_{\mu=1}^{n} \sum_{i=1}^{m} E^\mu_i E^i_\mu.$$

It is straightforward to prove[1] that on each irreducible gl$(m) \oplus$ gl(n) module $V^{(0)}(\mu) \subseteq V(\Lambda)$, Γ takes the eigenvalue

$$\gamma = \tfrac{1}{2}(\Lambda - \mu, \Lambda + \mu + 2\rho).$$

If $\Pi_+(\Lambda)$ denotes the set of gl$(m) \oplus$ gl(n) highest weights appearing in $\Pi(\Lambda)$, we have the following.

Proposition 3: $V(\Lambda)$ is an irreducible type (1) star module iff

$$(\Lambda - \mu, \Lambda + \mu + 2\rho) \geqslant 0, \quad \forall \mu \in \Pi_+(\Lambda). \tag{3}$$

Proof: The proof goes in exactly the same way as in the case of gl$(m|1)$[1] and $C(n)$[2]. Set $\eta = \epsilon = 0$. Given any v belonging to a definite gl$(m) \oplus$ gl(n) irreducible module $V^{(0)}(\mu) \subseteq V(\Lambda)$,

$$\gamma \langle v | v \rangle = \sum_{\mu,i} \langle E^i_\mu v | E^i_\mu v \rangle. \tag{4}$$

If $V(\Lambda)$ is type (1) star, $\langle v | v \rangle > 0$, $\forall v \neq 0$ and $\langle E^i_\mu v | E^i_\mu v \rangle \geqslant 0$, $\forall i,\mu$, thus, $\gamma \geqslant 0$, and this proves the necessity of (3).

To show its sufficiency, we use induction on the different Z-graded levels. Now $\langle | \rangle$ is positive definite on $V_0(\Lambda)$ by definition. Assume that it is also positive definite on $V_{K-1}(\Lambda)$, $k \geqslant 1$; then for any $0 \neq v \in V_k(\Lambda)$ Eq. (4) holds provided that v belongs to a (possible multiple of a) definite irreducible gl$(m) \oplus$ gl(n) module. Since $E^i_\mu v \in V_{k-1}(\Lambda)$,

following the induction hypothesis, we conclude that all the terms on the rhs of (4) are non-negative. In fact, they cannot vanish simultaneously, otherwise we would have

$$E^i_\mu v = 0, \quad \forall i,\mu,$$

contradicting the assumption $v \in V_k(\Lambda)$, $k \geqslant 1$. Therefore, the rhs of (4) is positive and hence $\gamma > 0$, from which we obtain $\langle v|v\rangle > 0$ which completes the first part of the argument. As to the general case, observe that every nonzero $v \in V_k(\Lambda)$ is expressible as a sum

$$v = \sum_\alpha v_\alpha, \quad v_\alpha \neq 0,$$

where each v_α belongs to a (possible multiple of a) irreducible $\mathrm{gl}(m) \oplus \mathrm{gl}(n)$ submodule. Hence, we obtain, in view of the first step,

$$\langle v|v\rangle = \sum_\alpha \langle v_\alpha|v_\alpha\rangle > 0,$$

where we have used the fact that $\mathrm{gl}(m) \oplus \mathrm{gl}(n)$ submodules with different highest weights are orthogonal under the induced form.[1,2] This proves $\langle v|v\rangle > 0$, for all nonzero $v \in V_k(\Lambda)$, from which the result follows by induction. ■

Now our task is to reformulate the above proposition for typical type (1) star irreps such that no reference is made to individual $\mathrm{gl}(m) \oplus \mathrm{gl}(n)$ submodules. Before doing this, we prove the following technical result.

Lemma 1: Suppose $\Lambda \in D_+$ satisfies $(\Lambda + \rho, \epsilon_m - \delta_n) \geqslant 0$. Then

(i) $(\Lambda + \rho, \beta) \geqslant 0, \quad \forall \beta \in \Phi_1^+$,

(ii) $(\nu, \beta) \geqslant 0, \quad \forall \nu \in \Pi(\Lambda), \beta \in \Phi_1^+$.

Proof: (i) For $\beta = \epsilon_i - \delta_\mu \in \Phi_1^+$, we have

$$\begin{aligned}(\Lambda + \rho, \beta) &= (\Lambda + \rho, \epsilon_m - \delta_n) + (\Lambda + \rho, \epsilon_i - \epsilon_m) \\ &\quad + (\Lambda + \rho, \delta_n - \delta_\mu) \\ &\geqslant (\Lambda + \rho, \epsilon_m - \delta_n),\end{aligned}$$

where we have used the fact that $\Lambda + \rho \in D_+$. This proves (i).

(ii) As to the second statement, we first consider a weight $\nu \in \Pi_+(\Lambda)$ which is a $\mathrm{gl}(m) \oplus \mathrm{gl}(n)$ highest weight in $V(\Lambda)$. Such a ν can always be written in the form

$$\nu = \Lambda - 2\rho_1(\theta), \quad \theta \subseteq \Phi_1^+$$

with $\rho_1(\theta)$ as in Eq. (2). Now observe that

$$\begin{aligned}(2\rho_1(\theta), \epsilon_m - \delta_n) &= \sum_{\epsilon_i - \delta_\mu \in \theta} (\epsilon_i - \delta_\mu, \epsilon_m - \delta_n) \\ &\leqslant \sum_{\epsilon_m - \delta_\mu \in \theta} (\epsilon_m - \delta_\mu, \epsilon_m - \delta_n) \\ &\leqslant n - 1.\end{aligned}$$

Thus

$$\begin{aligned}(\nu, \epsilon_m - \delta_n) &\geqslant (\Lambda, \epsilon_m - \delta_n) + 1 - n \\ &= (\Lambda + \rho, \epsilon_m - \delta_n) \geqslant 0.\end{aligned}$$

Now every weight $\mu \in \Pi(\Lambda)$ can be written as

$$\mu = \nu - \sum_{i=1}^{m-1} p_i(\epsilon_i - \epsilon_{i+1}) - \sum_{\mu=1}^{n-1} q_\mu(\delta_\mu - \delta_{\mu+1}),$$

with $p_i, q_\mu \in \mathbb{Z}^+$ and $\nu \in \Pi_+(\Lambda)$, therefore

$$(\mu, \epsilon_m - \delta_n) = (\nu, \epsilon_m - \delta_n) + p_{m-1} + q_{n-1} \geqslant 0.$$

Using the fact that every $\beta \in \Phi_1^+$ is conjugate under the Weyl group to the odd root $\epsilon_m - \delta_n$, i.e., $\beta = \sigma(\epsilon_m - \delta_n)$, $\sigma \in W$, we arrive at

$$(\mu, \beta) = (\sigma^{-1}(\mu), \epsilon_m - \delta_n) \geqslant 0, \quad \forall \mu \in \Pi(\Lambda), \quad \beta \in \Phi_1^+.$$

It is worth mentioning that if $(\Lambda + \rho, \epsilon_m - \delta_n) > 0$, then equalities will not occur above. ■

Equipped with Proposition 3 and the preceeding lemma, we are now ready to prove one of the main results of this section.

Theorem 1: Let $\Lambda \in D_+$. The irreducible highest weight $\mathrm{gl}(m|n)$ module $V(\Lambda)$ is typical and type (1)-star iff Λ is real and

$$(\Lambda + \rho, \epsilon_m - \delta_n) > 0. \tag{5}$$

Proof: Let us first show the necessity of (5). Consider the two vectors u and v defined by

$$v = E^{\bar{n}}_m u, \quad u = E^{n-1}_m E^{n-2}_m \cdots E^{\bar{1}}_m v^\Lambda$$

: here we have adopted the convention that indices with an overbar are odd. If $V(\Lambda)$ is typical and type (1)-star, we necessarily have

$$\langle u|u\rangle > 0, \quad \langle v|v\rangle > 0.$$

Note that

$$\langle v|v\rangle = (\Lambda + \rho, \epsilon_m - \delta_n)\langle u|u\rangle.$$

Thus

$$(\Lambda + \rho, \epsilon_m - \delta_n) > 0.$$

To show the sufficiency of (5), it suffices to prove

$$(\Lambda - \nu, \Lambda + \nu + 2\rho) \geqslant 0, \quad \forall \nu \in \Pi_+(\Lambda)$$

because of Proposition 3. We will carry out the proof in two steps. Recall first that every $\nu \in \Pi_+(\Lambda)$ can be expressed as

$$\nu = \Lambda - 2\rho_1(\theta), \quad \theta \subseteq \Phi_1^+$$

so it suffices to demonstrate that

$$(2\rho_1(\theta), 2(\Lambda + \rho) - 2\rho_1(\theta)) \geqslant 0. \tag{3'}$$

Step (a): Assume first that θ is of the special form

$$\theta = \{\epsilon_i - \delta_{\mu_1}, \epsilon_i - \delta_{\mu_2}, \ldots, \epsilon_i - \delta_{\mu_k}\}, \quad 1 \leqslant k \leqslant n, \tag{6}$$

where i is a fixed even index and the odd indices are all different, which we can assume to have the ordering $\mu_1 < \mu_2 < \cdots < \mu_k$ without losing generality. Now

$$\begin{aligned}&(\Lambda - \nu, \Lambda + \nu + 2\rho) \\ &\quad = (2\rho_1(\theta), 2(\Lambda + \rho)) - (2\rho_1(\theta), 2\rho_1(\theta)).\end{aligned}$$

We consider the two terms seperately. For the first term, we have

$$\begin{aligned}&2(\Lambda + \rho, 2\rho_1(\theta)) \\ &\quad = 2\sum_{\alpha=1}^{k} (\Lambda + \rho, \epsilon_i - \delta_{\mu_\alpha}) \\ &\quad = 2k(\Lambda + \rho, \epsilon_i - \delta_{\mu_k}) + 2\sum_{\alpha=1}^{k} (\Lambda + \rho, \delta_{\mu_k} - \delta_{\mu_\alpha}) \\ &\quad \geqslant 2k(\Lambda + \rho, \epsilon_i - \delta_{\mu_k}) + 2\sum_{\alpha=1}^{k} (\rho, \delta_{\mu_k} - \delta_{\mu_\alpha}).\end{aligned}$$

The definition of θ, i.e., (6), guarantees that there exist at least $k-\alpha+1$ distinct odd indices μ such that $\mu_\alpha \leqslant \mu \leqslant \mu_k$, thus

$$2\sum_{\alpha=1}^{k}(\rho,\delta_{\mu_k}-\delta_{\mu_\alpha}) \geqslant 2\sum_{\alpha=1}^{k}(k-\alpha)=k(k-1).$$

Therefore, $2(\Lambda+\rho,2\rho_1(\theta)) \geqslant k(k-1)$, where the first statement of Lemma 1 is used.

As to the second term, we have

$$(2\rho_1(\theta),2\rho_1(\theta))=\sum_{\alpha,\beta=1}^{k}(1-\delta_{\mu_\alpha\mu_\beta})$$

$$=\sum_{\alpha,\beta=1}^{k}(1-\delta_{\alpha,\beta})=k(k-1).$$

Therefore, we conclude that

$$(2\rho_1(\theta),2(\Lambda+\rho)-2\rho_1(\theta)) \geqslant 0,$$

when θ is of the special form (6).

Step (*b*): Now we consider an arbitrary $\theta \subseteq \Phi_1^+$. It is crucial to observe that every such θ can be partitioned into

$$\theta=\bigcup_{i=1}^{m}\theta_i,$$

with $\theta_i \subset \Phi_1^+$ being a set of the special form of (6), i.e.,

$$\theta_i=\{\epsilon_i-\delta_{\mu_{i,1}},\epsilon_i-\delta_{\mu_{i,2}},\ldots,\epsilon_i-\delta_{\mu_{i,k_i}}\},\quad 1 \leqslant \mu_{i,\alpha} \leqslant n.$$

We assume that $0 \leqslant k_i \leqslant m$ and interpret $k_i=0$ as θ_i being empty. Now utilizing the results of Step (a), we obtain

$$2(\Lambda+\rho,2\rho_1(\theta)) \geqslant \sum_{i=1}^{h_m}k_i(k_i-1) \tag{7}$$

and

$$(2\rho_1(\theta),2\rho_1(\theta))=\sum_{i=1}^{h_m}k_i(k_i-1)$$

$$+\sum_{i\neq j}(2\rho_1(\theta_i),2\rho_1(\theta_j)).$$

Since for $i \neq j$,

$$(2\rho_1(\theta_i),2\rho_1(\theta_j))=\sum_{\alpha=1}^{k_i}\sum_{\beta=1}^{k_j}(\epsilon_i-\delta_{\mu_{i,\alpha}},\epsilon_j-\delta_{\mu_{j,\beta}})$$

$$=\sum_{\alpha=1}^{k_i}\sum_{\beta=1}^{k_j}(\delta_{\mu_{i,\alpha}}-\delta_{\mu_{j,\beta}}) \leqslant 0,$$

we have

$$(2\rho_1(\theta),2\rho_1(\theta)) \leqslant \sum_{i=1}^{m}k_i(k_i-1). \tag{8}$$

Combining (7) and (8) together leads to the inequality (3'), thus proving the theorem. ■

There is a minor point worth mentioning: $(\Lambda+\rho,\epsilon_m-\delta_n)>0$ guarantees that Λ is typical.

Now we turn to the classification of typical star irreps of type (2). Following Proposition 1, $V(\Lambda)$ is a typical type (2) star module if and only if

$$(-\Lambda^-+\rho,\epsilon_m-\delta_n)>0,$$

where Λ^- is the lowest weight of $V(\Lambda)$. For a typical highest weight Λ defined by

$$\Lambda=(\lambda_1,\lambda_2,\ldots,\lambda_m|\omega_1,\omega_2,\ldots,\omega_n)$$

the corresponding lowest weight can be easily determined to be

$$\Lambda^-=(\lambda_m,\lambda_{m-1},\ldots,\lambda_1|\omega_n,\omega_{n-1},\ldots,\omega_1)-2\rho_1,$$

thus

$$(\Lambda^-,\epsilon_m-\delta_n)=(\Lambda,\epsilon_1-\delta_1)-(2\rho_1,\epsilon_1-\delta_1)$$

and

$$(\Lambda^--\rho,\epsilon_m-\delta_n)=(\Lambda+\rho,\epsilon_1-\delta_1).$$

Therefore, we arrive at the following theorem.

Theorem 2: For $\Lambda \in D_+$, $V(\Lambda)$ is a typical type (2)-star $\mathrm{gl}(m|n)$ module iff Λ is real and

$$(\Lambda+\rho,\epsilon_1-\delta_1)<0. \tag{9}$$

■

Theorems 1 and 2 classify the typical star modules of $\mathrm{gl}(m|n)$ completely. Now we turn to the classification of the atypical modules.

IV. ATYPICAL STAR IRREPS OF gl(m|n)

Having classified the typical star irreps of $\mathrm{gl}(m|n)$ we now turn our attention to the atypical star irreps. We will prove two theorems [(3) and (4)], which specify the necessary and sufficient conditions on the highest weight of a finite-dimensional irreducible $\mathrm{gl}(m|n)$ module $V(\Lambda)$ in order for it to be star.

A. Atypical type (1) star irreps of gl(m|n)

For atypical type (1) star irreps, we have the following result.

Proposition 4: Let $\Lambda \in D_+$ be atypical. If $V(\Lambda)$ is a type (1) star $\mathrm{gl}(m|n)$ module, there exists an odd index $1 \leqslant \mu \leqslant n$ such that

$$(\Lambda+\rho,\epsilon_m-\delta_\mu)=0,\quad (\Lambda,\delta_\mu-\delta_n)=0. \tag{10}$$

Proof: If $V(\Lambda)$ is type (1)-star, then [notation as in Theorem (1)]

$$0 \leqslant \langle E_m^{\bar{1}}v^\Lambda|E_m^{\bar{1}}v^\Lambda\rangle=(\Lambda+\rho,\epsilon_m-\delta_1)\langle v^\Lambda|v^\Lambda\rangle,$$

i.e., $(\Lambda+\rho,\epsilon_m-\delta_1) \geqslant 0$. If equality occurs, then $E_m^{\bar{1}}v^\Lambda=0$ and we take $\mu=1$, otherwise we consider

$$0<\langle E_m^{\bar{2}}E_m^{\bar{1}}v^\Lambda|E_m^{\bar{2}}E_m^{\bar{1}}v^\Lambda\rangle$$

$$=(\Lambda+\rho,\epsilon_m-\delta_2)\langle E_m^{\bar{1}}v^\Lambda|E_m^{\bar{1}}v^\Lambda\rangle.$$

We must again have either $(\Lambda+\rho,\epsilon_m-\delta_2)=0$ and $E_m^{\bar{2}}E_m^{\bar{1}}v^\Lambda=0$ or else $(\Lambda+\rho,\epsilon_m-\delta_2)>0$. In the former case, we take $\mu=2$; in the latter case we will continue the process until we find a vector u such that

$$u=E_m^{\mu}v,\quad v=E_m^{\mu-1}E_m^{\mu-2}\cdots E_m^{\bar{1}}v^\Lambda$$

with $\langle u|u\rangle=0$ but $\langle v|v\rangle>0$. Now

$$\langle u|u\rangle=(\Lambda+\rho,\epsilon_m-\delta_\mu)\langle v|v\rangle$$

thus

$$(\Lambda+\rho,\epsilon_m-\delta_\mu)=0. \tag{10a}$$

Such a vector u must exist, otherwise we would have $(\Lambda+\rho,\epsilon_m-\delta_n)>0$, and this makes Λ typical, contradicting the given conditions.

Assume we have found such an index μ which renders

the first condition of (10), i.e., (10a) satisfied; then we consider the vector

$$u' = E^{\bar{n}}_{m} v.$$

From $\langle u'|u'\rangle = [(\Lambda,\epsilon_m - \delta_n) - (\mu - 1)]\langle v|v\rangle \geqslant 0$ we deduce that

$$(\Lambda,\epsilon_m - \delta_n) \geqslant \mu - 1.$$

On the other hand, the lexicality of Λ requires

$$0 \geqslant (\Lambda,\delta_\mu - \delta_n) = (\Lambda,\epsilon_m - \delta_n) - (\Lambda,\epsilon_m - \delta_\mu)$$
$$= (\Lambda,\epsilon_m - \delta_n) - (\mu - 1) \geqslant 0, \quad (10b)$$

thus

$$(\Lambda,\delta_\mu - \delta_n) = 0$$

as we wished to prove. ■

Proposition 4 provides us with the necessary condition, i.e., (10), for a gl$(m|n)$ module $V(\Lambda)$ to be type (1)-star and atypical. It turns out that (10) is also sufficient. To prove this we need the following technical results.

Lemma 2: Given two type (1) [resp. (2)] star irreducible gl$(m|n)$ modules $V(\Lambda)$ and $V(\Lambda')$, the irreducible gl$(m|n)$ module $V(\Lambda'')$ is contained in the tensor product $V(\Lambda) \otimes V(\Lambda')$ if the maximal $\mathbb{Z}$-graded component of $V(\Lambda'')$, i.e., $V_0(\Lambda'')$ which is an irreducible gl$(m) \oplus$ gl(n) module, is contained in $V_0(\Lambda) \otimes V_0(\Lambda')$, and in this case $V(\Lambda'')$ is also type (1) [resp. (2)] star.

Proof: The lemma follows trivially from the fact[4] that the tensor product of two type (1) [resp. (2)] star irreps reduces completely into type (1) [resp. (2)] star irreps. ■

Lemma 3: Let $V_0(\Lambda)$ and $V_0(\Lambda')$ be finite-dimensional irreducible gl$(m) \oplus$ gl(n) modules and suppose $\nu \in \Pi_0(\Lambda)$ is W conjugate to Λ. Then the irreducible gl$(m) \oplus$ gl(n) module $V_0(\Lambda' + \nu)$ occurs (with unit multiplicity) in $V_0(\Lambda) \otimes V_0(\Lambda')$ if and only if $\nu + \Lambda' \in D_+$.

Proof: This result is proved for arbitrary semisimple Lie algebras in Ref. 11. ■

By considering the $(m+n)\times(m+n)$ matrices of the vector irrep of gl$(m|n)$, which has a highest weight $\Lambda_0 = (1,0,...,0|0,...,0)$, one can easily see that the vector module $V(\Lambda_0)$ is type (1)-star. However, by repeated application of Lemmas 2 and 3, we note that the irreducible gl$(m|n)$ module $V(\Lambda_1)$ with

$$\Lambda_1 = (\lambda_1,\lambda_2,...,\lambda_{m-1},0|0,0,...,0), \quad \lambda_i \in \mathbb{Z}^+, \quad \lambda_i \leqslant \lambda_{i-1},$$

appears in the tensor product

$$V(\Lambda_0) \otimes V(\Lambda_0) \otimes \cdots \otimes V(\Lambda_0) \quad \left(\sum_{i=1}^{m-1} \lambda_i \text{ copies}\right).$$

It thus follows that $V(\Lambda_1)$ is necessarily a type (1) star gl$(m|n)$ module.

Then, $V(\Lambda_1)$ serves as one of the building blocks of all the type (1) star modules of gl$(m|n)$. Another one is $V(\Lambda_2)$ with

$$\Lambda_2 = (1,1,...,1|\omega,0,...,0), \quad \omega \in \mathbb{Z}^+. \quad (11)$$

In Proposition 5 below we will show that the highest weight of the dual module of $V(\Lambda_2)$ is

$$\Lambda_2^* = (0,0,...,0|0,0,...,0,-(\omega+m)), \quad (12)$$

which is obviously type (2)-star since it is the rank $(\omega + m)$ symmetric tensor product of the contragrediant vector module $V(0,...,0|,...,0,-1) = V^*(\Lambda_0)$. Thus $V(\Lambda_2)$ is a type (1) star module.

Now let us construct the tensor product:

$$\bigotimes_{\alpha=1}^{\mu-1} V(1,1,...,1|\tilde{\omega}_\alpha,0,0,...,0), \quad \tilde{\omega}_\alpha \in \mathbb{Z}^+, \tilde{\omega}_\alpha \geqslant \tilde{\omega}_{\alpha+1}.$$

Following Lemmas 2 and 3, one of the component irreps occurring in this product is $V(\Lambda_3)$ with

$$\Lambda_3 = (\mu-1,\mu-1,...,\mu-1|\tilde{\omega}_1,\tilde{\omega}_2,...,\tilde{\omega}_{\mu-1},0,0,...,0). \quad (13)$$

Finally we form the tensor product $V(\Lambda_1) \otimes V(\Lambda_3)$, which contains the irreducible module $V(\tilde{\Lambda})$ with

$$\tilde{\Lambda} = (\tilde{\lambda}_1,\tilde{\lambda}_2,...,\tilde{\lambda}_{m-1},\mu-1|\tilde{\omega}_1,\tilde{\omega}_2,...,\tilde{\omega}_{\mu-1},0,0,...,0),$$
$$\tilde{\lambda}_i = \lambda_i + \mu - 1, \quad \tilde{\lambda}_i,\tilde{\omega}_\alpha \in \mathbb{Z}^+. \quad (14)$$

Since both $V(\Lambda_1)$ and $V(\Lambda_3)$ are type (1) star, so too is $V(\tilde{\Lambda})$.

Let us finally consider the representation of gl$(m|n)$ defined by

$$\tilde{\pi}_\Lambda(E^a_b) = \pi_\Lambda(E^a_b) + (-1)^{(a)}\delta^a_b\omega, \quad \omega \in \mathbb{R} \quad (15)$$

where π_Λ is the irrep of gl$(m|n)$ with highest weight

$$\Lambda = (\lambda_1,\lambda_2,...,\lambda_m|\omega_1,...,\omega_{\mu-1},\omega,\omega,...,\omega), \quad (16)$$

satisfying $\lambda_m + \omega = \mu - 1$. Then $\tilde{\pi}_\Lambda$ gives rises to an irrep with highest weight (14) which, as we have seen, is type (1) star. This implies that

$$\pi^\dagger_\Lambda(E^a_b) = \tilde{\pi}^\dagger_\Lambda(E^a_b) - (-1)^{(a)}\delta^a_b\omega$$
$$= \tilde{\pi}_\Lambda(E^b_a) - (-1)^{(a)}\delta^a_b\omega$$
$$= \pi_\Lambda(E^b_a),$$

so that $V(\Lambda)$, with Λ as in Eq. (16), gives rise to a type (1)-star irrep of gl$(m|n)$.

Note that (16) is the most general form of a real $\Lambda \in D_+$ satisfying condition (10). Thus we have proved the following.

Theorem 3: Suppose $\Lambda \in D_+$ is real. The gl$(m|n)$ irreducible module $V(\Lambda)$ is atypical and type (1)-star iff there exists an odd index $\mu \in \{1,2,...,n\}$ such that

$$(\Lambda + \rho,\epsilon_m - \delta_\mu) = (\Lambda,\delta_\mu - \delta_n) = 0. \quad (10')$$

■

Now we determine the lowest weight of $V(\Lambda)$ with Λ satisfying (10'). Define a sequence of odd indices μ_i, $1 \leqslant i \leqslant m$, by

$$\mu_i = [\mu_m + (\Lambda,\epsilon_i - \epsilon_m)] \wedge n, \quad a \wedge b = \min(a,b), \quad (17)$$

where

$$\mu_m = \mu - 1.$$

For each $i \in \{1,2,...,m\}$, we set

$$T_i = \prod_{\nu=1}^{\mu_i} E^\nu_i$$

and consider the sequence of vectors

$$w_i = T_i w_{i+1}, \quad w_m = T_m v^\wedge.$$

Note first that $w_m \neq 0$ since

$$\langle w_m|w_m\rangle = \langle v^\Lambda|v^\Lambda\rangle \cdot \prod_{\nu=1}^{\mu_m} (\Lambda+\rho,\epsilon_m-\delta_\nu) > 0,$$

where $\langle|\rangle$ is the nondegenerate sesquilinear form induced on $V(\Lambda)$. However,

$$w = E_m^{\mu_m+1} T_m v^\Lambda = 0,$$

because there are at most μ_m odd raising operators E_1^m, $E_2^m,\ldots,E_{\mu_m}^m$ which do not annihilate w, hence w, if nonzero, would generate a submodule of $V(\Lambda)$ not containing the maximal state v^Λ. Therefore, the existence of a nonvanishing w contradicts the irreducibility of $V(\Lambda)$. Using the property of v^Λ that

$$E_{\nu_1}^{\nu_2} v^\Lambda = 0, \quad \forall \nu_1,\nu_2 \geqslant \mu_m + 1, \quad \nu_1 \neq \nu_2, \tag{18}$$

we obtain

$$E_{\mu_{m+1}}^{\nu} w = E_m^\nu w_m = 0, \quad \forall \nu \geqslant \mu_m + 1.$$

Thus

$$E_m^\nu w_m = 0, \quad \forall \nu.$$

More generally consider the vector

$$w_i = T_i T_{i+1} \cdots T_m v^\Lambda,$$

which is nonvanishing since (by direct calculation)

$$\langle w_i|w_i\rangle > 0,$$

and $\mathrm{gl}(m) \oplus \mathrm{gl}(n)$ maximal. Observe that, for $\mu_i < n$,

$$w' = E_i^{\mu_i+1} w_i = 0,$$

otherwise w' would be a maximal $\mathrm{gl}(m) \oplus \mathrm{gl}(n)$ state with a nonlexical weight. Using (18) again, we can easily show that

$$E_i^\nu w_i = 0, \quad \forall \nu,$$

and this implies

$$E_j^\nu w_i = 0, \quad \forall j \geqslant i, \nu = 1,2,\ldots,n,$$

because the odd lowering operators anticommute.

Particularly interesting is

$$w_1 = T_1 T_2 \cdots T_m v^\Lambda,$$

which is nonvanishing, $\mathrm{gl}(m) \oplus \mathrm{gl}(n)$ maximal and satisfying

$$E_i^\nu w_1 = 0, \quad \forall i,\nu,$$

and therefore must be the highest weight vector of the minimal $\mathbb{Z}$-graded component of $V(\Lambda)$. We have thus proved the following.

Proposition 5: Let $\Lambda \in D_+$ satisfy

$$(\Lambda+\rho,\epsilon_m-\delta_\mu) = (\Lambda,\delta_\mu-\delta_n) = 0,$$

for a fixed odd index μ. Then,

(i) the highest weight of the minimal $\mathbb{Z}$-graded component of the irreducible $\mathrm{gl}(m|n)$ module $V(\Lambda)$ is

$$\bar{\Lambda} = \Lambda - \sum_{i=1}^{m} \sum_{\nu=1}^{\mu_i} (\epsilon_i - \delta_\nu);$$

(ii) the lowest weight of $V(\Lambda)$ is

$$\Lambda^- = \tau(\bar{\Lambda}),$$

where τ is the unique Weyl group element sending the positive even roots into negative ones;

(iii) $V(\Lambda)$ admits $d_\Lambda + 1$ levels with

$$d_\Lambda = \sum_{i=1}^{m} \mu_i. \qquad \blacksquare$$

We want to point out that our proof of Proposition 5 does not depend on the fact that $V(\Lambda)$ is type (1)-star, thus it is legitimate to use this result in proving Theorem 3. Applying (i) and (ii) to the highest weight defined by (11) we can easily see that the corresponding lowest weight is the negative of (12).

B. Atypical type (2) star-irreps of gl(m|n)

Proposition 5 also enables us to classify the atypical type (2) star irreps explicitly in terms of their highest weights. Since there is a one-to-one correspondence between type (1) and type (2)-star modules via duality, the classification of atypical type (2) star irreps is actually accomplished by Theorem 3, although in an implicit way. Our task now is to make this classification explicit.

Let us assume that $V(\Lambda)$ is an atypical type (1)-star $\mathrm{gl}(m|n)$ module, so that its highest weight Λ satisfies (10'). Consider the indices μ_i, $i = 1,2,\ldots,m$, defined by (17). In the most general case, we have

$$\mu_1 = \mu_2 = \cdots = \mu_l = n > \mu_{l+1} \geqslant \mu_{l+2} \geqslant \cdots \geqslant \mu_m \geqslant 0,$$

with l a fixed even index. We will allow the possibility that $l = 0$, in which case we understand that $n > \mu_i$, $\forall_i$.

Using Proposition 5 we obtain the highest weight of the minimal $\mathbb{Z}$-graded component of $V(\Lambda)$,

$$\bar{\Lambda} = \Lambda - \sum_{i=1}^{m} \sum_{\nu=1}^{\mu_i} (\epsilon_i - \delta_\nu). \tag{19}$$

Its even components can be easily read from this formula

$$\bar{\lambda}_i = \lambda_i - \mu_i, \quad \forall_i.$$

More explicitly,

$$\bar{\lambda}_j = \lambda_j - n, \quad 1 \leqslant j \leqslant l$$

and

$$\begin{aligned}\bar{\lambda}_j &= \lambda_j - \mu_m - (\Lambda,\epsilon_j - \epsilon_m) \\ &= \lambda_m - \mu + 1, \quad j > l.\end{aligned}$$

Thus the last $m - l$ even components of $\bar{\Lambda}$ are all equal, i.e.,

$$(\bar{\Lambda},\epsilon_{l+1} - \epsilon_m) = 0.$$

Now let τ be the unique Weyl group element sending positive even roots to negative ones, so that the lowest weight of $V(\Lambda)$ is

$$\Lambda^- = \tau(\bar{\Lambda}).$$

Then the highest weight of $V^*(\Lambda)$ is

$$\Lambda^* = -\Lambda^-,$$

satisfying

$$(\Lambda^*,\epsilon_1 - \epsilon_{m-l}) = 0. \tag{20}$$

Observe that

$$\begin{aligned}&-(\Lambda^* + \rho,\epsilon_{m-l} - \delta_1) \\ &\quad = (\tau(\bar{\Lambda}),\epsilon_{m-l} - \delta_1) - (\rho,\epsilon_{m-l} - \delta_1) \\ &\quad = (\bar{\Lambda},\epsilon_{l+1} - \delta_n) - l \\ &\quad = \bar{\lambda}_{l+1} + \bar{\omega}_n - l,\end{aligned}$$

where $\bar{\omega}_n$ is the nth odd component of $\bar{\Lambda}$, which is determined by (19) to be

$$\bar{\omega}_n = \omega_n + l,$$

hence,

$$(\Lambda^* + \rho, \epsilon_{m-l} - \delta_1) = 0. \tag{21}$$

Equations (20) and (21) are satisfied by the highest weight of the dual module of every atypical type (1)-star module, from which we deduce that if $V(\Lambda)$ is atypical and type (2)-star then its highest weight $\Lambda \in D_+$ must satisfy

$$(\Lambda + \rho, \epsilon_k - \delta_1) = 0, \quad (\Lambda, \epsilon_1 - \epsilon_k) = 0, \quad 1 \leqslant k \leqslant m \text{ fixed}. \tag{22}$$

To establish the converse, we need to show that condition (22) implies (10′) for the highest weight of the dual module of $V(\Lambda)$. Let $\Lambda \in D_+$ satisfy (22), and define even indices, $m+1 \geqslant i_1 \geqslant i_2 \geqslant \cdots \geqslant i_n \geqslant 1$, by

$$i_\mu = [i_1 + (\Lambda, \delta_1 - \delta_\mu)] \vee 1, \quad a \vee b = \max(a,b), \tag{23}$$

with

$$i_1 = k + 1.$$

Using exactly the same argument as for proving Proposition 5, we can show the following.

Proposition 6: Let $\Lambda \in D_+$ satisfy

$$(\Lambda + \rho, \epsilon_k - \delta_1) = 0 \quad (\Lambda, \epsilon_1 - \epsilon_k) = 0,$$

for a fixed even index k. Then,

(i) the highest weight of the minimal $\mathbb{Z}$-graded component of the irreducible $\mathrm{gl}(m/n)$ module $V(\Lambda)$ is

$$\bar{\Lambda} = \Lambda - \sum_{\nu=1}^{n} \sum_{j=i_\nu}^{m} (\epsilon_j - \delta_\nu);$$

(ii) the lowest weight of $V(\Lambda)$ is

$$\Lambda^- = \tau(\bar{\Lambda});$$

(iii) $V(\Lambda)$ admits $d_\Lambda + 1$ levels with

$$d_\Lambda = \sum_{\nu=1}^{n} (m + 1 - i_\nu). \qquad \blacksquare$$

From (i) we can see that the odd components of $\bar{\Lambda}$ are given by

$$\bar{\omega}_\mu = \omega_\mu + m + 1 - i_\mu.$$

In the most general case, we have

$$i_1 \geqslant i_2 \geqslant \cdots \geqslant i_{\xi-1} > i_\xi = i_{\xi+1} = \cdots = i_n = 1,$$

so that

$$\bar{\omega}_\mu = \omega_\mu - i_\mu + m + 1 = \omega_1 + m - k, \quad \text{for } \mu < \xi,$$

$$\bar{\omega}_\mu = \omega_\mu + m, \quad \text{for } \mu \geqslant \xi.$$

Therefore, the first $\xi - 1$ odd components of $\bar{\Lambda}$ are all equal, i.e.,

$$(\bar{\Lambda}, \delta_1 - \delta_{\xi-1}) = 0,$$

which implies

$$(\tau(\bar{\Lambda}), \delta_{n+2-\xi} - \delta_n) = 0. \tag{24}$$

Now observe that

$$\begin{aligned} -(\tau(\bar{\Lambda}) - \rho, \epsilon_m - \delta_{n+2-\xi}) &= -(\bar{\Lambda}, \epsilon_1 - \delta_{\xi-1}) + (\rho, \epsilon_m - \delta_{n+2-\xi}) \\ &= -\bar{\lambda}_1 - (\omega_1 + m - k) - (n + 1 - \xi), \end{aligned}$$

where $\bar{\lambda}_1$ can be read off from (i) to be

$$\bar{\lambda}_1 = \lambda_1 - (n + 1 - \xi).$$

Therefore,

$$(-\tau(\bar{\Lambda}) + \rho, \epsilon_m - \delta_{n+2-\xi}) = -(\lambda_1 + \omega_1 + m - k)$$

and using the fact that $\lambda_1 = \lambda_k$, we obtain

$$(\Lambda^* + \rho, \epsilon_m - \delta_{n+2-\xi}) = -(\Lambda + \rho, \epsilon_k - \delta_1) = 0, \tag{25}$$

where

$$\Lambda^* = -\tau(\bar{\Lambda}).$$

Combining (24) and (25) together we see that the highest weight of the dual of $V(\Lambda)$, i.e., $V^*(\Lambda)$, satisfies

$$(\Lambda^* + \rho, \epsilon_m - \delta_{n+2-\xi}) = (\Lambda^*, \delta_{n+2-\xi} - \delta_n) = 0, \tag{26}$$

i.e., $V^*(\Lambda)$ is a type (1)-star module following Theorem 3. Therefore, we have proved the following.

Theorem 4: Let $\Lambda \in D_+$. The irreducible $\mathrm{gl}(m|n)$ module $V(\Lambda)$ is atypical and type (2)-star iff there exists an even index $i \in \{1,2,\ldots,m\}$ such that

$$(\Lambda + \rho, \epsilon_i - \delta_1) = (\Lambda, \epsilon_1 - \epsilon_i) = 0. \tag{27}$$

$\blacksquare$

V. TENSOR IRREPS OF gl($m|n$)

Having classified all the star irreps of $\mathrm{gl}(m|n)$, we now apply this classification to the so-called contravariant and covariant tensor irreps,[7] which have been studied quite extensively using the Young diagram method,[6-8] and Schur functions.[9] We will show the following.

Proposition 7: The contravariant and covariant tensor irreps of $\mathrm{gl}(m|n)$ are star irreps of type (1) and type (2), respectively.

$\blacksquare$

In view of the discussions leading to Theorem 3, we can easily see that the proposition holds. However, to make our argument more precise, recall that a contravariant tensor irrep[7] of $\mathrm{gl}(m|n)$ is by definition a tensor product of the vector irrep satisfying certain symmetry properties. Such an irrep can be characterized by a partition

$$P = (p_1, p_2, \ldots, p_m, p_{m+1} p_{m+2}, \ldots, p_N),$$

where $p_a \in \mathbb{Z}^+$, $p_a \geqslant p_{a+1}$, $\forall a$ and $p_{m+1} \leqslant n$ is assumed. Associated with each partition P, there exists a unique [up to the automorphism (15)] highest weight

$$\Lambda = (\lambda_1, \lambda_2, \ldots, \lambda_m | \omega_1, \omega_2, \ldots, \omega_n)$$

with

$$\lambda_i = p_i, \quad i = 1,2,\ldots,m$$

and the ω_α's defined by

$$\sum_{\nu=1}^{n} \omega_\nu \delta_\nu = \sum_{\nu_1=1}^{p_{m+1}} \delta_{\nu_1} + \sum_{\nu_2=1}^{p_{m+2}} \delta_{\nu_1} + \cdots + \sum_{\nu_{N-m}=1}^{p_N} \delta_{\nu_{N-m}}.$$

When $p_m \geqslant n$, we have

$$(\Lambda+\rho,\epsilon_m-\delta_n)=p_m+\omega_n+1-n\geqslant\omega_n+1>0,$$

hence, the irreducible gl$(m|n)$ module associated with the partition P is typical and type (1)-star. If $p_m<n$, let us define

$$\mu=p_m+1.$$

Since $p_a<\mu$, $\forall a>m$, we immediately see that

$$(\Lambda,\delta_\nu-\delta_\mu)=0,\quad \forall\nu\geqslant\mu$$

and

$$(\Lambda+\rho,\epsilon_m-\delta_\mu)=p_m+1-\mu=0.$$

Therefore, the irreducible gl$(m|n)$ module $V(\Lambda)$ is atypical and type (1)-star.

In a similar way, the covariant tensor irreps arise from tensor products of the contragredient vector module $V^*(\Lambda_0)$, which is type (2)-star. Thus the covariant tensor irreps must all be type (2)-star.

A direct consequence of the proposition is that the Young diagram method always produces irreducible irreps of gl$(m|n)$ when applied to a given type of tensor irreps (i.e., not a mixture of covariant and contravariant tensors!). As far as we are aware, a convincing proof was not given before for this generally accepted fact, which insures the applicability of the Young diagram method to the two types of tensor irreps.

It should be pointed out that the converse of Proposition 7 is not true. Let us consider for example the type (1) star irreps. It can be easily seen that the atypical type (1) star irreps and those typical ones with

$$c=(\Lambda+\rho,\epsilon_m-\delta_n)>0$$

being integral constitute the contravariant tensor irreps and those irreps obtained by tensoring them with a one-dimensional (1-D) representation. However, those irreps with c greater than zero but nonintegral are intrinsically different. In this case, we may write

$$c=\gamma+\Delta,\quad \gamma\in\mathbb{Z}^+,\quad \Delta\in(0,1).$$

Thus, corresponding to each non-negative integer γ, there exists a one-parameter family of typical type (1) star irreps, which are not tensorial, thus, cannot be dealt with by the Young diagram techniques.

For type (2) star irreps the situation is the same. The irreps with $c'=(\Lambda+\rho,\epsilon_1-\delta_1)$ negative and nonintegral are nontensorial, and the rest consists of the covariant tensor irreps and those obtained by tensoring them with a one-dimensional representation.

VI. CONCLUSION

We have systematically classified all the finite-dimensional star irreps of gl$(m|n)$ in terms of their highest weights. In particular we have shown in Theorems 1 and 3 that a given gl$(m|n)$ module $V(\Lambda)$, with $\Lambda\in D_+$, is type (1)-star iff (i) $(\Lambda+\rho,\epsilon_m-\delta_n)>0$; or (ii) there exists an odd index $\mu\in\{1,2,\ldots,n\}$ such that

$$(\Lambda+\rho,\epsilon_m-\delta_\mu)=(\Lambda,\delta_\mu-\delta_n)=0.$$

In the former case the given condition also enforces typicality on $V(\Lambda)$, while in the latter case all irreps are obviously atypical. Except for the typical type (1) star irreps with $c=(\Lambda+\rho,\epsilon_m-\delta_n)>0$ being nonintegral, the rest of the type (1) star irreps constitute the contravariant tensor irreps studied in Refs. 6–9 and their tensor products with 1-D irreps. The lowest weights of the atypical type (1) star modules are determined explicity in Proposition 5 (the lowest weight of a typical module is trivial to obtain).

Theorems 2 and 4 give a complete classification of the type (2) star modules of gl$(m|n)$. It is shown that a gl$(m|n)$ module $V(\Lambda)$, with $\Lambda\in D_+$, is type (2)-star iff (i) $(\Lambda+\rho,\epsilon_1-\delta_1)<0$; or (ii) there exists an even index $k\in\{1,2,\ldots,m\}$ such that

$$(\Lambda+\rho,\epsilon_k-\delta_1)=(\Lambda,\epsilon_1-\epsilon_k)=0.$$

In the former case $V(\Lambda)$ is typical, while in the latter case $V(\Lambda)$ is atypical, the lowest weight of which is given in Proposition 6. The atypical type (2) star irreps and the typicals with $c'=(\Lambda+\rho,\epsilon_1-\delta_1)$ being a negative integer, include the covariant tensor irreps of gl$(m|n)$, which have been studied quite extensively in the literature.[6–9]

It has been shown by Scheunert *et al.*[4] and Van der Jeugt[5] that only the type I basic classical Lie superalgebras admit finite-dimensional star irreps, thus, the results of this paper together with those of Refs. 1 and 2 provide a complete classification of the finite-dimensional star irreps of all the basic classical Lie superalgebras. We have also classified all the finite-dimensional grade star irreps of such superalgebras; these results will be reported in another publication.

ACKNOWLEDGMENTS

The authors take great pleasure in thanking Dr. A. J. Bracken for many fruitful discussions.

We also gratefully acknowledge the financial support of an ARC research grant.

[1] M. D. Gould and R. B. Zhang, "Classification of All Star and Grade Star Irreps of gl$(m|1)$," J. Math. Phys. (to be published).
[2] R. B. Zhang and M. D. Gould, "Classification of All Star and Grade Star Irreps of $C(n)$," J. Math. Phys. (to be published).
[3] V. G. Kac, Adv. Math. **26**, 8 (1977); *Lecture Notes in Mathematics*, Vol. 676 (Springer, Berlin, 1978), p. 597.
[4] M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 146 (1977).
[5] J. Van der Jeugt, J. Math. Phys. **26**, 913 (1985).
[6] P. H. Dondi and P. D. Jarvis, J. Phys. A **14**, 547 (1981).
[7] A. B. Balantekin and I. Bars, J. Math. Phys. **22**, 1149 (1981).
[8] R. J. Farmer and P. D. Jarvis, J. Phys. A **16**, 473 (1983); **17**, 2365 (1984); J. P. Hurni, J. Phys. A **20**, 5755 (1987); A. Berele and A. Regev, Adv. Math. **64**, 118 (1987).
[9] B. G. Wybourne, J. Phys. A **17**, 1573 (1984).
[10] J. W. B. Hughes, J. Math. Phys. **22**, 245 (1981); Ann. Israel Phys. Soc. **3**, 320 (1980); J. Van der Jeugt, J. Math. Phys. **25**, 3334 (1984); **28**, 758 (1987); W. Heidenreich, Phys. Lett. B **110**, 461 (1982); M. Scheunert, W. Nahm, and V. Rittenberg, J. Math. Phys. **18**, 155 (1977).
[11] M. D. Gould, Rep. Math. Phys. **27**, 73 (1989).

Automorphisms and general charge conjugations

Grzegorz Cieciura
Department of Mathematical Methods in Physics, Warsaw University, Hoża 74, 00-682 Warsaw, Poland
Igor Szczyrba
Department of Mathematics and Applied Statistics, University of Northern Colorado, Greeley, Colorado 80639

(Received 5 December 1989; accepted for publication 13 June 1990)

It is shown how outer automorphisms of a semisimple Lie algebra and automorphisms of the corresponding root system describe conjugations of general charges. New results about representations of semidirect products of non-Abelian Lie groups and about projective representations of finite groups are derived.

I. INTRODUCTION

In many symmetry theories antiparticles are described by contragredient representations. Yet, a systematic study of a particle–antiparticle (or a general charge) conjugation was never done within the representation theory of a considered symmetry Lie algebra $\mathfrak{g}$. Some authors[1] describe a conjugation of general charges using Tits' group.[2] This group, however, provides only inner automorphisms from Int $\mathfrak{g}$. So it *cannot* link, e.g., inequivalent contragredient representations.

Our main aim is to show how two automorphisms groups: the group Aut $\mathfrak{g}$ and the group $A(R)$, consisting of all automorphisms of the corresponding root system R, can be used to describe discrete symmetries mentioned above. We treat Aut $\mathfrak{g}$ as a symmetry group whose representations describe physical states. If the physical interpretation requires, we replace Aut $\mathfrak{g}$ by $\tilde{G} \rtimes$ Out $\mathfrak{g}$, where $\tilde{G}$ is the simply connected Lie group corresponding to $\mathfrak{g}$, $\rtimes$ denotes *an outer* semidirect product, and Out $\mathfrak{g}$: = Aut $\mathfrak{g}$/Int $\mathfrak{g}$, cf. Sec. V below and Ref. 3.

For simplicity of the exposition, we shall consider only a semisimple Lie algebra $\mathfrak{g}$. The generalization to a reductive case can be easily done following Sec. II in Ref. 4.

A common belief is that physical relations are implied by transformation properties of observables with respect to the group Aut $\mathfrak{g}$ or $\tilde{G} \rtimes$ Out $\mathfrak{g}$. We shall show that rather $A(R)$ transformation properties of an observable are responsible for relations among its eigenvalues.

The group Aut $\mathfrak{g}$ [resp. $A(R)$] is a semidirect product of groups Int $\mathfrak{g}$ [resp. $W(R)$] and Out $\mathfrak{g}$, i.e., the following exact short sequences split

$$1 \to \text{Int } \mathfrak{g} \to \text{Aut } \mathfrak{g} \to \text{Out } \mathfrak{g} \to 1,$$

$$1 \to W(R) \to A(R) \to \text{Out } \mathfrak{g} \to 1.$$

More detailed description of groups Aut $\mathfrak{g}$, Out $\mathfrak{g}$, and $A(R)$ is given in Appendix A.

II. REPRESENTATIONS OF SEMIDIRECT PRODUCTS

To simplify the notation we shall assume that a Lie group A is *an inner* semidirect product of its Lie subgroups W and D, i.e.,

$$A = W \cdot D, \text{ where } W \lhd A \text{ and } W \cap D = \{\mathbf{1}\}.$$

For our purpose it is sufficient to deal with a finite group D but most of the following results do not require this assumption.

Let (ρ, V) be a complex finite-dimensional representation of W. For any $d \in D$, the composition of the inner automorphism $d \cdot d^{-1}$ with ρ defines a W representation in V. The set

$$D_\rho := \{d \in D \mid \rho(d \cdot d^{-1}) \simeq \rho(\cdot)\}, \tag{2.1}$$

consisting of elements that provide representations equivalent to ρ, is a subgroup. For any $d \in D_\rho$, there exists an operator $\sigma(d) \in \mathrm{GL}(V)$ such that

$$\rho(d \cdot d^{-1}) = \sigma(d)\rho(\cdot)\sigma(d)^{-1}. \tag{2.2}$$

If ρ is irreducible, each operator $\sigma(d)$ is determined up to a scalar factor. Then any particular choice of operators $\sigma(d)$, $d \in D_\rho$, defines a *lifting* (σ, V) of a projective D_ρ representation that is characterized by a two cocycle $\mu \in Z^2(D_\rho, \mathbb{C}^*)$. More precisely, the mapping $\sigma: D_\rho \to \mathrm{GL}(V)$ satisfies

$$\sigma(d_1)\sigma(d_2) = \mu(d_1, d_2)\sigma(d_1 d_2), \tag{2.3}$$

where

$$\mu(d_1, d_2)\mu(d_1 d_2, d_3) = \mu(d_1, d_2 d_3)\mu(d_2, d_3). \tag{2.4}$$

The lifting σ is usually called in short a *projective* D_ρ representation. We shall use this simple terminology in the sequel, except for Appendix B where a detailed analysis of *both* notions is given. Due to the fact that σ fulfills condition (2.2), it will be called a *projective* ρ *intertwining* D_ρ *representation*. Of course, such ρ intertwining D_ρ representations exist also for reducible ρ's.

Now let (τ, M) be an arbitrary projective D_ρ representation corresponding to the same two cocycle $\mu(\cdot, \cdot)$ as σ, and let $L := V \otimes M^* = \mathscr{L}(M, V)$. Relations (2.2)–(2.4), and Eq. (2.3) with τ instead of σ, imply that

$$\kappa(wd)F := \rho(w)\sigma(d)F\tau(d)^{-1}, \quad w \in W, \ d \in D_\rho, \ F \in L, \tag{2.5}$$

defines an *ordinary* representation (κ, L) of the group

$$A_\rho := W \cdot D_\rho \subset A. \tag{2.6}$$

The following theorem is a refinement of the results[5–7] proven for the case of general group extensions.

Theorem 1: (a) The representation $\mathrm{Ind}_{A_\rho}^{A}(\kappa, L)$ of a

 0022-2488/90/112560-07$03.00

group A induced from a representation (κ,L) of A_ρ is irreducible iff representations (ρ,V) and (τ,M) are irreducible. (b) Any irreducible representation of A is equivalent to a representation defined in part (a).

Proof: (a) To use the Mackey's irreducibility criterion,[8] we must check that (i) the A_ρ representation (κ,L) is irreducible, (ii) for all $a\in A$ such that $a\notin A_\rho$, representations $\kappa(\cdot)$ and $\kappa(a\cdot a^{-1})$ of the group $H_a := A_\rho \cap (a^{-1}A_\rho a)$ do not contain any common irreducible component.

Ad (i): The irreducibility of ρ and Eq. (2.5) imply that an operator intertwining the W representation $(\kappa|_W,L)$ with itself is of the form $L\ni F\to F\circ J$, where $J\in\mathrm{End}\ M$. Moreover, such operator intertwines the D_ρ representation $(\kappa|_{D_\rho},L)$ iff J intertwines (τ,M). Hence, the irreducibility of τ implies that $J=\mathrm{const}\cdot id_M$.

Ad (ii): Since $W\subset H_a$, $\forall a\in A$, it is sufficient to prove that the restrictions $\kappa(\cdot)|_W$ and $\kappa(a\cdot a^{-1})|_W$ are disjoint. Due to relations

$$\kappa(a\cdot a^{-1}) = \kappa(w)\kappa(d\cdot d^{-1})\kappa(w)^{-1} \simeq \kappa(d\cdot d^{-1}), \quad a=wd,$$

the first restriction is a multiple of $\rho(\cdot)$, whereas the second one is a multiple of $\rho(d\cdot d^{-1})$. But from Eq. (2.1) it follows that

$$\rho(\cdot)\simeq\rho(d\cdot d^{-1}) \Leftrightarrow d\in D_\rho \Leftrightarrow a\in A_\rho.$$

So $a\notin A_\rho$ means that $\rho(\cdot)\not\simeq\rho(d\cdot d^{-1})$. The only if part is true because if representations ρ or τ were reducible so would be κ.

(b) Let $(\hat\pi,U)$ be an irreducible representation of A, let (ρ,V), $V\subset U$, be an irreducible W subrepresentation of (π,U), where $\pi := \hat\pi|_W$, and let $L\supset V$ carry the corresponding isotypic component, i.e., (π,L), is a multiple of (ρ,V). Any subspace $\hat\pi(a)L\subset U$, $a\in A$, carries also an isotypic component of $\hat\pi|_W$ because

$$\hat\pi(a)\hat\pi(w) = \hat\pi(awa^{-1})\hat\pi(a), \quad w\in W,$$

and W is a normal subgroup. Subspaces $\hat\pi(a)L$, $a\in A$, span U due to the irreducibility of $\hat\pi$. In addition to an isotypic W representation, the subspace L carries a representation $(\kappa,L) := (\hat\pi|_{\tilde A_\rho},L)$ of the group

$$\tilde A_\rho := \{a\in A \mid \hat\pi(a)L = L\}. \tag{2.7}$$

This means, *ex definitione*, that $(\hat\pi,U)$ is induced by (κ,L). The representation (κ,L) must be irreducible since it induces the irreducible $(\hat\pi,U)$. On the other hand, the equation

$$\pi(w)\hat\pi(d) = \hat\pi(d)\pi(d^{-1}wd), \quad d\in D,$$

implies that the isotypic W component carried by $\hat\pi(d)L$ is equivalent to a multiple of $\rho(d^{-1}\cdot d)$. It shows that D_ρ coincides with the set $\{d\in D \mid \hat\pi(d)L = L\}$. Thus $\tilde A_\rho$ defined in (2.7) is equal to A_ρ given by (2.6). Moreover, we have operators $\sigma(d)\in\mathrm{GL}(V)$ providing a projective ρ intertwining D_ρ representation (σ,V).

The space L can be represented as a tensor product $L = V\otimes M^*$, where $\dim M$ is the multiplicity of (ρ,V), i.e., $\kappa(w) = \rho(w)\otimes id_{M^*}$ for $w\in W$. From Eq. (2.2) we get

$$(\sigma(d)\otimes id_{M^*})\kappa(w) = \sigma(d)\rho(w)\otimes id_{M^*} = \rho(dwd^{-1})\sigma(d)\otimes id_{M^*} = \kappa(dwd^{-1})(\sigma(d)\otimes id_{M^*}), \quad d\in D_\rho.$$

It proves that the operator $\kappa(d^{-1})(\sigma(d)\otimes id_{M^*})$ commutes with operators $\kappa(w)$, $w\in W$. The irreducibility of ρ implies that this operator has the form $id_V\otimes\tau(d)^*$, where $\tau(d)\in\mathrm{GL}(M)$. So formula (2.5) is fulfilled and operators $\tau(d)$, $d\in D_\rho$, provide a projective D_ρ representation (τ,M) corresponding to the same two cocycle as (σ,V). The representation τ is irreducible since so is κ. ■

Corollary 1: The restriction $\pi := \hat\pi|_W$ is a direct sum of $|D/D_\rho|$ isotypic components that are multiples of inequivalent irreducible representations $\rho(d^{-1}\cdot d)$, $d\in D$, with the same multiplicity $\dim M$. □

Remarks: (1) If $A = W\cdot D$ is a *direct* product, we have $dwd^{-1} = w$. Consequently, (2.1) implies that $D_\rho = D$, whereas (2.2) implies that any operator $\sigma(d)$ commutes with all $\rho(w)$'s. Thus for an irreducible ρ we can set $\sigma(d) = id_V$ and we obtain the known result: $\hat\pi = \kappa$ is a tensor product of irreducible representations ρ and τ.

(2) In some applications, see Sec. V, it is sufficient to consider projective representations of A. Then cocycles corresponding to τ and σ need not to coincide.

(3) A representation analogous to $\mathrm{Ind}_{A_\rho}^{A}(\kappa,L)$ can be also constructed if we replace D_ρ by a proper subgroup. But in such case Theorem 1 does not hold.

(4) The group A_ρ is called the inertia group of ρ, cf. Ref. 9. □

The results obtained in this section will be applied in Secs. IV and V for the case where $D\subset\mathrm{Out}\ \mathfrak{g}$, and $W := \mathrm{Int}\ \mathfrak{g}$ or $W := W(R)$. Let us also notice that if $\mathfrak{g}$ is simple then any subgroup $D\subseteq\mathrm{Out}\ \mathfrak{g}$ has trivial second cohomology group $H^2(D,\mathbb{C}^*)$, cf. Appendix A. It means that in the case of a simple $\mathfrak{g}$, any projective D representation can be replaced, nonuniquely in general, by an ordinary representation. See Appendix B for details.

III. REALIZATIONS OF THE REPRESENTATION $\mathrm{Ind}_{A_\rho}^{A}(\kappa,L)$

The representation $\mathrm{Ind}_{A_\rho}^{A}(\kappa,L)$, where κ is given by (2.5), can be realized on a subspace of functions from $L^D := \{F(\cdot): D\to L\}$. Namely, the formula

$$({}^aF)(\delta) := \rho(\delta^{-1}w\delta)F(d^{-1}\delta), \quad a := wd\in A = W\cdot D, \quad F(\cdot)\in L^D,\ \delta\in D,$$

defines a left A action on L^D, whereas the formula

$$({}_{d_0}F)(\delta) := \sigma(d_0)F(\delta d_0)\tau(d_0)^{-1}, \quad d_0\in D_\rho,$$

provides a left D_ρ action. Both actions commute and as a result the subspace

$$U := \{F(\cdot)\in L^D \mid {}_{d_0}F(\cdot) = F(\cdot), \forall d_0\in D_\rho, \text{ i.e., } F(\cdot d_0) = \sigma(d_0)^{-1}F(\cdot)\tau(d_0)\} \tag{3.1}$$

carries an A representation $\hat\pi$, where

$$\hat\pi(a)F(\cdot) := {}^aF(\cdot), \quad a\in A,\ F(\cdot)\in U. \tag{3.2}$$

Isotypic W subspaces in U are given by

$$U_x := \{F(\cdot)\mid F(\delta) = 0 \quad \text{for } \delta\notin x\in D/D_\rho\},$$

and it is clear that $\hat\pi\simeq\mathrm{Ind}_{A_\rho}^{A}(\kappa,L)$. Compare Corollary 1.

We can also treat the group D as a space of a principal bundle with a base D/D_ρ, and define an action of the structural group D_ρ in the space L by formula

$$\kappa(d_0)F := \sigma(d_0)F\tau(d_0)^{-1}, \quad d_0 \in D_\rho, \ F \in L.$$

The associate vector bundle $E := D \times_{D_\rho} L$ consists of equivalence classes $[\delta,F]$ that are defined in $D \times L$ by the equation

$$[\delta,F] = [\delta d_0, \kappa(d_0)^{-1}F], \quad \delta \in D.$$

Sections in E are mappings $\phi:\delta \cdot D_\rho \mapsto [\delta, F(\delta)]$, where $F(\cdot) \in L^D$ satisfies

$$F(\delta d_0) = \sigma(d_0)^{-1}F(\delta)\tau(d_0).$$

So the space of sections coincides with U given by (3.1). A left A action on E is defined by

$$a \cdot [\delta,F] := [d\delta, \rho(\delta^{-1}w_1\delta)F],$$

where $a = dw_1 = dd^{-1}wd \in A$. This action provides an A action on sections that corresponds to the representation $\hat{\pi}$ defined by (3.2).

IV. ACTIONS OF A GROUP $A \subseteq A(R)$ IN ZERO-WEIGHT SPACES

In Ref. 4 we proved that zero-weight spaces of $\mathfrak{g}$ representations carry *a canonical* action Π of the corresponding Weyl group $W(R)$. We shall now show how to extend this action onto any subgroup $A \subseteq A(R)$ that contains $W(R)$.

Let $\mathfrak{g}$ be a split semisimple Lie algebra, let $\mathfrak{h}$ be its Cartan subalgebra, and let us set

$$\mathrm{Aut}_0(\mathfrak{g},\mathfrak{h}) := \mathrm{Aut}(\mathfrak{g},\mathfrak{h}) \cap \mathrm{Int}\,\mathfrak{g},$$

where $\mathrm{Aut}(\mathfrak{g},\mathfrak{h})$ consists of all $\mathfrak{g}$ automorphisms that preserve $\mathfrak{h}$. See Ref. 10. Three classes of groups: (a) subgroups G such that $\mathrm{Int}\,\mathfrak{g} \subseteq G \subseteq \mathrm{Aut}\,\mathfrak{g}$, (b) subgroups $G_\mathfrak{h}$ such that $\mathrm{Aut}_0(\mathfrak{g},\mathfrak{h}) \subseteq G_\mathfrak{h} \subseteq \mathrm{Aut}(\mathfrak{g},\mathfrak{h})$, and (c) subgroups A such that $W(R) \subseteq A \subseteq A(R)$ are in bijective correspondences. Indeed, it holds

$$G_\mathfrak{h} = G \cap \mathrm{Aut}(\mathfrak{g},\mathfrak{h}), \quad G = \mathrm{Int}\,\mathfrak{g} \cdot G_\mathfrak{h},$$

and

$$A = \epsilon(G_\mathfrak{h}), \quad G_\mathfrak{h} = \epsilon^{-1}(A),$$

where $\epsilon:\mathrm{Aut}(\mathfrak{g},\mathfrak{h}) \to A(R)$ is the canonical homomorphism. All these subgroups are semidirect products. Moreover, each of these classes can be identified with the class of subgroups in Out $\mathfrak{g}$. See Appendix A for details.

Since we are interested in representations of a group A in zero-weight spaces, we may consider only those $\mathfrak{g}$ representations all of whose irreducible components have nontrivial zero-weight spaces. Proposition 1 in Ref. 4 implies that any such representation is *radical*, i.e., it is integrable to a representation of Int $\mathfrak{g}$.

However, the existence of an Int $\mathfrak{g}$ representation (π,U) is not sufficient to determine an action of a group A in the zero-weight space of U. We can only achieve this by considering a representation $(\hat{\pi},U)$ of the group G corresponding to A. In fact, let $(\dot{\pi},U)$ be the derivative of $\pi := \hat{\pi}|_{\mathrm{Int}\,\mathfrak{g}}$, thus also of $\hat{\pi}$, and let Λ denote the weight diagram of $\dot{\pi}$, i.e., a function $\mathfrak{h}^* \to \mathbb{Z}$ whose values $\Lambda(\lambda)$ are dimensions of weight spaces $U(\lambda)$, $\lambda \in \mathfrak{h}^*$. The support $\underline{\Lambda}$ of Λ forms the set of weights. Similarly as is done in Ref. 10 for the group $W(R)$, one can show that

$$\hat{\pi}(g)U(\lambda) = U(\epsilon(g)\lambda), \quad g \in G_\mathfrak{h}, \ \lambda \in \underline{\Lambda} \subset \mathfrak{h}^*.$$

It implies that the zero-weight space $U(0)$, $0 \in \underline{\Lambda}$, is $G_\mathfrak{h}$ invariant. Moreover, the kernel $\ker \epsilon|_{G_\mathfrak{h}} = \exp_G \mathfrak{h}$ acts trivially on $U(0)$ because

$$\exp_G X|_{U(\lambda)} = e^{\langle \lambda, X \rangle} id_{U(\lambda)} \quad \text{for } X \in \mathfrak{h}.$$

So operators

$$\hat{\Pi}_a := \hat{\pi}(g)|_{U(0)} \in \mathrm{End}\,U(0), \quad a \in A, \ g \in \epsilon^{-1}\{a\} \subset G_\mathfrak{h}, \tag{4.1}$$

are correctly defined. They form an A representation $(\hat{\Pi},U(0))$ that is an extension of the canonical $W(R)$ representation $(\Pi,U(0))$ constructed in Ref. 4 by means of the representation (π,U).

Example 1: Let $d \in \mathrm{Aut}(\mathfrak{g},\mathfrak{h})$ be an involutive outer automorphism and let us set $G := \mathrm{Int}\,\mathfrak{g} \cdot D$, where $D := \{1,d\}$. Given an Int $\mathfrak{g}$ representation (ρ,V) and a parameter $0 \neq z \in \mathbb{C}$, the formulas

$$\hat{\pi}(g) := \begin{bmatrix} \rho(g) & 0 \\ 0 & \rho(dgd^{-1}) \end{bmatrix}, \quad \hat{\pi}(d) := \begin{bmatrix} 0 & z \\ z^{-1} & 0 \end{bmatrix},$$
$$g \in \mathrm{Int}\,\mathfrak{g}, \ z := z \cdot id_V,$$

define a G representation $(\hat{\pi},V \oplus V)$. The Int $\mathfrak{g}$ representation $\pi = \rho \oplus \rho(d \cdot d^{-1})$ does not depend on z, and Theorem 1 implies that $\hat{\pi}$ is irreducible iff ρ is irreducible and $\rho \not\simeq \rho(d \cdot d^{-1})$. In the latter case $\hat{\pi}$ is the only possible extension of π onto the group G. Any representation $\hat{\pi}$ determines a representation $(\hat{\Pi},V(0) \oplus V(0))$ of the group $A = W(R) \cdot \epsilon(D)$ that satisfies

$$\hat{\Pi}_w = \begin{bmatrix} \Pi_w & 0 \\ 0 & \Pi_{\delta w \delta^{-1}} \end{bmatrix}, \quad \hat{\Pi}_\delta = \begin{bmatrix} 0 & z \\ z^{-1} & 0 \end{bmatrix},$$
$$w \in W(R), \ \delta := \epsilon(d),$$

where Π denotes the canonical $W(R)$ representation acting in the zero-weight space $V(0)$ of (ρ,V). If $\mathfrak{g} = \mathrm{sl}(n,\mathbb{C})$ and $d(X) := -X^t$ (transposed) for $X \in \mathfrak{g}$ then $\rho(d \cdot d^{-1})$ is equivalent to the representation contragredient to ρ. □

Example 2: Let G, ρ, and Π be as above, and let $J \in \mathrm{End}\,V$ be an involution. If

$$\rho(d \cdot d^{-1}) = J\rho(\cdot)J^{-1}, \quad d \in D,$$

we can define *two distinct* G representations $\hat{\pi}$ in V by formulas

$$\hat{\pi}(g) := \rho(g), \quad \hat{\pi}(d) := \pm J, \quad g \in \mathrm{Int}\,\mathfrak{g}, \quad d \in D.$$

The corresponding A representations $(\hat{\Pi},V(0))$ satisfy

$$\hat{\Pi}_w = \Pi_w, \quad \hat{\Pi}_\delta = \pm J|_{V(0)}, \quad w \in W(R), \quad \delta := \epsilon(d).$$

Therefore, contrary to the $W(R)$ case, zero-weight spaces of *equivalent* $\mathfrak{g}$ representations may carry *inequivalent* representations of $A \neq W(R)$. This fact is important for applications—see Example 3 in the next section. □

Theorem 2: Let $(\hat{\Pi},U(0))$ be an arbitrary extension to a group A of the canonical $W(R)$ action $(\Pi,U(0))$ that is provided by an Int $\mathfrak{g}$ representation (π,U). Then there exists at most one extension $(\hat{\pi},U)$ of (π,U) to the group $G \subset \mathrm{Aut}\,\mathfrak{g}$ corresponding to A that provides $(\hat{\Pi},U(0))$.

Proof: Since each π irreducible component in U has a

nontrivial zero-weight space, U is spanned by vectors $\pi(g_0)u$, $g_0 \in \mathrm{Int}\,\mathfrak{g}$, $u \in U(0)$. Furthermore, $G = \mathrm{Int}\,\mathfrak{g} \cdot G_{\mathfrak{h}}$ implies that for any $g \in G$ we can choose $h \in G_{\mathfrak{h}}$ such that $g \in (\mathrm{Int}\,\mathfrak{g}) \cdot h$, i.e., there exists $h \in G_{\mathfrak{h}}$ such that $gg_0h^{-1} \in \mathrm{Int}\,\mathfrak{g}$. Consequently, we obtain

$$\hat{\pi}(g)\pi(g_0)u = \pi(gg_0h^{-1})\hat{\pi}(h)u = \pi(gg_0h^{-1})\hat{\Pi}_{\epsilon(h)}u.$$

Thus a given extension $\hat{\pi}$ is uniquely determined by representations π and $\hat{\Pi}$. ■

We shall analyze now the structure of a representation $(\hat{\Pi}, U(0))$ given by (4.1) in the case where the representation $(\hat{\pi}, U)$ of a group $G = \mathrm{Int}\,\mathfrak{g} \cdot D$, $D \subset \mathrm{Aut}(\mathfrak{g}, \mathfrak{h})$, is obtained by construction described in Sec. II. So let (ρ, V) be a representation of $\mathrm{Int}\,\mathfrak{g}$. For any automorphism $d \in \mathrm{Aut}(\mathfrak{g}, \mathfrak{h})$, the $\mathrm{Int}\,\mathfrak{g}$ representation $\rho(d \cdot d^{-1})$ acting in the space $V^d = V$ corresponds to a $\mathfrak{g}$ representation $\dot{\rho} \circ d$, and the weight spaces satisfy $V(\epsilon(d)\lambda) = V^d(\lambda)$, $\lambda \in \mathfrak{h}^*$. As a result we have

$$\Lambda_{\dot{\rho}\circ d} = \Lambda_{\dot{\rho}} \circ \epsilon(d) \quad \text{and} \quad \underline{\Lambda}_{\dot{\rho}\circ d} = \epsilon(d)^{-1}\underline{\Lambda}_{\dot{\rho}}.$$

Moreover, for the canonical $W(R)$ representations Π and Π^d acting, respectively, in zero-weight spaces of representations ρ and $\rho(d \cdot d^{-1})$, we have

$$\Pi^d_w = \Pi_{\epsilon(d)w\epsilon(d)^{-1}}, \quad w \in W(R).$$

Now if an operator $\sigma(d) \in \mathrm{GL}(V)$ fulfills Eq. (2.2) then $\dot{\rho} \circ d = \sigma(d)\dot{\rho}(\cdot)\sigma(d)^{-1}$. Hence, we get

$$\sigma(d)V(\lambda) = V(\epsilon(d)\lambda) = V^d(\lambda).$$

Accordingly, the space $V(0)$ is $\sigma(d)$ invariant. All that implies the following facts.

(1) The group $D_\rho = \{d \in D \,|\, \dot{\rho} \circ d \simeq \dot{\rho}\}$, whereas $\epsilon(D_\rho) = \{\delta \in \epsilon(D) \,|\, \Lambda_{\dot{\rho}} \circ \delta = \Lambda_{\dot{\rho}}\}$; compare Lemma 2 in Ref. 3.

(2) The zero-weight space $V(0)$ is $\sigma(D_\rho)$ invariant; we denote the corresponding D_ρ representation by $(\sigma_0, V(0))$.

(3) The representation $(\hat{\Pi}, U(0))$ of the group $A = W(R) \cdot \epsilon(D)$ corresponding to $G = \mathrm{Int}\,\mathfrak{g} \cdot D$ can be *directly* obtained from the canonical $W(R)$ representation $(\Pi, U(0))$ by the construction described in Sec. II if we substitute $W(R)$ for W, $(\Pi, V(0))$ for (ρ, V), and $(\sigma_0, V(0))$ for (σ, V); we identify here $\epsilon(D)$ with D.

Let us notice that the group D_ρ (that has to be used in order to obtain $\hat{\Pi}$ directly from Π) is, in general, a *proper* subgroup of the group

$$D_\Pi := \{d \in D \,|\, \Pi(\epsilon(d)(\cdot)\epsilon(d^{-1})) \simeq \Pi(\cdot)\},$$

which is given by Definition (2.1) if we substitute Π for ρ. Indeed, zero-weight spaces of *inequivalent* $\mathfrak{g}$ representations may carry *equivalent* $W(R)$ actions.[4,11] See also Remark 3 in Sec. II.

V. ACTIONS OF THE GROUP $A \subseteq A(R)$ AND PHYSICAL RELATIONS

Our results from Ref. 4 show that the canonical $W(R)$ representation Π in zero-weight spaces is responsible for many physical relations that traditionally are believed to be implied by $\mathfrak{g}$ transformation properties of observables.

More precisely, let $\tilde{G}$ be a simply connected Lie group corresponding to a split pair $(\mathfrak{g}, \mathfrak{h})$, let (π, U) be a $\tilde{G}$ representation describing physical states, and let $\dot{\pi}$ be the derivative of π. The zero-weight space End $U(0)$ of the $\mathfrak{g}$ representation ad $\dot{\pi}$, acting in the space End U, coincides with the commutant $\dot{\pi}(\mathfrak{h})'$ of the image $\dot{\pi}(\mathfrak{h}) \subset \mathrm{End}\,U$. The operators from $\dot{\pi}(\mathfrak{h})$ represent *basic observables* given by the Cartan subalgebra $\mathfrak{h}$. Consequently, the assumption that an observable $\mathcal{O} \in \mathrm{End}\,U$ commutes with this basic observables implies that the canonical $W(R)$ action Π is defined on $\mathcal{O}$. Moreover, in the considered case the canonical $W(R)$ action is implemented by operators from End U since we have

$$\Pi_w(\mathcal{O}) = \pi(g)\mathcal{O}\pi(g)^{-1}, \quad w \in W(R), \tag{5.1}$$

where $g \in (\epsilon \circ \mathrm{Ad})^{-1}\{w\} \subset \tilde{G}$ and $\mathrm{Ad}: \tilde{G} \to \mathrm{Int}\,\mathfrak{g}$.

According to Proposition 2 in Ref. 4, it follows from (5.1) that any particular $W(R)$ transformation properties of an observable $\mathcal{O}$ provide specific linear relations among its eigenvalues or, more generally, among its diagonal elements. These relations depend on $\mathfrak{g}$ transformation properties of $\mathcal{O}$ only in a sense that the zero-weight space of a given *irreducible* $\mathfrak{g}$ component in ad $\dot{\pi}$ carries a specific, in general *reducible*,[11] $W(R)$ representation. It is important to realize that because *the same* irreducible $W(R)$ representations may appear in zero-weight spaces of *inequivalent* irreducible components of ad $\dot{\pi}$, the same physical relations may be imposed by *distinct* $\mathfrak{g}$ transformation properties of a given observable.

As we explained in Sec. IV, to define the canonical $W(R)$ action $(\Pi, \mathrm{End}\,U(0))$ we do not need a $\tilde{G}$ representation acting in the space End U. In fact, it is sufficient to have only an $\mathrm{Int}\,\mathfrak{g}$ representation acting in the radical subspace of End U, i.e., in the subspace carrying all these irreducible subrepresentations of ad $\dot{\pi}$ that have nontrivial zero-weight spaces. Note that if the representation π is irreducible then the space End U is radical.

Now if we want to consider a bigger group $A = W(R) \cdot \epsilon(D)$, where $D \neq \{1\}$, the situation becomes more complicated. Namely, even if the $\mathrm{Int}\,\mathfrak{g}$ representation acting in the radical subspace of End U can be extended to a representation of the group $G = \mathrm{Int}\,\mathfrak{g} \cdot D$, the corresponding A action *need not* be implemented by operators from End U. However if, for instance, (π, U) can be extended to a representation $(\hat{\pi}, U)$ of the group $\tilde{G} \rtimes D$ then $\hat{\pi}$ implements a natural A action on End $U(0)$, i.e., any D action on states from U determines a D action on $\dot{\pi}(\mathfrak{h})'$ that provides additional relations among eigenvalues of observables.

Let us notice that an ordinary G representation in the radical part of End U, and hence an A action on observables, can also be implemented by a *projective* representation of $\tilde{G} \rtimes D$ (or G) in U, compare Remark 2 in Sec. II.

Since a D action in U can link both: equivalent and inequivalent $\mathfrak{g}$ irreducible subspaces, it is suitable to describe general charge conjugations. Moreover, contrary to the case of the group $W(R)$, zero-weight spaces of *equivalent* $\mathfrak{g}$ components in End U may carry *inequivalent* A representations, cf. Example 2. This fact enables two observables to have the same $\mathfrak{g}$ transformation properties and, at the same time, to provide distinct physical relations, e.g., the opposite change of signs under a transformation from D. To make it clear let us consider the following example.

Example 3: In a flavor theory based on the Lie algebra

$\mathfrak{g} = \mathrm{su}(n)$, antiparticles are described by contragredient representations. Thus transformations of general charges corresponding to the particle–antiparticle conjugation are characterized by the group $D \simeq \mathrm{Out}\,\mathfrak{g} \simeq Z_2$. In particular, for any self-contragredient $\tilde{G}$ representation (π, U) describing particle and antiparticle states, there exists a natural Z_2 action on the commutant $\dot{\pi}(\mathfrak{h})'$ that is implemented by operators from End U. Furthermore, if π is not trivial, the representation (ad $\dot{\pi}$, End U) contains the adjoint representation (ad,g) at least twice,[12] and basic observables, i.e., elements from $\dot{\pi}(\mathfrak{h}) \subset \dot{\pi}(\mathfrak{h})'$, transform according to the representation ad.

The Z_2 action on basic observables leads to an equation for eigenvalues of the type $e = -\bar{e}$. This explains why isospin, electric charge, etc., change signs under the particle–antiparticle conjugation. At the same time, it is possible to choose in this case a Z_2 invariant mass operator (i.e., providing *equal* masses for particles and antiparticles) and such that its main component transforms also according to the representation ad. For instance, if states are described by the adjoint representation, i.e., $(\dot{\pi}, U) = (\mathrm{ad}, \mathfrak{g})$, the representation (ad $\dot{\pi}$,End U) contains ad exactly twice. If we identify the space End U with the tensor product $U \otimes U$, the $\mathfrak{g}$ isomorphic subspace that contains basic observables is included in the skew-symmetrical product $U \wedge U$, whereas the second $\mathfrak{g}$ isomorphic subspace can be chosen from the symmetrical product $U \odot U$, cf. Ref. 13. The group Z_2 acts in $U \otimes U$ by permuting factors and leads to equations $e = \pm \bar{e}$. □

The next example provides more sophisticated mathematical structure.

Example 4: The Lie algebra $\mathfrak{g} = \mathrm{so}(8)$ is the only simple Lie algebra with a non-Abelian group $\mathrm{Out}\,\mathfrak{g} \simeq S_3$. A physical theory based on so(8) can link two equivalent $\mathfrak{g}$ irreducible subspaces of an irreducible representation $(\hat{\pi}, U)$ of $\tilde{G} \rtimes S_3$ because S_3 has a two-dimensional irreducible representation (τ, M), compare Example 1.

Moreover, since there are three inequivalent eight-dimensional irreducible representations of so(8), this algebra can describe conjugations of a three-valued general charge; and the action of the group S_3 can be used to characterize chargeless states, e.g., colorless particles.[14] Finally, the existence of three inequivalent one-dimensional Z_3 representations enables a physical model to include three distinct phase transformations. □

ACKNOWLEDGMENTS

The research was supported in part by the UNC Graduate School Grant. The authors would also like to thank Dr. C. Cotter for his remarks and discussions.

APPENDIX A: THE STRUCTURE OF THE GROUPS AUT $\mathfrak{g}$ AND $A(R)$

Let $\mathfrak{h}$ be a Cartan subalgebra of a semisimple split Lie algebra $\mathfrak{g}$, and let $\mathrm{Aut}(\mathfrak{g},\mathfrak{h})$ consist of all $\mathfrak{g}$ automorphisms that preserve $\mathfrak{h}$. The canonical homomorphism[10] $\epsilon: \mathrm{Aut}(\mathfrak{g},\mathfrak{h}) \to A(R)$ is defined by formula

$$\epsilon(s)\alpha = \alpha \circ s^{-1}|_{\mathfrak{h}}, \quad s \in \mathrm{Aut}(\mathfrak{g},\mathfrak{h}), \quad \alpha \in R \subset \mathfrak{h}^*.$$

The subgroup

$$\mathrm{Aut}_0(\mathfrak{g},\mathfrak{h}) := \mathrm{Aut}(\mathfrak{g},\mathfrak{h}) \cap \mathrm{Int}\,\mathfrak{g}$$

is normal and coincides with $\epsilon^{-1}(W(R))$. In Ref. 10 it is proved that $\mathrm{Aut}(\mathfrak{g},\mathfrak{h})$ is a semidirect product of subgroups $\mathrm{Aut}_0(\mathfrak{g},\mathfrak{h})$ and

$$\Delta := \mathrm{Aut}(\mathfrak{g},\mathfrak{h},B,(X_\alpha)_{\alpha\in B}) \simeq \mathrm{Out}\,\mathfrak{g}.$$

This implies that $A(R)$ is a semidirect product of groups $W(R)$ and $\epsilon(\Delta) \simeq \Delta$.

To prove that Aut $\mathfrak{g}$ is a semidirect product of groups Int $\mathfrak{g}$ and Δ, note that Int $\mathfrak{g}$ acts transitively on a set of all Cartan subalgebras of $\mathfrak{g}$. Thus for any automorphism $s \in \mathrm{Aut}\,\mathfrak{g}$, there exists $r \in \mathrm{Int}\,\mathfrak{g}$ such that $r \circ s \in \mathrm{Aut}(\mathfrak{g},\mathfrak{h})$, i.e.,

$$\mathrm{Aut}\,\mathfrak{g} = \mathrm{Int}\,\mathfrak{g} \cdot \mathrm{Aut}(\mathfrak{g},\mathfrak{h}).$$

Since $\mathrm{Aut}(\mathfrak{g},\mathfrak{h}) = \mathrm{Aut}_0(\mathfrak{g},\mathfrak{h}) \cdot \Delta$, we obtain hence

$$\mathrm{Aut}\,\mathfrak{g} = \mathrm{Int}\,\mathfrak{g} \cdot \Delta.$$

Moreover, we also have

$$\mathrm{Int}\,\mathfrak{g} \cap \Delta = \mathrm{Int}\,\mathfrak{g} \cap \mathrm{Aut}(\mathfrak{g},\mathfrak{h}) \cap \Delta = \mathrm{Aut}_0(\mathfrak{g},\mathfrak{h}) \cap \Delta = \{\mathbf{1}\},$$

that concludes the proof.

The subgroups listed in Sec. IV are semidirect products because if an extension

$$1 \to W \to A \to A/W \to 1$$

splits, i.e., $A = W \cdot D$, $D \subset A$, $W \lhd A$, $W \cap D = \{\mathbf{1}\}$, then for any subgroup $A' \subset A$ containing the group W, the extension

$$1 \to W \to A' \to A'/W \to 1$$

splits too. Indeed, for $D' := D \cap A'$, it holds $W \cap D' = \{\mathbf{1}\}$ and $W \cdot D' \subset A'$. To show that $A' \subset W \cdot D'$, let us notice that if $a \in A' \subset A$ then $a = wd$, where $w \in W$ and $d \in D$. Accordingly, we get $d = w^{-1}a \in A' \cdot A'$, i.e., $d \in D'$.

Let us change slightly our notation. Let $\tilde{\mathfrak{g}}$ denote a semisimple Lie algebra, let $\mathfrak{g}$ be its simple component, and let $n \times \mathfrak{g} := \mathfrak{g} \oplus \cdots \oplus \mathfrak{g}$ denote the corresponding isotypic constituent. The groups Aut $\tilde{\mathfrak{g}}$, Int $\tilde{\mathfrak{g}}$, Out $\tilde{\mathfrak{g}}$, $A(R_{\tilde{\mathfrak{g}}})$, $W(R_{\tilde{\mathfrak{g}}})$ are direct sums of groups $\mathrm{Aut}(n \times \mathfrak{g}), \ldots, W(R_{n\times\mathfrak{g}})$, respectively.

Lemma 1: Let $\mathfrak{g}$ be a simple Lie algebra. Then $\mathrm{Aut}(n \times \mathfrak{g}) = (\mathrm{Aut}\,\mathfrak{g})^n \rtimes S_n$.

Proof: For any automorphism $s \in \mathrm{Aut}(n \times \mathfrak{g})$ and $i \in \overline{1,n}$, an epimorphism $s_i := pr_i \circ s : n \times \mathfrak{g} \to \mathfrak{g}$ maps a component of $n \times \mathfrak{g}$ on a subalgebra in $\mathfrak{g}$. Because s_i images of all components commute one with another, they must be equal to $\mathfrak{g}$ or 0 (an endomorphism of a simple Lie algebra is either onto or trivial). Therefore all but one component must be mapped on 0, i.e., s_i defines an automorphism $r_i \in \mathrm{Aut}\,\mathfrak{g}$. In consequence, the automorphism s has the form

$$s(X_1 \oplus \cdots \oplus X_n) = r_1 X_{\sigma^{-1}(1)} \oplus \cdots \oplus r_n X_{\sigma^{-1}(n)}, \quad \sigma \in S_n. \ \blacksquare$$

Corollary 2: For any simple split pair $(\mathfrak{g},\mathfrak{h})$ it holds

(a) $\mathrm{Aut}(n \times \mathfrak{g}, n \times \mathfrak{h}) \simeq \mathrm{Aut}(\mathfrak{g},\mathfrak{h})^n \rtimes S_n$,

(b) $A(R_{n\times\mathfrak{g}}) \simeq A(R_{\mathfrak{g}})^n \rtimes S_n$,

(c) $\mathrm{Aut}_0(n \times \mathfrak{g}, n \times \mathfrak{h}) \simeq \mathrm{Aut}_0(\mathfrak{g},\mathfrak{h})^n$,

(d) $W(R_{n\times\mathfrak{g}}) \simeq W(R_{\mathfrak{g}})^n$,

(e) $\mathrm{Out}(n \times \mathfrak{g}) \simeq (\mathrm{Out}\,\mathfrak{g})^n \rtimes S_n$,

where

$$\mathrm{Out}\,\mathfrak{g} \simeq S_3, Z_2, \text{ or } \{\mathbf{1}\}.$$

Proof: The assertions follow from Lemma 1 and the fact that $\mathrm{Int}(n\times\mathfrak{g})$ is isomorphic to $(\mathrm{Int}\,\mathfrak{g})^n$. ■

For a simple Lie algebra $\mathfrak{g}$, the second cohomology group $H^2(D,\mathbb{C}^*)$, see Appendix B, of any subgroup $D\subseteq \mathrm{Out}\,\mathfrak{g}$ is trivial due to the following lemma.

Lemma 2:

(a) $H^2(Z_n,\mathbb{C}^*)=\{\mathbf{1}\}$,

(b) $H^2(S_3,\mathbb{C}^*)=\{\mathbf{1}\}$,

(c) $H^2(S_n,\mathbb{C}^*)=Z_2$ for $n>3$.

Proof: (a) A projective representation (ρ,V) of $Z_n=\{0,1,...,n-1\}$ is determined by an endomorphism $\rho(1)\in\mathrm{GL}(V)$ such that $\rho(1)^n=c\cdot id_V$, $c\in\mathbb{C}^*$. Replacing operators $\rho(k)$, $k\in Z_n$, by operators $b^k\cdot\rho(1)^k$, where $b\in\mathbb{C}^*$ and $cb^k=1$, we get an ordinary representation.

(b) The group S_3 is generated by two transpositions $\tau_1=(1,2)$ and $\tau_2=(2,3)$ that satisfy $\tau_1^2=\tau_2^2=(\tau_1\tau_2)^3=\mathbf{1}$. So a projective S_3 representation (ρ,V) is defined by endomorphisms $P:=\rho(\tau_1)$, $Q:=\rho(\tau_2)\in\mathrm{End}\,V$ such that P^2, Q^2, and $(PQ)^3$ are scalar operators. Multiplying P and Q by scalar coefficients we can substitute them by operators fulfilling $P^2=Q^2=id_V$. Hence, for $(PQ)^3=b\cdot id_V$ we have

$$(PQ)^3QPQ=bQPQ\Rightarrow PQP=bQPQ,$$

$$PQP(PQ)^3=bPQP\Rightarrow QPQ=bPQP.$$

As a result we get $b=1$ or $b=-1$. Replacing in the second case P by $-P$, we obtain $(PQ)^3=id_V$, i.e., an ordinary representation. Compare Ref. 15.

(c) The group S_n is generated by transpositions $\tau_i=(i,i+1)$, $i\in\overline{1,n-1}$, satisfying

$$\tau_i^2=(\tau_j\tau_{j+1})^3=(\tau_k\tau_l)^2=\mathbf{1}\quad\text{for } i,k,l\in\overline{1,n-1},$$

$$j\in\overline{1,n-2},\quad |k-l|\geqslant 2.$$

Therefore, a projective S_n representation (ρ,V) is defined by operators $T_i\in\mathrm{End}\,V$ such that

$$T_i^2=id_V,\quad (T_jT_{j+1})^3=b_j\cdot id_V,\quad (T_kT_l)^2=c_{kl}\cdot id_V.$$

Computations done in part (b) for $P=T_j$ and $Q=T_{j+1}$ imply that $b_j=\pm 1$. By a proper change of signs for T_j's we obtain $b_j=1$. Now from the relations

$$id_V=(T_kT_lT_k)^2=(c_{kl}T_l)^2=c_{kl}^2\cdot id_V,$$

it follows that $c_{kl}=\pm 1$. Moreover, for $|k-l|\geqslant 2$ there exists $\sigma\in S_n$ such that $\sigma\tau_k\sigma^{-1}=\tau_1$ and $\sigma\tau_l\sigma^{-1}=\tau_3$. So we get

$$ST_kS^{-1}=dT_1,\quad ST_lS^{-1}=d'T_3,$$

where $S\in\mathrm{End}\,V$ corresponds to σ, and $d^2=d'^2=1$. It shows that

$$S(T_kT_l)^2S^{-1}=(dT_1d'T_3)^2.$$

Hence, constants $c_{kl}=c_{13}=c$ do not depend on k,l. Therefore, there exist at most two classes of projective S_n representations corresponding to $c=\pm 1$. This means that $|H^2(S_n,\mathbb{C}^*)|\leqslant 2$.

To prove that the representation with $c=-1$ exists, let us choose in a Euclidean space E a set of vectors $e_1,...,e_{n-1}$ such that $(e_i|e_j)$ is equal to 1, $-\frac{1}{2}$, or 0 if $|i-j|=0,1$, or $\geqslant 2$, respectively. Then, in the Clifford algebra $\mathscr{C}(E)$, we obtain $e_i^2=(e_je_{j+1})^3=-(e_ke_l)^2=1$. Thus for any representation $\phi:\mathscr{C}(E)\to\mathrm{End}\,V$, the operators $T_i=\phi(e_i)$ satisfy the required relations. ■

APPENDIX B: PROJECTIVE REPRESENTATIONS

Let D be a finite group. The group $C^2(D,\mathbb{C}^*):=(\mathbb{C}^*)^{D\times D}$ consisting of two cochains contains a subgroup

$$Z^2(D,\mathbb{C}^*):=\{\mu\in(\mathbb{C}^*)^{D\times D}\,|\,\mu(d_1,d_2)\mu(d_1d_2,d_3)=\mu(d_1,d_2d_3)\mu(d_2,d_3)\}$$

of two cocycles that in turn contains a subgroup of two coboundries

$$B^2(D,\mathbb{C}^*):=\{\mu\in(\mathbb{C}^*)^{D\times D}\,|\,\exists\nu\in(\mathbb{C}^*)^D:\mu(d_1,d_2)=\nu(d_1)\nu(d_2)/\nu(d_1d_2)\}.$$

Two cocycles μ and μ' are equivalent if their quotient is a two coboundry, and

$$H^2(D,\mathbb{C}^*):=Z^2(D,\mathbb{C}^*)/B^2(D,\mathbb{C}^*)$$

is called the second cohomology group.[16,9]

In this section by a projective representation of D, acting in a finite-dimensional vector space V, we mean a homomorphism

$$\mathscr{S}:D\to PGL(V).$$

Any projective D representation $\mathscr{S}$ has a *lifting*

$$\sigma:D\to\mathrm{GL}(V)$$

such that

$$\sigma(d_1)\sigma(d_2)=\mu(d_1,d_2)\sigma(d_1d_2),\tag{B1}$$

where $\mu\in Z^2(D,\mathbb{C}^*)$ due to the associativity of a product $\sigma(d_1)\sigma(d_2)\sigma(d_3)$.

Another lifting $\sigma'=\nu\sigma$, $\nu\in(\mathbb{C}^*)^D$, defines an equivalent two cocycle

$$\mu'(d_1,d_2)=[\nu(d_1)\nu(d_2)/\nu(d_1d_2)]\mu(d_1,d_2).$$

So with any projective representation we can associate an element $[\mu]\in H^2(D,\mathbb{C}^*)$. In particular, the unit element in the group $H^2(D,\mathbb{C}^*)$ corresponds to a class of projective representations that can be lifted to ordinary representations. (Note that two ordinary representations provide the same projective representation iff they differ by a character of D.) Furthermore, the tensor multiplication of projective representations becomes the multiplication of elements in $H^2(D,\mathbb{C}^*)$, whereas the contragredient representation is related to the inverse element.

We shall prove now that every element $[\mu]\in H^2(D,\mathbb{C}^*)$ is associated with a nonempty class of projective representations.

Lemma 3: (a) For any two cocycle $\mu\in Z^2(D,\mathbb{C}^*)$, the formula

$$(\rho_\mu(c)f)(d):=\mu(d,c)f(dc),\quad f\in\mathbb{C}^D,\ c,d\in D,$$

defines a lifting $(\rho_\mu,\mathbb{C}^D)$ of a projective D representation $(\mathscr{R}_\mu,\mathbb{C}^D)$ that is characterized by $[\mu]\in H^2(D,\mathbb{C}^*)$.

(b) Equivalent two cocycles μ and μ' provide equivalent projective representations $\mathscr{R}_\mu$ and $\mathscr{R}_{\mu'}$.

Proof: (a) Let us set $f' := \rho_\mu(c)f$. Then we have

$$\begin{aligned}(\rho_\mu(c_1)\rho_\mu(c)f)(d) &= (\rho_\mu(c_1)f')(d)\\ &= \mu(d,c_1)f'(dc_1)\\ &= \mu(d,c_1)\mu(dc_1,c)f(dc_1c)\\ &= \mu(d,c_1c)\mu(c_1,c)f(dc_1c)\\ &= \mu(c_1,c)(\rho_\mu(c_1c)f)(d).\end{aligned}$$

(b) The liftings corresponding to equivalent two cocycles μ and μ' are related by an endomorphism $F\in\mathrm{End}\,\mathbb{C}^D$, where $(Ff)(c) := \nu(c)f(c)$, in the following way:

$$(F\rho_\mu(c)F^{-1}f)(d) = 1/\nu(c)\cdot(\rho_{\mu'}(c)f)(d).$$

The operators $F\rho_\mu(c)F^{-1}$ and $\rho_{\mu'}(c)$ give rise to the same projective representation because they differ by a scalar factor $\nu(c)$. ■

A mapping $\sigma: D\to \mathrm{GL}(V)$ satisfying the condition (B1) will be called a μ *representation* of D. Similarly, as for ordinary representations, such notions as direct sum, irreducibility, intertwining operator, or equivalence are well defined for μ representations.

Lemma 4: An irreducible μ representation (σ,V) of D, $\mu\in Z^2(D,\mathbb{C}^*)$, is contained in the μ representation $(\rho_\mu,\mathbb{C}^D)$ with the multiplicity equal to $\dim V$.

Proof: Schur's Lemma holds also for μ representations and any μ representation is fully reducible.[17,15] So the considered multiplicity coincides with the dimension of the space $\mathcal{L}_D(V,\mathbb{C}^D)$ consisting of operators intertwining (σ,V) and $(\rho_\mu,\mathbb{C}^D)$. It remains to prove that $\mathcal{L}_D(V,\mathbb{C}^D)\simeq V$.

To this end let us note that any intertwining operator $F: V\to\mathbb{C}^D$ has the form

$$(Fv)(d) = \langle\psi_d,v\rangle,\quad v\in V,\ d\in D,\ \psi_d\in V^*,$$

and that the intertwining property $F\circ\sigma(c) = \rho_\mu(c)\circ F$, $c\in D$, implies

$$\langle\psi_d,\sigma(c)v\rangle = \mu(d,c)\langle\psi_{dc},v\rangle.$$

Hence, setting $d = \mathbf{1}\in D$, we obtain $\langle\psi_c,v\rangle = \langle\psi_1,\sigma(c)v\rangle$. Thus the linear mapping

$$\mathcal{L}_D(V,\mathbb{C}^D)\ni F\mapsto\psi_1\in V^*$$

is injective. It is also surjective because any $\psi\in V^*$ defines an intertwining operator by the formula

$$(Fv)(c) := \langle\psi,\sigma(c)v\rangle.$$ ■

Now note that for a given projective representation $\mathcal{S}$ and a corresponding μ representation σ, the group

$$\Gamma(\mathcal{S}) := \{\chi\in\mathrm{Hom}(D,\mathbb{C}^*)\,|\,\chi\cdot\sigma\simeq\sigma\}$$

does not depend on a choice of σ. Moreover, for a fixed two cocycle μ, the number of nonequivalent μ representations that correspond to the representation $\mathcal{S}$ coincides with the index

$$n(\mathcal{S}) := (\mathrm{Hom}(D,\mathbb{C}^*):\Gamma(\mathcal{S})).$$

In fact, two μ representations σ and σ' provide equivalent projective representations iff $\sigma'\simeq\chi\cdot\sigma$ for a certain character $\chi\in\mathrm{Hom}(D,\mathbb{C}^*)$.

Corollary 3: Any irreducible projective D representation $\mathcal{S}$ corresponding to an element $[\mu]\in H^2(D,\mathbb{C}^*)$ is contained in the D projective representation $\mathcal{R}_\mu$ [defined in part (a) of Lemma 3] with the multiplicity $n(\mathcal{S})\cdot\dim\mathcal{S}$. So the rank of D is given by

$$|D| = \sum_{\mathcal{S}} n(\mathcal{S})(\dim\mathcal{S})^2,$$

where $\mathcal{S}$ assumes all irreducible pairwise inequivalent projective representations corresponding to a fixed class in $H^2(D,\mathbb{C}^*)$. □

[1] R. V. Moody and J. Patera, J. Math. Phys. **25**, 2838 (1984).
[2] J. Tits, *Théorie des Groúpes*, "Résume des cours et travaux," Annuaire du Collége de France, 1980–1981.
[3] I. Szczyrba, Nucl. Phys. B, Proc. Suppl. **6**, 440 (1989).
[4] G. Cieciura and I. Szczyrba, J. Math. Phys. **28**, 2115 (1987).
[5] A. H. Clifford, Ann. Math. Ser. 2 **38**, 533 (1937).
[6] J.-P. Serre, *Linear Representations of Finite Groups* (Springer, New York, 1977).
[7] G. Karpilovsky, Proc. Edinburgh Math. Soc. **27**, 7 (1984).
[8] G. W. Mackey, *Induced Representations of Groups and Quantum Mechanics* (Benjamin, New York, 1968).
[9] G. Karpilovsky, *Clifford Theory for Group Representations* (North-Holland, Amsterdam, 1989).
[10] N. Bourbaki, *Éléments de Mathématique, Groupes et Algèbres de Lie* (Hermann, Paris, 1975).
[11] G. Cieciura and I. Szczyrba, J. Math. Phys. **28**, 2128 (1987).
[12] I. Szczyba, Lett. Math. Phys. **4**, 249 (1980).
[13] L. Michel, *Applications of Group Theory in Quantum Physics, Algebraic Aspects*, Lecture Notes in Physics, Vol. 6 (Springer, Berlin, 1970).
[14] I. Szczyrba, "An SO (8)-subquark Model," to be published.
[15] I. Schur, J. Reine Angew. Math. **139**, 155 (1911).
[16] A. A. Kirillov, *Elements of the Theory of Representations* (Springer, Berlin, 1976).
[17] Ch. W. Curtis and W. Reiner, *Representation Theory of Finite Groups and Associative Algebras* (Wiley, New York, 1962).

On a Fourier integral over SO(3)

Gaoyuan Wei and B. E. Eichinger
Department of Chemistry, BG-10, University of Washington, Seattle, Washington 98195

(Received 23 January 1990; accepted for publication 11 July 1990)

A two-matrix function of general interest in the areas of configuration statistics of macromolecules, number theory, harmonic analysis, and multivariate statistics is studied. The function is defined as a Fourier integral over SO(3), the Lie group of orthogonal 3×3 matrices with unit determinant. This six-variable function is first expressed as a product of a three-variable function and an exponential function of an additional variable, thereby reducing the total number of independent variables by 2. The new function with three parameters is expressible either as a double integral or as a series in one of the variables with the coefficients being polynomials in the other two. A special, nontrivial case where one of three arguments of the function takes a particular value is explored thoroughly. The resulting two-variable function is real valued and is an oscillating function of one of the variables when the other is fixed. When this function is expanded as a power series in one of the two variables, it generates polynomials in the other variable. Numerical analysis of this series shows it to be rapidly convergent and it is of practical use in the numerical evaluation of the function. Although the connection between these newly found polynomials and zonal polynomials has not been investigated, the parametrization for the four new variables of the two-matrix function studied may well prove useful in the effective numerical evaluation of the function when expressed alternatively as a series in zonal polynomials with an exponential part factored out.

I. INTRODUCTION

A Fourier integral over SO(3), the special orthogonal group, of the form

$$F(\mathbf{a},\mathbf{b}) = \frac{1}{8\pi^2}\int_{SO(3)} \mathrm{etr}(i\mathbf{a}\mathbf{h}\mathbf{b}\mathbf{h}')\,d\,\mathbf{h}, \tag{1}$$

where $\mathrm{etr}(\mathbf{a}) = \exp[\mathrm{tr}(\mathbf{a})]$, arises in the general formulation of shape distribution functions for Gaussian molecules[1-5] as well as in number theory,[6] harmonic analysis,[7] and multivariate statistics.[8,9] Here $\mathbf{h}\in SO(3)$, $\mathbf{h}'$ is the transpose of $\mathbf{h}$, $d\mathbf{h}$ is the unnormalized Haar measure on SO(3), $\mathbf{a} = \mathrm{diag}(a_1,a_2,a_3)$ and $\mathbf{b} = \mathrm{diag}(b_1,b_2,b_3)$ with $a_i,b_j\in\mathbf{R}$. That the function $F(\mathbf{a},\mathbf{b})$ defined above is essentially a six-dimensional Fourier integral can be seen from Hua's work[7] on integrals over the orthogonal group. A common approach to evaluate the above integral has been to express $F(\mathbf{a},\mathbf{b})$ as a series in zonal polynomials since $F(\mathbf{a},\mathbf{b}) = {}_0F_0^{(3)}(i\mathbf{a},\mathbf{b})$, where ${}_0F_0^{(m)}(\mathbf{X},\mathbf{Y})$ is the two-matrix hypergeometric function.[8-10] Another study[11] on $F(\mathbf{a},\mathbf{b})$ made use of a parametrization of $\mathbf{h}$ in terms of its skew symmetric part, and subsequently expressed the function as a double series. This approach yields an exact asymptotic expansion of the function. Similar evaluation of $F(-i\mathbf{a},\mathbf{b})$ by use of a different parametrization of $\mathbf{h}$ has also appeared in the literature with the result expressed in a series of multiple sums of Beta functions.[1]

In this paper, we begin by expressing $F(\mathbf{a},\mathbf{b})$ as a new function of four variables instead of the original six matrix components. One component of this factorization, $(1/3)\mathrm{tr}(\mathbf{a})\mathrm{tr}(\mathbf{b})$, has appeared before.[3,6,11] We then study in detail the nontrivial part of this redefined four-argument function, that is, the part of $F(\mathbf{a},\mathbf{b})$ with $\exp[i(1/3)\mathrm{tr}(\mathbf{a})\mathrm{tr}(\mathbf{b})]$ factored out. We have found that this part of $F(\mathbf{a},\mathbf{b})$, being itself an entirely new and well defined three-variable function, can be conveniently expressed either as a simple double integral or as a series in one of the three variables with the coefficients being symmetric polynomials in the other two. These polynomials can easily be written down to arbitrary degree, especially with the use of symbolic integration on a computer. The application of the double-integral representation to the evaluation of shape distribution functions for Gaussian macromolecules[5] has been one of its most powerful advantages over many of the earlier approaches.

II. THEORY

We first make use of the classical Euler angles parametrization of the group SO(3), i.e.,

$$\mathbf{h} = \begin{pmatrix} c_\alpha c_\gamma - s_\alpha c_\beta s_\gamma & c_\alpha s_\gamma - s_\alpha c_\beta c_\gamma & s_\alpha s_\beta \\ -s_\alpha c_\gamma - c_\alpha c_\beta s_\gamma & -s_\alpha s_\gamma + c_\alpha c_\beta c_\gamma & c_\alpha s_\beta \\ s_\beta s_\gamma & -s_\beta c_\gamma & c_\beta \end{pmatrix}, \tag{2}$$

where $c_\alpha = \cos\alpha$, $s_\beta = \sin\beta$, etc., and then transform the original Lebesgue integral to a multiple Riemann one with the invariant measure given by[12] $d\,\mathbf{h} = \sin\beta\,d\alpha\,d\beta\,d\gamma$, i.e.,

$$F(\mathbf{a},\mathbf{b}) = \frac{1}{8\pi^2}\int_0^{2\pi} d\alpha \int_0^{\pi} \sin\beta\,d\beta \int_0^{2\pi} d\gamma\,\mathrm{etr}(i\mathbf{a}\mathbf{h}\mathbf{b}\mathbf{h}'). \tag{3}$$

To evaluate the above triple integral, we first write[4,13]

$$\mathrm{tr}(\mathbf{a}\mathbf{h}\mathbf{b}\mathbf{h}') = \tfrac{1}{3}[\mathrm{tr}(\mathbf{a})\mathrm{tr}(\mathbf{b}) - \mathrm{tr}(\tilde{\mathbf{a}})\mathrm{tr}(\tilde{\mathbf{b}})] + \mathrm{tr}(\tilde{\mathbf{a}}\mathbf{h}\tilde{\mathbf{b}}\mathbf{h}'), \tag{4}$$

where $\tilde{\mathbf{a}} = \mathrm{diag}(a_{13},a_{23},0)$ with $a_{ij} = a_i - a_j$, and $\tilde{\mathbf{b}} = \mathrm{diag}(b_{13},b_{23},0)$. In the nontrivial case where $\mathbf{a},\mathbf{b}\neq\mathbf{0},\mathbf{1}$, one can always rearrange the elements of $\mathbf{a}$ and $\mathbf{b}$ so that they are in descending order without changing the value of the

 0022-2488/90/112567-03$03.00

integral. We shall therefore take $x_{ij} = x_i - x_j > 0$ if $i < j$, with $x:(a,b)$. The integral now depends on only four distinct combinations of the six elements of **a** and **b**, thereby reducing the number of independent variables of $F(\mathbf{a},\mathbf{b})$ by 2. For convenience, we choose the following four parameters: $\epsilon_+ = (\frac{1}{3})\mathrm{tr}(\mathbf{a})\mathrm{tr}(\mathbf{b})$, $\epsilon_- = (\frac{1}{3})\mathrm{tr}(\tilde{\mathbf{a}})\mathrm{tr}(\tilde{\mathbf{b}})$, $\epsilon_a = a_{12}/\mathrm{tr}(\tilde{\mathbf{a}})$ and $\epsilon_b = b_{12}/\mathrm{tr}(\tilde{\mathbf{b}})$ with $-\infty \leqslant \epsilon_+ \leqslant \infty$, $0 < \epsilon_- \leqslant \infty$ and $0 \leqslant \epsilon_a$, $\epsilon_b \leqslant 1$. In terms of these new variables, $\mathrm{tr}(\mathbf{ahbh}')$ takes the form

$$\mathrm{tr}(\mathbf{ahbh}') = \epsilon_+ + (\epsilon_-/4)[\delta_0 + 3\epsilon_b(\delta_1 \cos 2\gamma - \delta_2 \sin 2\gamma)], \tag{5}$$

where

$$\delta_0 = 3r^2 - 1 + 3\epsilon_a(1 - r^2)\cos 2\alpha,$$
$$\delta_1 = 1 - r^2 + \epsilon_a(1 + r^2)\cos 2\alpha,$$
$$\delta_2 = 2\epsilon_a r \sin 2\alpha$$

and

$$r = -\cos\beta.$$

We now evaluate the integral in Eq. (3) by using the following equalities:

$$\int_0^{2\pi} f(\cos 2\theta,\ \sin 2\theta)d\theta = \int_0^{2\pi} f(\cos\theta,\ \sin\theta)d\theta \tag{6}$$

and

$$\int_0^{2\pi} g[\cos(\theta + \theta_0)]d\theta = \int_0^{2\pi} g(\cos\theta)d\theta, \tag{7}$$

where θ_0 is real and independent of θ and $f(x,y)$ and $g(x)$ are any functions defined on $-1 \leqslant x,\ y \leqslant 1$. The result is

$$F(\mathbf{a},\mathbf{b}) = e^{i\epsilon_+} W(\epsilon_-,\epsilon_a,\epsilon_b), \tag{8}$$

where the function $W(x_1,x_2,x_3)$ takes the form of a double integral, one convenient choice of which is

$$W(x_1,x_2,x_3) = \frac{1}{2\pi}\int_0^1 dr \int_0^{2\pi} d\theta \exp\left\{i\frac{x_1}{4}[3\eta(r,\theta,x_2) - 4]\right\} \times J_0\left[\frac{3x_1x_3}{4}\sqrt{\xi(r,\theta,x_2)}\right], \tag{9}$$

where $J_0(x)$ is the Bessel function of the first kind of zero order, and the functions $\eta(r,\theta,x_2)$ and $\xi(r,\theta,x_2)$ are defined as

$$\eta(r,\theta,x_2) = 1 + r^2 - x_2(1 - r^2)\cos\theta \tag{10}$$

and

$$\xi(r,\theta,x_2) = \eta^2 - 4(1 + x_2^2)r^2. \tag{11}$$

We note also that with a change of variables to $x = r^2$ and $y = \eta(r,\theta,x_2)$, Eq. (9) can be cast into the form

$$W(x_1,x_2,x_3) = \frac{1}{2\pi}\int_0^1 \frac{dx}{\sqrt{x}} \int_{y_1}^{y_2} dy \frac{\exp[i(x_1/4)(3y - 4)]}{\sqrt{(y_2 - y)(y - y_1)}} \times J_0\left[\frac{3x_1x_3}{4}\sqrt{y^2 - 4(1 - x_2^2)x}\right], \tag{12}$$

where $y_1 = 1 - x_2 + (1 + x_2)x$ and $y_2 = 1 + x_2 + (1 - x_2)x$. The above integral may be evaluated numerically by use of the Fejér quadrature rule.[14]

Further analysis of the double integral in Eq. (9) may be carried out by using power-series representations of $\sin x$, $\cos x$, and $J_0(x)$. The result is

$$W(x_1,x_2,x_3) = \sum_{n=0}^{\infty}(-1)^n x_1^{2n} \times \sum_{m=0}^{n}[\omega_{m,2l}(x_2^2) - ix_1\omega_{m,2l+1}(x_2^2)]x_3^{2m}, \tag{13}$$

where $l = n - m$ and $\omega_{ml}(x_2^2)$ is given by

$$\omega_{ml}(x_2^2) = \frac{3^{2m}}{4^{3m+l}(2\pi)(m!)^2 l!} \times \int_0^1 dr \int_0^{2\pi} d\theta\, \xi^m(3\eta - 4)^l. \tag{14}$$

Notice that the ω_{ml} are functions of x_2^2 as a result of the inner integral over $\theta \in [0,2\pi]$. We further note that $\omega_{ml}(x)$ is a polynomial of order n in x, i.e.,

$$\omega_{ml}(x) = \sum_{k=0}^{n} \gamma_{mlk} x^k. \tag{15}$$

The coefficients γ_{mlk} are calculated from the double integral in Eq. (14), which may be evaluated analytically for any given m and l.

For the special value of $\epsilon_b = 1/3$, the imaginary part of $W(\epsilon_-,\epsilon_a,\epsilon_b)$ vanishes as it follows that $(\epsilon_-,\epsilon_a,1/3)$ is equal to its complex conjugate (see Ref. 4 for a proof of this property of the function). Thus one has

$$W\left(4x,\epsilon,\frac{1}{3}\right) = \frac{1}{2\pi}\int_0^1 dr \int_0^{2\pi} d\theta \cos\{x[4\eta(r,\theta,\epsilon) - 3]\} \times J_0[x\sqrt{\xi(r,\theta,\epsilon)}], \tag{16}$$

which generates polynomials $\omega_n(\epsilon^2)$ of degree n, i.e.,

$$W\left(4\sqrt{x},\sqrt{\epsilon},\frac{1}{3}\right) = \sum_{n=0}^{\infty}(-1)^n x^n \omega_n(\epsilon) \equiv W(x,\epsilon). \tag{17}$$

The coefficients $\omega_n(\epsilon)$ can be obtained analogously to calculating γ_{mlk} in Eq. (15). For $n = 0$ and 1, for example, one has $\omega_0(\epsilon) = 1$ and $\omega_1(\epsilon) = \frac{8}{17}(1 + 3\epsilon)$. The function $W(x,\epsilon)$ is of use in the newly-developed computational algorithm for the computation of shape distribution functions for macromolecules of arbitrary complexity.[5]

In Fig. 1, we plot $W(x,\epsilon)$ against x for five selected values of $\epsilon \in [0,1]$. As can be seen, $W(x,\epsilon)$ is an oscillating function of x for any given ϵ, decreasing very rapidly as x increases initially from zero to the first zero of the function and reaching a zero limit as $x \to \infty$. This oscillating behavior implies that $W(x,\epsilon)$ may be expressed in terms of classical orthogonal polynomials. Unfortunately, we have not been able to accomplish this. It is also seen from Fig. 1 that the numerical value of the first zero of $W(x,\epsilon)$ decreases as ϵ increases from zero to 1. Our calculations show that the series representation given in Eq. (17) for $W(x,\epsilon)$, when truncated at $n = 10$, yields results that are accurate to at least two digits after the decimal for x lying within the first zero of the

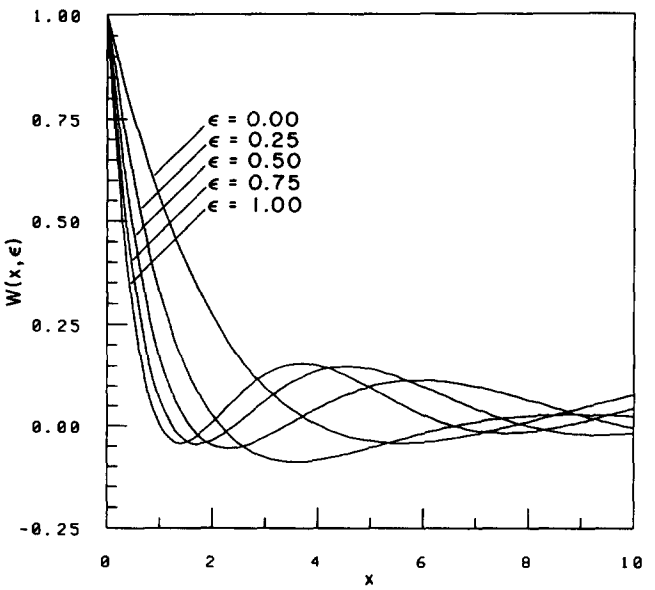


FIG. 1. Plots of $W(x,\epsilon)$ as a function of x for five values of ϵ.

function for any given ϵ when compared with those obtained by numerical quadrature of the double integral in Eq. (16). For small ϵ, the truncated series can be used very effectively in the computation of the function for a wide range of values of x. In a sample calculation of $W(x,\epsilon)$ for $\epsilon = 0$, for example, an absolute error of less than 10^{-6} was found for x varying from zero to about 6.7 when only the first 11 expressions for $\omega_n(\epsilon)$ were used. For larger values of ϵ and x, more terms in the series are required.

III. CONCLUSION

The two-matrix function $F(\mathbf{a},\mathbf{b})$, when expressed as the three-variable function $W(\epsilon_-,\epsilon_a,\epsilon_b)$ multiplied by $\exp(i\epsilon_+)$, shows a number of interesting new features that have not been revealed in previous works. This three-variable function can be represented either as a double integral whose integrand involves the Bessel function of the first kind of zero order or as a series in one of the variables with each coefficient being a polynomial in the other two parameters. For $\epsilon_b = 1/3$, this function is real-valued and when it is expanded as a power series in ϵ_-^2 it generates polynomials in ϵ_a^2. The series for $W(x,\epsilon)$ has been numerically shown to be rapidly convergent. Although the connection between these polynomials and zonal polynomials has not been explored, the parametrization used in this work for the four independent variables of the F function may prove useful for the effective numerical computation of the function expressed alternatively as a series in zonal polynomials multiplied by an exponential factor.

ACKNOWLEDGMENTS

We gratefully acknowledge the financial support for this work by the Department of Energy through Grant No. DE-FG06-84ER45123.

One of us (G.W.) thanks Professor Donald Richards for many critical comments. The referee's comments on this paper are also appreciated.

[1] K. Šolc, J. Chem. Phys. **55**, 335 (1971).
[2] B. E. Eichinger, SIAM Rev. **26**, 117 (1984).
[3] B. E. Eichinger, Macromolecules **18**, 211 (1985).
[4] G. Wei and B. E. Eichinger, Macromolecules **22**, 3429 (1989).
[5] G. Wei and B. E. Eichinger, to appear in Macromolecules.
[6] I. Satake, Notices Am. Math. Soc. **27**, 442 (1980).
[7] L. K. Hua, *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains* (Translation of American Mathematical Society, Providence, RI, 1963).
[8] A. T. James, Ann. Math. Stat. **35**, 475 (1964).
[9] R. J. Muirhead, *Aspects of Multivariate Statistical Theory* (Wiley, New York, 1982).
[10] D. Richards, SIAM Rev. **27**, 81 (1985).
[11] C. Bingham, Ann. Math. Stat. **43**, 1948 (1972).
[12] H. Dym and H. P. Mckean, *Fourier Series and Integrals* (Academic, New York, 1972).
[13] C. Bingham, J. Multivariate Anal. **4**, 210 (1974).
[14] W. Gautschi, Math. Comput. **22**, 251 (1968).

A parametrization for independent variables of a two-matrix hypergeometric function

Gaoyuan Wei
Department of Chemistry, BG-10, University of Washington, Seattle, Washington 98195

(Received 6 February 1990; accepted for publication 11 July 1990)

A new parametrization for the independent variables of the two-matrix hypergeometric function ${}_0F_0^{(k)}(\mathbf{a},\mathbf{b})$ is given. This parametrization results in a net reduction of two variables. The new arguments for the zonal polynomials in the series representation of ${}_0F_0^{(k)}(\mathbf{a},\mathbf{b})$, multiplied by an exponential factor, are all in the interval [0,1], which greatly improves the convergence of the series. The results are of practical use in a number of applications, such as the configuration statistics of macromolecules, shapes of random walks, and multivariate statistics.

I. INTRODUCTION

The two-matrix hypergeometric function ${}_0F_0^{(k)}(\mathbf{a},\mathbf{b})$ (Ref. 1) where $\mathbf{a}$ and $\mathbf{b}$ are $k\times k$ diagonal matrices, has found wide applications in the formulations of both the joint density function of the latent roots of the Wishart matrix in multivariate statistics[2] and the shape distribution function in the configuration statistics of macromolecules.[3] The computation of ${}_0F_0^{(k)}(\mathbf{a},\mathbf{b})$ in the latter case has become increasingly important due to the useful information it contains about the shapes of chain molecules in the unperturbed state or the trails left by a random walker.[3-5] It is well known that when the integral representation of ${}_0F_0^{(k)}(\mathbf{a},\mathbf{b})$ is used, the function may be numerically evaluated by quadrature methods with a proper choice of parametrization of an element of the special orthogonal group.[6,7] More generally, however, ${}_0F_0^{(k)}(\mathbf{a},\mathbf{b})$ is expressed as a series in zonal polynomials.[1,2] This representation provides an elegant analytic solution of the integral which is otherwise difficult to solve in a simple form except for $k=2$ and 3. Unfortunately, such series is generally slowly convergent, although some effort has been made to improve the convergence.[8] It is therefore desirable to accelerate the speed of convergence of the series if it is to be used effectively in the numerical evaluation of ${}_0F_0^{(k)}(\mathbf{a},\mathbf{b})$. This short paper deals with a new parametrization of the independent variables of ${}_0F_0^{(k)}(\mathbf{a},\mathbf{b})$, and it shows that much faster convergence for the series can be achieved.

The integral representation of ${}_0F_0^{(k)}(\mathbf{a},\mathbf{b})$ is[1-3,7,9]

$$ {}_0F_0^{(k)}(\mathbf{a},\mathbf{b}) = \frac{1}{V[\mathrm{SO}(k)]}\int_{\mathrm{SO}(k)} \mathrm{etr}(\mathbf{ahbh}')\,d\mathbf{h}, \tag{1}$$

where $\mathbf{h}\in\mathrm{SO}(k)$, $\mathbf{h}'$ is the transpose of $\mathbf{h}$, and $d\mathbf{h}$ is the unnormalized Haar measure on $\mathrm{SO}(k)$. Without loss of generality, the diagonal elements of $\mathbf{a}$ and $\mathbf{b}$ may be assumed to be in descending order for the nontrival case of $\mathbf{a},\mathbf{b},\neq\mathbf{0},\mathbf{1}$. We then make use of the identity[4,7,10,11]

$$\begin{aligned}\mathrm{tr}(\mathbf{ahbh}') &= (1/k)[\mathrm{tr}(\mathbf{a})\mathrm{tr}(\mathbf{b}) - \mathrm{tr}(\tilde{\mathbf{a}})\mathrm{tr}(\tilde{\mathbf{b}})] \\ &\quad + \mathrm{tr}(\tilde{\mathbf{a}}\mathbf{h}\tilde{\mathbf{b}}\mathbf{h}'),\end{aligned} \tag{2}$$

where $\tilde{\mathbf{x}} = \mathbf{x} - x_k\mathbf{1}$ with $\mathbf{x}$: $(\mathbf{a},\mathbf{b})$. Substituting Eq. (2) into Eq. (1), one obtains

$$\begin{aligned}{}_0F_0^{(k)}(\mathbf{a},\mathbf{b}) &= \exp\{(1/k)[\mathrm{tr}(\mathbf{a})\mathrm{tr}(\mathbf{b}) \\ &\quad - \mathrm{tr}(\tilde{\mathbf{a}})\mathrm{tr}(\tilde{\mathbf{b}})]\}{}_0F_0^{(k)}(\tilde{\mathbf{a}},\tilde{\mathbf{b}}).\end{aligned} \tag{3}$$

Now rewrite $\tilde{\mathbf{x}}$ as

$$\tilde{\mathbf{x}} = (x_1 - x_k)\begin{pmatrix} 1 & & & \\ \dfrac{x_2-x_k}{x_1-x_k} & & & \\ & \ddots & & \\ & & \dfrac{x_{k-1}-x_k}{x_1-x_k} & \\ & & & 0 \end{pmatrix} \tag{4}$$

or

$$\tilde{\mathbf{y}} = (y_1 - y_k)\begin{pmatrix} \mathbf{x}_y & \\ & 0 \end{pmatrix}, \tag{5}$$

where $\mathbf{x}_y = \mathrm{diag}(1,z_1,z_2,..,z_{k-2})$ with $z_j = (y_{j+1}-y_k)/(y_1-y_k)$ and y:(a,b). Note that $0\leqslant z_j < 1$. We further define $x_+ = \mathrm{tr}(\mathbf{a})\mathrm{tr}(\mathbf{b})$ and $x_- = (a_1-a_k)(b_1-b_k)$ with $-\infty\leqslant x_+\leqslant\infty$ and $0<x_-\leqslant\infty$. In terms of these new variables and with use of standard results,[1-3,9] Eq. (3) becomes

$$\begin{aligned}{}_0F_0^{(k)}(\mathbf{a},\mathbf{b}) &= \exp\{(1/k)[x_+ - x_-\,\mathrm{tr}(\mathbf{x}_a)\mathrm{tr}(\mathbf{x}_b)]\} \\ &\quad\times\sum_{m=0}^{\infty}\frac{x_-^m}{m!}\sum_{\kappa}\frac{C_\kappa(\mathbf{x}_a)C_\kappa(\mathbf{x}_b)}{C_\kappa(\mathbf{1}_k)},\end{aligned} \tag{6}$$

where $C_\kappa(\mathbf{x}_a)$ is a zonal polynomial.

Zonal polynomials $C_\kappa(\mathbf{x}_a)$ are symmetric homogeneous functions of degree $k-1$ in the elements of $\mathbf{x}_a$ with $k=(m_1,m_2,...,m_{k-1})$ being a partition of m into no more than $k-1$ parts such that $m_i\geqslant 0$ and $\Sigma_i m_i = m$ (Refs. 1 and 2). Note that Eq. (6) defines a new function of $2(k-1)$ independent variables. For small x_-, the series in Eq. (6) is rapidly convergent since all the elements of $\mathbf{x}_a$ and $\mathbf{x}_b$ are in the interval [0,1]. This improved convergence was first seen in the numerical evaluation of ${}_0F_0^{(3)}(i\mathbf{a},\mathbf{b})$ expressed in a form similar to that given in Eq. (6) with a slightly different choice of the parameters.[11] For large, x_-, however, the improvement on the convergence of the series is not substan-

 0022-2488/90/112570-02$03.00

tial. Nevertheless, the choice of the new set of variables is always useful in effective numerical evaluations of ${}_0F_0^{(k)}(\mathbf{a},\mathbf{b})$.

II. CONCLUSION

In summary, a new parametrization for the independent variables of ${}_0F_0^{(k)}(\mathbf{a},\mathbf{b})$ has been found, which reduces the number of independent variables of the function by two for any k. The new parametrization results in the improved convergence of the series in zonal polynomials representing part of ${}_0F_0^{(k)}(\mathbf{a},\mathbf{b})$, with the exponential part factored out. This improved convergence may well prove useful for the numerical evaluation of ${}_0F_0^{(k)}(\mathbf{a},\mathbf{b})$ by a truncated series of zonal polynomials. Computations of these functions find use in a number of areas, such as configuration statistics of macromolecules, shapes of random walks, and multivariate statistics.

ACKNOWLEDGMENTS

This work is supported in part by the Department of Energy through Grant DE-FG06-84ER45123.

The author wishes to acknowledge the encouragement and comments from Professor Michael D. Perlman and Professor Bruce E. Eichinger.

[1] A. T. James, Ann. Math. Stat. **35**, 475 (1964).
[2] R.J. Muirhead, *Aspects of Multivariate Statistical Theory* (Wiley, New York, 1982).
[3] B. E. Eichinger, Macromolecules **18**, 211 (1985).
[4] G. Wei and B. E. Eichinger, Macromolecules **22**, 3429 (1989).
[5] G. Wei and B. E. Eichinger, Macromolecules (in press).
[6] L. K. Hua, *Harmonic Analysis of Functions of Several Complex Variables in the Classical Domains* (Translation of American Mathematical Society, Providence, RI, 1963).
[7] C. Bingham, Ann. Math. Stat. **43**, 1498 (1972).
[8] A. T. James, Ann. Math. Stat. **31**, 151 (1960).
[9] D. Richards, SIAM Rev. **27**, 81 (1985).
[10] C. Bingham, J. Multivariate Anal. **4**, 210 (1974).
[11] G. Wei and B. E. Eichinger, J. Math. Phys. **31**, 2567 (1990).

A search for integrable bilinear equations: The Painlevé approach

B. Grammaticos
LPN, Université Paris VII, Tour 24-14, 5ème étage, 2 Place Jussieu, 75251 Paris, France

A. Ramani
CPT, Ecole Polytechnique, 91128 Palaiseau, France

J. Hietarinta
Department of Physical Sciences, University of Turku, 20500 Turku, Finland

(Received 17 January 1990; accepted for publication 16 May 1990)

The possibility of the existence of new integrable partial differential equations is investigated, using the tools of singularity analysis. The equations treated are written in the Hirota bilinear formalism. It is shown here how to apply the Painlevé method directly under the bilinear form. Just by studying the dominant part of the equations, the number of cases to be considered can be limited drastically. Finally, the partial differential equations identified in a previous work [J. Hietarinta, J. Math. Phys. **28**, 1732, 2096, and 2586 (1987); **29**, 628 (1988)] as possessing at least four soliton solutions, are shown to pass the Painlevé test as well, which is a strong indication of their integrability.

I. INTRODUCTION

In the 20 years since the discovery of the soliton and of the integrable character of the Korteweg–de Vries (KdV) equation, a large number of integrable partial differential equations (PDEs) have been found by various methods. The notion of integrability has also been refined during this time. Apart from the really exceptional cases where a PDE is integrable through quadratures, two kinds of integrabilities can be distinguished: In the terminology of Calogero,[1] these are C integrability, when the PDE is linearized through a local (Cole–Hopf type) transformation, and S-integrability when the linearization proceeds through a "spectral" or inverse scattering transform.

There are several properties that are encountered in the integrable PDEs and are often taken as synonyms to integrability:

(1) *algebraic properties*, existence of an infinite sequence of conserved quantities, imbedding in a hierarchy of equations. In addition, the equation can be written as a consistency condition for a pair of linear equations, the so-called Lax pair.

(2) *existence of a certain type of special solutions*, solitary waves and multisoliton solutions.

(3) *analytical properties*, the solutions of the PDEs possess the Painlevé property, i.e., they are free from "movable" critical singularities.

The search for new integrable nonlinear equations proceeds in various ways. One can study the algebraic properties that characterize a given PDE and try to obtain other equations that have these same properties. Whole hierarchies of equations have been obtained in this way.[2] Alternatively, one can start with an equation, often suggested by some physical problem, and investigate whether this equation possesses the special analytical properties (Painlevé) and/or particular solution of multisoliton type.

The advantage of investigating the Painlevé property and the existence of multisoliton solutions lies in the fact that their study can be made algorithmically. Thus, although no rigorous theorems exist, one can consider these two approaches as providing integrability criteria. Their combined use[3] constitutes a powerful integrability prognosticator: Equations that pass both tests are most probably integrable. In fact, no counterexample to this last statement is known to date.

In a series of papers,[4] one of us (JH) has presented a systematic search for equations having multisoliton solutions. Several candidates for integrability have been identified this way. This study will be complemented here by the application of singularity analysis methods (Painlevé property) to the equations isolated by the previous approach. The equations that satisfy both integrability criterions are either known integrable equations or new ones the integrability of which is a "safe bet."

This paper is organized as follows. In Sec. II we recall briefly the Hirota formalism that allows for a compact expression for the nonlinear equations studied here. Sections III and IV are devoted to the implementation of the Painlevé analysis to the various equations obtained in Ref. 4. The main improvement over previous treatments is that the analysis is performed directly under the bilinear form of the equation: There is no need to write the equations in their "usual" nonlinear form in order to perform the singularity analysis. After presenting a summary of our findings, comparing them to the conclusions of Ref. 4 concerning the existence of soliton solutions, we discuss the problems that remain open and cannot be settled by the present approach.

II. THE HIROTA BILINEAR FORMALISM

The Hirota bilinear formalism was introduced initially as a compact way of writing nonlinear PDEs and is based on a dependent variable transformation.[5] Let us start with the KdV equation, in its so-called "potential" form:

$$u_{xxx} - 3u_x^2 + u_t = 0 \tag{2.1}$$

and introduce the transformation:

$$u = -2\,\partial_x \log F. \tag{2.2}$$

 0022-2488/90/112572-07$03.00

Note that under this transformation the one-soliton (or, rather, one-kink) solution:

$$u_1 = -p(\tanh(\eta/2) + 1) \tag{2.3}$$

(where $\eta = px + \Omega t + \eta_0$ and $\Omega + p^3 = 0$) writes:

$$u_1 = -2\,\partial_x \log F_1, \tag{2.4}$$

where

$$F_1 = 1 + e^{\eta}. \tag{2.5}$$

We remark that in terms of F_1 the one-soliton solution is written as a finite sum of exponentials. The usual KdV equation is the x derivative of Eq. (2.1), written in terms of $v = \partial_x u$. Note also that in terms of the scattering matrix M, given by the solution of the Gel'fand–Levitan–Martchenko equation which is the final step of the inverse scattering transform, the variable v is expressed as $v = -2\,\partial_x{}^2 \log \det M$.

In terms of the new variable F, Eq. (2.1) is written as

$$FF_{xxxx} - 4F_xF_{xxx} + 3F_{xx}{}^2 + FF_{xt} - F_xF_t = 0. \tag{2.6}$$

Hirota has introduced a new bilinear operator D that helps express (2.6) in a more compact form. The symbol D is defined as

$$D_x^m D_y^n \cdots F\cdot G = (\partial_x - \partial_{x'})^m(\partial_y - \partial_{y'})^n \cdots F(x,y,\ldots) \times G(x',y',\ldots)|_{x'=x,y'=y,\ldots}, \tag{2.7}$$

where the dots stand for possible additional independent variables. Note that the parity of a monomial in D under the interchange of F and G is just the same as that of its total degree. Using the D operator Eq. (2.6) can be rewritten as

$$(D_x^4 + D_xD_t)F\cdot F = 0. \tag{2.8}$$

Initially the Hirota formalism has been just a most convenient tool for the construction of multisoliton solutions of nonlinear PDEs. Recently, however, through the works of the Kyoto group, its deep mathematical meaning in the theory of the integrability of bilinear PDEs has been brought into light.[2]

As mentioned earlier, various nonlinear PDEs were examined in Ref. 4 from the point of view of the "Hirota" integrability, i.e., the existence of multisoliton solutions. In these studies, no recursive proof of the existence of N-soliton solutions for all N was attempted, but rather the investigation was limited to the existence of three- and four-soliton solutions. The constraints linked to the existence of these highly nontrivial solutions are nevertheless so strong that we believe that the PDEs which pass this first test are very good candidates for integrability. In the next section, we will apply a second integrability detector on these candidates.

The following five classes of nonlinear PDEs were studied in Ref. 4:

(1) KdV type $A\ F\cdot F = 0$; (2.9)

(2) modified KdV (MKdV) and

(3) sine–Gordon (SG) types $A(F\cdot F - G\cdot G) = 0,\quad B\ F\cdot G = 0$; (2.10)

(4) nonlinear Schrödinger (NLS) type $A\ F\cdot F = GG^*,\quad B\ F\cdot G = 0,\quad B^*\ F\cdot G^* = 0$; (2.11)

(5) Benjamin–Ono (B–O) type $B\ G\cdot G^* = 0$, (2.12)

where A and B are polynomials in D_x, D_t, etc., A being always even. The distinction between the MKdV and the SG types is based on the parity of B which is odd in the former and even in the latter. For the NLS and B–O cases, the polynomial B is complex, with an even "real part" and an odd "imaginary part" i.e., $B(iD_x, iD_t, \ldots)$ is real. These five classes were chosen because they apply to well-known equations and because they are "generic" in the sense that for these bilinear equations the existence of two-soliton solutions implies only parity conditions on the polynomials A and B.

III. THE PAINLEVÉ TEST IN BILINEAR FORM

A. General

In order to test a given PDE for the Painlevé property, Ablowitz, Ramani, and Segur introduced an algorithm[6] that deals with the reductions to ordinary differential equations (ODEs) of the initial equation. This intermediate step was not, in fact, necessary and later Weiss[7] and co-workers proposed the following formulation of the Painlevé conjecture: A PDE possesses the Painlevé property if its solutions are single-valued about a (noncharacteristic) arbitrary singularity manifold. To be more specific, if the singularity manifold is given by $\phi(x,y,z,\ldots) = 0$ then the solution of the PDE must have the expansion:

$$u(x,y,z,\ldots) = \phi^{\alpha}(x,y,z,\ldots) \times \sum u_n(x,y,z,\ldots)\phi^n(x,y,z,\ldots), \tag{3.1}$$

where the u_n are analytic functions in the neighborhood of the singularity manifold and α is a (negative) integer. If the equation is singular, i.e., if the coefficient of the highest-order derivative may vanish, the values of u which cancel this coefficient must be analyzed as well. In the case of bilinear equations this means that one must consider the zeroes of F (and G, G^*), i.e., α positive integer in (3.1). The arbitrariness of the singularity manifold plays the role of the movable character of the singularities of nonlinear ODEs.

The only drawback of this method is that the calculations it leads to are quite lengthy. To reduce the amount of computations, Kruskal[8] proposed to solve locally for one of the variables and write, for instance

$$\phi(x,y,z,\ldots) \approx x + \psi(y,z,\ldots). \tag{3.2}$$

Similarly, in the expansion (3.1), the coefficients u_n would depend only on the remaining variables y, z,..., but not on x, in the case shown above. (The initial formulation of the algo-

rithm by Weiss is still quite useful as it allows to obtain Bäcklund transformations for the solutions of the PDEs.)

Here, we will apply the Weiss–Kruskal algorithm to various bilinear equations. First, we must start with a suitable ansatz for F. By analogy to the KdV example above in which F was an entire function we expect that for integrability, F must be entire. Therefore, we look for F of the form

$$F=\phi^n\sum_{m=0}^{\infty} f_m\phi^m, \quad n\geqslant 0. \tag{3.3}$$

Thus to a singularity of u corresponds a (possibly multiple) zero of F. If there are several functions F, G, etc., then a "singular" expansion would correspond to at least some of them having a zero of some multiplicity.

B. The leading behavior for KdV type equations

The algorithm for the Painlevé test proceeds in the usual ARS steps. First, one must find all the leading behaviors. Let us analyze in detail the case of a single component, KdV type equation. We suppose that the equation is of order $2p$, i.e., the highest degree in D of a monomial in A is $2p$:

$$(D_x^{2p}+\text{lower-order terms})F\cdot F=0. \tag{3.4}$$

To find the leading behavior we take just $F\approx\phi^n$. We have

$$\begin{aligned} D_x^{2p}F\cdot F &= D_x^{2p}\phi^n\cdot\phi^n \\ &= \sum_{k=0}^{2p}(-1)^k\binom{2p}{k}\partial_x^k\phi^n\,\partial_x^{2p-k}\phi^n \\ &= (2p)!\phi^{2n-2p}\sum_{k=0}^{2p}(-1)^k\binom{n}{k}\binom{n-k}{2p-k} \\ &= (2p)!\phi^{2n-2p}(-1)^p\binom{n}{p}, \end{aligned} \tag{3.5}$$

where we have used well-known identities on the binomial coefficients. Thus it is clear that the lhs of (3.4) will vanish whenever $\binom{n}{p}$ vanishes, i.e., for $n=0,1,\ldots,p-1$. This gives the possible leading behaviors of a KdV type equation.

It may happen that there exist several terms of the same homogeneity with a different x, y,... dependence. As far as the leading behaviors and resonances are concerned, this fact is of no concern. Assuming that the singularity manifold has the form (3.2) it is straightforward to show that the effect of such terms is to multiply the lhs of the dominant term of the equation by a factor that does not vanish outside of characteristics. For instance, the KdV type equation

$$(D_x^4+D_xD_t^3)F\cdot F=0 \tag{3.6}$$

has the same highest-order term as

$$D_x^4F\cdot F=0 \tag{3.7}$$

up to multiplicative factor $(1+\psi_t^3)$.

Let us now proceed to the resonances of Eq. (3.4). Starting from a given leading behavior $F\approx\phi^n$ $(n=0,1,\ldots,p-1)$, we expand as

$$F=\phi^n(1+\omega\phi^r) \tag{3.8}$$

and collect terms linear in ω in Eq. (3.4). We find

$$\begin{aligned} D_x^{2p}F\cdot F &= D_x^{2p}\phi^n\cdot\phi^n \\ &\quad+\omega(D_x^{2p}\phi^{n+r}\cdot\phi^n+D_x^{2p}\phi^n\cdot\phi^{n+r})+O(\omega^2). \end{aligned} \tag{3.9}$$

So the equation for the resonances r is given by

$$Q_n(r)\equiv\frac{D_x^{2p}\phi^{n+r}\cdot\phi^n}{\phi^{2n+r-2p}}=0, \tag{3.10}$$

or, equivalently

$$\begin{aligned} Q_n(r) &= \sum_{k=0}^{2p}(-1)^k\binom{2p}{k}\frac{(n+r)!}{(n+r-k)!}\frac{n!}{(n-2p+k)!} \\ &= (2p)!\sum_{k=0}^{2p}(-1)^k\binom{n+r}{k}\binom{n}{2p-k}. \end{aligned} \tag{3.11}$$

The quantity $\tilde{Q}_n(r)=Q_n(r)/(2p)!$ can be rewritten as

$$\tilde{Q}_n(r)=\sum_{k=0}^{n}(-1)^k\binom{n}{k}\binom{r}{2p-2k}. \tag{3.12}$$

At this stage, no further simplification appears possible for general n. Still, for a given n expression (3.12) can be computed as a finite sum. For $n=0$, we have simply

$$\tilde{Q}_0(r)=\binom{r}{2p} \tag{3.13}$$

with roots $r=0,1,\ldots,2p-1$. Note that as $n=0$, the leading behavior is *not* a zero but a regular point (note that u is also regular at that point), and thus $r=-1$ is *not* a resonance.

For $n=1$, we have

$$\begin{aligned} \tilde{Q}_1(r) &= \binom{r}{2p}-\binom{r}{2p-2} \\ &= \binom{r}{2p-2}\frac{(r+1)(r-4p+2)}{2p(2p-1)}, \end{aligned} \tag{3.14}$$

with resonances $r=-1,0,\ldots,2p-3$, and $4p-2$.

For $n=2$, we find $r=-2,\ -1,\ 0,\ldots,2p-5$, and $(8p-7\pm\sqrt{16p+1})/2$. Thus unless $2p$ is of the form

$$2p=m(m-1)/2, \tag{3.15}$$

with integer m, two irrational resonances exist and the equation does not possess the Painlevé property. Thus only PDEs with leading terms of degree $2p=6,\ 10,\ 28,\ 36,\ldots$, can possibly survive (and of course also $2p=2,4$ for which the possibility $n=2$ does *not* exist!).

For $n=3$, we have $r=-3,\ -2,\ -1,\ 0,\ldots,2p-7$, $4p-6$, and $(8p-9\pm\sqrt{48p+1})/2$. For the last two resonances to be integer, it is necessary that $2p$ be of the form

$$2p=k(k-1)/6, \tag{3.16}$$

for some integer k. This does not exclude $2p=6$, however, because n can be at most 2 in that case. Together with (3.15), Eq. (3.16) severely restricts the possible values of $2p$ beyond 8. Indeed, the solutions can be expressed in terms of the continued fraction expansion of $\sqrt{3}$, and increase exponentially. The first six solutions are

$$2p=210,\quad 2926,\quad 7906276,\quad 110120220,$$
$$297544793910,\quad 4144264359690.$$

For $n=4$, the resonances are $r=-4,\ -3,\ -2,\ -1,\ 0,\ldots,2p-9$ and the four roots of a fourth-order equation:

$$r^4 + 2r^3(-8p+13) + r^2(96p^2 - 336p + 251) + 2r(-128p^3 + 720p^2 - 1184p + 533) + 16(16p^4 - 128p^3 + 344p^2 - 352p + 105) = 0. \tag{3.17}$$

Using the REDUCE symbolic algebra system,[9] we have checked that this equation does not have integer roots for any of the six first values above. Thus if any value of $2p > 6$ passes the first three nontrivial conditions for $n = 2, 3$, and 4, it must at least equal to the seventh value of the above sequence which is larger than 10^{16}. But by that time, there are an additional 5×10^{15} conditions to be met, so although this does not constitute a rigorous proof it is extremely improbable that any of these higher values of p meet the requirement that *all* resonances be integers. Therefore, we conclude that 2, 4, and 6 are the only possible values of $2p$ for an integrable KdV type equation.

Once the nonintegrable character of the KdV type equations with $2p > 6$ has been established, we can extend the above treatment to the other types of equations. For the MKdV, SG, and NLS types, since we can always choose G (and G^*) to vanish, it follows that A is at most of degree 6 in D.

C. The leading part for the SG type equations

For the SG case where B is also even, an important remark is that through the change of variables $F' = F + G$, $G' = F - G$, A and B are interchanged. This shows that, since one can take $G' = 0$, that B is also of degree 6 at most and moreover that one can choose without loss of generality

$$d^\circ B \leqslant d^\circ A \leqslant 6.$$

where d° denotes the degree in D.

If $A = B$, the system decouples into two equations for the quantities $F_\pm = F \pm iG$, namely

$$A\ F_+ \cdot F_+ = 0,$$
$$A\ F_- \cdot F_- = 0.$$

Therefore, if for some A_0, the equation $A_0\ F \cdot F = 0$ is integrable, then the SG system with $A = B = A_0$ is also integrable. Since different A's and B's of the same degree in D cannot be distinguished at the two first steps (leading behaviors and resonances) of the analysis, but only at the third step (resonance conditions) it is clear that, since such A_0's exist for $2p = 2$, 4 and 6, we cannot eliminate the cases

$$d^\circ B = d^\circ A = 2,4,6, \tag{3.18}$$

until we check the resonance conditions.

We must now consider the cases

$$d^\circ B = 2, \quad d^\circ A = 4,$$
$$d^\circ B = 2, \quad d^\circ A = 6,$$
$$d^\circ B = 4, \quad d^\circ A = 6. \tag{3.19}$$

In the first case, although there are quite a few possible leading behaviors to be studied one by one, it turns out that for all of them, both the leading exponents and the resonances are all integers. For the two other there is at least one wrong leading behavior.

For $d^\circ A = 6$, there is a leading behavior where $F \approx x^2$, $G \approx x^q$, $q > 2$. Indeed this does satisfy the lowest-order term of the A equation, because

$$D^6 x^2 \cdot x^2 = 0, \tag{3.20}$$

and G only appears at higher orders. The B equation in (2.10) now leads to an equation for q that is

$$q^2 - 5q + 2 = 0, \tag{3.21}$$

if $d^\circ B = 2$, and

$$q(q-1)(q^2 - 13q + 34) = 0, \tag{3.22}$$

if $d^\circ B = 4$. Both of these equations have one *irrational* root larger than 2, corresponding to a non-Painlevé type leading behavior.

We are left with only the case $d^\circ B = 2$, $d^\circ A = 4$, for which, as we will see, there are indeed integrable subcases, and the three possibilities of (3.18) where this is also true for the reason stated above. This concludes the analysis of the SG type at this point.

D. The leading part for the MKdV type equations

We now turn to the MKdV type equations. We have $d^\circ A = 2, 4, 6$ but we do not have any limitation on $d^\circ B$ yet. (However, we never consider the case $d^\circ B = 1$, as it is too trivial. In fact, the equation $D_x F \cdot G = 0$ just means that F/G does not depend on x.) Because B is *odd*, under the change $F' = F + G$, $G' = F - G$ the B equation remains

$$B\ F' \cdot G' = 0, \tag{3.23}$$

while the A equation becomes

$$A\ F' \cdot G' = 0. \tag{3.24}$$

We will use (2.10) or (3.24) depending on the case at hand, dropping the primes.

Let us first examine the case $d^\circ A = 2$. In the form (3.24) the leading behaviors of F and G are $F \approx x^{n(n-1)/2}$, $G \approx x^{n(n+1)/2}$ with integer n. For $d^\circ B = 3$ the only choices are $n = 0$ and 1 which lead to integer resonances. As soon as $d^\circ B = 2k + 1 \geqslant 5$ one can take $n = 2$ in which case the resonances are: -3, -1, 0 (double), 1,...,$2k-4$, plus $4k-2$ and the two roots of the equation

$$r^2 - (4k+1)r + 8(k^2 - k - 1) = 0 \tag{3.25}$$

with discriminant $\Delta = -16k^2 + 40k + 33$. For $k = 2$ we have $\Delta = 49$ and for $k = 3$, $\Delta = 9$ but for $n \geqslant 4$, Δ is negative and the resonances are complex. This shows that for $d^\circ B \geqslant 9$ the system does not have the Painlevé property. We have checked that for $d^\circ B = 5$ and 7 all the other leading behaviors have integer resonances only.

Let us now turn to $d^\circ A = 4$. We use Eq. (2.10) and choose $F \approx x^1$, $G \approx x^4$ which is indeed a possibility whenever $d^\circ B \geqslant 7$. Then G is subdominant in the A equation and $G \cdot G$ enters with an arbitrary coefficient exactly at resonance 6. Because of this arbitrariness, the resonance condition will not always be satisfied and therefore the Painlevé property does not hold for $d^\circ B > 5$. For $d^\circ B = 5$, the leading behavior $F \approx x^0$, $G \approx x^3$ has resonances -3, -2, -1, 0 (double), 1, 6 and the irrational roots of $r^2 - 15r + 32$, so it does not have the Painlevé property either. We have checked that for

$d^\circ B = 3$ all dominant behaviors lead to integer resonances.

The $d^\circ A = 6$ case is treated along the same lines. A leading behavior $F \approx x^1, G \approx x^6$ is possible whenever $d^\circ B \geqslant 9$ while the resonance condition at resonance 10 cannot always be satisfied. For $d^\circ B = 3$ there are only integer resonances throughout. For $d^\circ B = 5$ $(F \approx x^0, G \approx x^4)$ and $d^\circ B = 7$ $(F \approx x^0, G \approx x^5)$ we find irrational resonances.

Summarizing, the only candidates for the Painlevé property are

$$\begin{aligned} &d^\circ A = 2, \quad d^\circ B = 3,5,7, \\ &d^\circ A = 4, \quad d^\circ B = 3, \\ &d^\circ A = 6, \quad d^\circ B = 3. \end{aligned} \tag{3.26}$$

As a matter of fact we will see in the next chapter that there exist at least one example of an integrable PDE for each of these choices.

E. The leading part for the NLS type equations

Finally, we consider the NLS type systems. As far as singularity analysis is concerned G^* must be considered as a new variable H independent from G, as the singularity manifold need not be real. On the other hand, the *highest* degree monomial of B has a definite parity, and thus at that order one can take $B^* = \pm B$. Thus at highest order we can rewrite the NLS type equation as

$$D^{2p}F \cdot F = GH, \quad D^q F \cdot G = 0, \quad D^q F \cdot H = 0. \tag{3.27}$$

Again, we have $2p = 2, 4, 6$ but q may be even or odd.

We begin with $2p = 2$. Then for all $q(\geqslant 2)$ there is always a leading behavior of the form $F \approx x, G \approx a, H \approx -2/a$, the resonances of which are $-1, \{0,\ldots,q-2\}$ (all double), $2q-1$ and the two roots of

$$r^2 - r(2q+1) + 4q - 2 - 2(-1)^q = 0. \tag{3.28}$$

Consider now the discriminant of this equation. For q even, it is $4q^2 - 12q + 17$. This odd number is equal to $(2q-3)^2 + 8$. In order for it to be a perfect square it must be equal to $(2q-1)^2$ and this is the case for $q = 2$ only. For q odd, the discriminant is $4q^2 - 12q + 1 = (2q-3)^2 - 8$, which, to be a perfect square can only be $(2q-5)^2$ and thus q must be 3. Summarizing, in order for all the resonances for this particular leading behavior to be integer, it is necessary that q be either 2 or 3. We have checked that for these two values, all the other leading behaviors are also of Painlevé type, having integer resonances only.

The case $d^\circ A = 4$ is easier. We start with $F \approx x$, and note that for $q \geqslant 4$, one can take G, $H \approx x^2$ with arbitrary coefficients. Then, the quantity GH on the rhs of the A equation enters exactly at order 6, which is a resonance. Because this term is arbitrary, the resonance condition cannot always be satisfied. Thus the Painlevé property is violated for $q \geqslant 4$. On the other hand for $q = 2, 3$ all leading behaviors lead to integer resonances.

A similar argument can be used for $d^\circ A = 6$. Taking again $F \approx x$, but now with G, $H \approx x^3$ in order for the rhs to enter at resonance 10. This will be possible whenever $q \geqslant 5$, and also happens to be a solution for $q = 2$. Only $q = 3, 4$ are not excluded by this argument, but $q = 4$ has a non-Painlevé leading behavior $F \approx x^2$, G, $H \approx x^s$ with irrational s satisfying $s^2 - 13s + 34 = 0$. Finally, $q = 3$ has noninteger leading behaviors $F \approx x^{10/3}$, $G \approx x^{1/3}$, $H \approx x^{1/3}$, and $F \approx x^{28/3}$, $G \approx x^{10/3}$, $H \approx x^{28/3}$ (and $F \approx x^{28/3}$, $G \approx x^{28/3}$, $H \approx x^{10/3}$). In summary, *no* NLS case with $2p = 6$ can have the Painlevé property.

So the only Painlevé candidates for NLS type equations are

$$\begin{aligned} &d^\circ A = 2, \quad d^\circ B (= d^\circ B^*) = 2,3, \\ &d^\circ A = 4, \quad d^\circ B (= d^\circ B^*) = 2,3. \end{aligned} \tag{3.29}$$

F. Summary

At this point, having used only the constraints from the integer character of the leading exponents and resonances (plus some very general arguments on the resonance conditions) we have narrowed down our investigation of possible Painlevé cases to the following:

KdV type, $d^\circ A = 2,4,6$;

SG type, $d^\circ A = d^\circ B = 2,4,6$,

$d^\circ A = 4, \quad d^\circ B = 2$;

MKdV type, $d^\circ A = 2, \quad d^\circ B = 3,5,7$,

$d^\circ A = 4, \quad d^\circ B = 3$,

$d^\circ A = 6, \quad d^\circ B = 3$,

NLS type, $d^\circ A = 2, \quad d^\circ B (= d^\circ B^*) = 2,3$

$d^\circ A = 4, \quad d^\circ B (= d^\circ B^*) = 2,3$.

IV. BILINEAR EQUATIONS HAVING THE PAINLEVÉ PROPERTY

Before proceeding to a systematic examination of the equations presented in Ref. 4 and which possess soliton solutions a remark is in order. Once the general form of an even polynomial A for KdV-type equation which passes the Painlevé test is found this helps limit the search for the other types. In fact, as G (and G^*) can be taken equal to zero, only those A's given by the KdV-type can be acceptable.

We also note that there are various transformations that change the form of the bilinear equations but not the existence of N-soliton solutions or the Painlevé property. These include linear transformations between dependent and independent variables, which we will freely use in the following. There is also the so-called "gauge transformation" $(F, G, G^*) \to \exp(\phi)\,(F, G, G^*)$ which has no effect if ϕ is linear in the independent variables. If ϕ is quadratic then the bilinear form changes and the transformation make take us also outside of the form assumed in Ref. 4 for N-soliton solutions. However, such transformations do not change the Painlevé

TABLE I. KdV-type equations.

A	F	Resonances
$X^4 + XT + Y^2$	x^1	$-1, 0, 1, 6$
$X^3T + YT + X^2$	x^1	$-1, 0, 1, 6$
$X^4 + XT^3 + aX^2 + bXT + cT^2$	x^1	$-1, 0, 1, 6$
$X^6 + 5X^3T - 5T^2 + XY$	x^1	$-1, 0, 1, 2, 3, 10$
	x^2	$-2, -1, 0, 1, 5, 12$

TABLE II. SG-type equations. In the first (resp. second) equation the constant a may be removed through the gauge transformation $F \to \exp(-axy/2)F$ [resp. $F \to \exp(-ayt/2)F$] unless y is proportional to t (resp. x). In Ref. 4, the form $A = X^3T + 3bX^2 + YT$, $B = XT + b$ was given. However, the constant b can be removed through the gauge transformation $F \to \exp(-bxt/2)F$ and this form reduces to the second equation above. In addition all equations with $A = B$ where A is included in the list of the KdV acceptable equations automatically satisfy the Painlevé criterion. We did not find any Painlevé case with $A \neq B$ for $d^\circ A = d^\circ B = 4$ or 6.

A, B	F, G	Resonances
$XT, XY + a$	x^0, x^1	$-1, 0, 0, 1$
$XT, X^3T + YT + a$	x^0, x^1	$-1, 0, 0, 1, 2, 3$
	x^1, x^3	$-3, -1, 0, 0, 1, 6$

property, so we will use them to reduce the number of equations to check.

So, let us start with the KdV-type equations. For $d^\circ A = 2$ there are no constraints: written in the usual form the equation is linear. For $d^\circ A = 4$ we have shown that the most general *homogeneous* form which can have the Painlevé property involves at most two independent variables. If we write the condition at the resonance $r = 6$ for

$$(c_0 D_x^4 + c_1 D_x^3 D_t + c_2 D_x^2 D_t^2 + c_3 D_x D_t^3 + c_4 D_t^4) F \cdot F = 0, \tag{4.1}$$

we find $c_2^2 - 3c_1 c_3 + 12 c_0 c_4 = 0$. Thus through a rotation in the x-t plane this can be rewritten as

$$D_x (a D_x^3 + b D_t^3) F \cdot F = 0. \tag{4.2}$$

Now we can consider the quadratic and constant terms that can be added to the above homogeneous form. In fact, depending on whether a, b vanish or not, Painlevé analysis (at the resonance condition step) shows that there are essentially three different possibilities:

$$(D_x^4 + D_x D_t + D_y^2 + c) F \cdot F = 0, \tag{4.3}$$

$$(D_x^3 D_t + D_y D_t + D_x^2 + c) F \cdot F = 0, \tag{4.4}$$

$$(D_x^4 + D_x D_t^3 + \lambda D_x^2 + \mu D_x D_t + \nu D_t^2 + c) F \cdot F = 0, \tag{4.5}$$

where in (4.4) we have exchanged x and t from (4.2) in order to write it in a more familiar form.

For $d^\circ A = 6$, Painlevé analysis shows that there is only one possibility for the homogeneous term, which must be exactly D_x^6, and the general nonhomogeneous form is

$$(D_x^6 + 5 D_x^3 D_t - 5 D_t^2 + D_x D_y + c) F \cdot F = 0. \tag{4.6}$$

In all these equations, the constant c can be removed by a "gauge transformation" of F, which in this case consists in multiplying F by the exponential of a quadratic polynomial in the independent variables. Such a transformation only adds a constant to the "usual" variable $\partial_x^2 \log F$. Since it is precisely in the gauge where c vanishes that the one-soliton solution has the usual form (2.5), we will always take $c = 0$ from now on.

Equation (4.3) is the Kadomtsev–Petviashvili (KP) equation (and contains the Boussinesq and KdV equations), (4.4) is known as the Ito equation, (4.5) was first given in Ref. 4, while (4.6) is a combination of the Sawada–Kotera equation and of an equation given by one of us (AR) in Ref. 3. Equations (4.2)–(4.6) are precisely those identified in Ref. 4 as having at least four-soliton solutions. In fact, (4.2), (4.3), and (4.6) are well-known integrable equations belonging to the KP hierarchy of the Kyoto group.[2]

We summarize our findings on the four classes of equations (2.9)–(2.11) in Tables I–IV. In order to simplify the notations, we will write, following Ref. 4, X instead of D_x, T instead of D_t, etc. For all cases, we give the leading exponents, as well as the corresponding resonances. The resonance conditions have been verified using the REDUCE[9] symbolic manipulation language. We must stress the fact that for the equations of the last three types our investigation was not the most general one (and even so, this turned out to be a formidable task, straining the available computer memory to its limits): we have analyzed essentially equations identified in Ref. 4 as having multisoliton solutions. We did try some modifications and generalizations of them but in all cases the Painlevé analysis forced us back to the precise form obtained in Ref. 4.

TABLE III. MKdV-type equations.

A, B	F, G	Resonances
$XT, XTY + X + T$	x^0, x^1	$-1, 0, 0, 3, 4$
$XT, X^3 + T^3 + Y$	x^0, x^1	$-1, 0, 0, 3, 4$
$X^2, X^5 + X^2T + Y$	x^0, x^1	$-1, 0, 0, 1, 2, 3, 6$
	x^1, x^3	$-3, -1, 0, 0, 1, 6, 8$
$X^2, X^7 + X^5 + X^2T + Y$	x^0, x^1	$-1, 0, 0, 1, 2, 3, 4, 5, 8$
	x^1, x^3	$-3, -1, 0, 0, 1, 2, 5, 8, 10$
	x^3, x^6	$-5, -3, -1, 0, 0, 1, 8, 10, 12$
$X^3T + YT + X^2, X^3 + Y$	x^0, x^1	$-1, 0, 0, 1, 1, 2, 6$
	x^1, x^1	$-1, -1, 0, 0, 1, 4, 6$
	x^0, x^2	$-2, -1, 0, 0, 1, 3, 8$
$X^6 + 5X^3T - 5T^2 + XY, X^3 + T$	x^0, x^1	$-1, 0, 0, 1, 1, 2, 3, 4, 8$
	x^1, x^1	$-1, -1, 0, 0, 1, 2, 3, 4, 10$
	x^0, x^2	$-2, -1, 0, 0, 1, 2, 3, 7, 8$
	x^2, x^2	$-2, -2, -1, 0, 0, 1, 5, 5, 12$
	x^2, x^7	$-7, -5, -1, 0, 0, 1, 8, 10, 12$

TABLE IV. NLS-type equations.

A, B	F, G, G^*	Resonances
$X^2, XT + iY + a$	x^1, x^0, x^0	$-1, 0, 0, 3, 4$
$XT, X^2 + T^2 + iY + a$	x^1, x^0, x^0	$-1, 0, 0, 3, 4$
$X^2, iX^3 + XT + iY + a$	x^1, x^0, x^0	$-1, 0, 0, 1, 1, 3, 4, 5$
	x^2, x^2, x^0	$-1, 0, 0, 1, 2, 4, 5, 6$
$X^4 + XY - 3T^2, X^2 + iT + a$	x^1, x^0, x^0	$-1, 0, 0, 0, 1, 3, 3, 6$
	x^3, x^1, x^1	$-3, -1, 0, 0, 1, 5, 6, 8$
$X^4 + XY - 3T^2 + kX^2,$	x^1, x^0, x^0	$-1, 0, 0, 0, 1, 1, 1, 5, 5, 6$
$iX^3 - 3XT - iY/2 + a$	x^2, x^0, x^0	$-2, -1, 0, 0, 1, 2, 3, 6, 7, 8$
	x^5, x^5, x^1	$-5, -4, -1, 0, 0, 1, 6, 7, 8, 12$
$X^3T + aX^2 + bXT + cT^2,$	x^1, x^0, x^0	$-1, 0, 0, 0, 1, 1, 1, 5, 5, 6$
$iX^3 + 3dX^2 + i(b - 3d^2)X - 2icT + e$	x^2, x^0, x^0	$-2, -1, 0, 0, 1, 2, 3, 6, 7, 8$
	x^5, x^5, x^1	$-5, -4, -1, 0, 0, 1, 6, 7, 8, 12$

V. DISCUSSION

Tables I–IV summarize our knowledge on bilinear equations that possess a minimum of four-soliton solutions and at the same time have the Painlevé property. In fact, all the PDEs identified in Ref. 4 through the existence of nontrivial multisoliton solutions also satisfy the Painlevé criterion (unless all their soliton solutions are trivial). The situation is less clear concerning the ODEs obtained in Ref. 4. In fact, quite a few of them do not pass the Painlevé test. Thus the existence of localized solitonlike solutions for ODEs, or of solely trivial soliton solutions for PDEs, does not appear to be intimately related to integrability.

One type of bilinear equations could not be treated by our Painlevé algorithm: namely, the equations of the form

$$D_x D_y (F\cdot F - G\cdot G) = 0, \quad D_z D_t \; F\cdot G = 0. \tag{5.1}$$

If both y and z differ from both x and t, then no multisoliton solution exists, but the Painlevé criterion is automatically satisfied, because there are no positive resonances.

Moreover, a whole class of equations, namely the B–O type, $B\,F\cdot F^* = 0$ (2.12), is outside the range of the present treatment. The difficulty does not arise from the fact that they are integrodifferential equations, when written under their usual nonlinear form. Indeed, we have shown in Ref. 10 that the ARS algorithm can be applied to the B–O equation itself, which does pass the Painlevé test. Moreover, one can also test other similar nonlinear equations (involving the Hilbert transform) and show that they do not pass the test. However, whenever they can be written under bilinear form, the equations of this family trivially pass the test. Keeping in mind that, for singularity analysis purposes, F and F^* must be considered as independent functions, one has only one equation for two unknowns. The usual procedure is to analyze which singularities F, say, can have when F^* is considered as known (and arbitrary). For a nonlinear equation, this is already nontrivial. But since Eq. (2.12) is now linear in F, there are no movable singularities at all, and the Painlevé property is trivially satisfied.

We believe that the equations given in Tables I–IV are excellent candidates for integrability (and the integrable character of many of them is already established). The combination of two powerful integrability criteria allows us to make this statement with a high degree of confidence. Of course, the ultimate proof of integrability will be given only by obtaining the Lax pair, for those cases where it is not yet known. Although this can be done in the Hirota bilinear formalism, it remains a formidable task.

The extension of our method to more complicated bilinear systems is straightforward: it suffices to add more components to the systems and classify the new equations according to the rules we have already set. This method is one of the most promising for finding new integrable PDEs in higher dimensions.

[1] F. Calogero and W. Eckhaus, Inverse Problems **3** 229 (1987).
[2] M. Jimbo and T. Miwa, Publ. RIMS, Kyoto Univ. **19** 943 (1983).
[3] A. Ramani, in *Fourth International Conference on Collective Phenomena*, edited by J. L. Lebowitz (New York Academy of Sciences, New York, 1981), p. 54.
[4] J. Hietarinta, J. Math. Phys. **28**, 1732, 2094, and 2586 (1987); **29**, 628 (1988); work in preparation.
[5] R. Hirota, "Direct Methods in Soliton Theory," in *Solitons*, edited by R. K. Bullough and P. J. Caudrey (Springer, New York, 1980).
[6] M. J. Ablowitz, A. Ramani, and H. Segur, J. Math. Phys. **21**, 715 (1980).
[7] J. Weiss, M. Tabor, and J. Carnevale, J. Math. Phys. **24**, 522 (1983).
[8] M. D. Kruskal, private communication (1982); and in M. Jimbo, M. D. Kruskal, and T. Miwa, Phys. Lett. A **92**, 59 (1982).
[9] A. C. Hearn, *REDUCE User's Manual Version 3.2* (Rand, Santa Monica, 1985).
[10] B. Grammaticos, B. Dorizzi, and A. Ramani, Phys. Rev. Lett. **53**, 1 (1984).

Nonperturbative square-well approximation to a quantum theory

Stefan Boettcher and Carl M. Bender
Department of Physics, Washington University, St. Louis, Missouri 63130

(Received 1 March 1990; accepted for publication 20 June 1990)

The possibility of expressing the solution to a ϕ^{2P} quantum field theory as a series in powers of $1/P$ is proposed. Such a series would be nonperturbative in its dependence on the fundamental parameters of the theory such as the mass and the coupling constant. The first term in such a series describes a field in an infinite-dimensional square-well potential. In this paper, the quantum-mechanical Hamiltonian $H = p^2 + q^{2P}$ is studied as a model calculation and the expansion of the energy levels as series in powers of $1/P$ is examined. The method of matched asymptotic expansions to determine the first five terms in the series for all energy levels is used. The results are compared with extensive numerical calculations of the ground-state energy and it is found that the series is extremely accurate: When $P = 2$, the five-term series has a relative error of 6%, when $P = 10$ the relative error is 0.009%, and when $P = 200$ the relative error is 3.4×10^{-9} %.

I. INTRODUCTION

There have been many attempts to obtain nonperturbative approximations to quantum systems. Such approximations have a clear advantage in that they do not express the content of the theory as series in powers of a physical parameter such as a coupling constant. Thus a nonperturbative approximation may reveal the true dependence of the structure of the theory on the physical parameters. Standard nonperturbative approaches include the $1/N$ expansion in which the field ϕ has N components and the Lagrangian has $O(N)$ symmetry, mean-field, and random-phase approximations, and ϵ expansions. More recently, a nonperturbative expansion called the δ expansion[1] was proposed in which a ϕ^4 field theory is approximated by a $(\phi^2)^{1+\delta}$ theory. In this approach, the Green's functions for a $(\phi^2)^{1+\delta}$ field theory are expanded as series in powers of δ assuming that $\delta \ll 1$. Then δ is allowed to tend to 1 to obtain the solution to a ϕ^4 theory. Note that in this approximation scheme one is expanding about a free-field theory because $(\phi^2)^{1+\delta}$ becomes a mass term in the Lagrangian when $\delta = 0$.

In the present paper we propose the possibility of solving a ϕ^4 quantum theory by expanding the solution to a ϕ^{2P} theory as a series in powers of $1/P$ for P large. Note that the $1/P$ series is *not* an expansion about a free-field theory. Rather, the leading term in the $1/P$ series corresponds to a free field confined to an infinite-dimensional square well.[2] Of course, it is not clear *a priori* whether such an expansion will be numerically accurate. Furthermore, it is not at all obvious how to calculate such an expansion term by term because when $P = \infty$ the field is completely confined to a finite domain while for P finite there is no such confinement. In other words, the leading term in a $1/P$ series is relatively easy to calculate but subsequent terms may be extremely hard to find.

The purpose of this paper is to illustrate the computation of higher-order terms in the $1/P$ series for a simple quantum-mechanical model. We consider the Schrödinger equation

$$\left[-\frac{d^2}{dx^2} + x^{2P} - E(P)\right]\Psi(x) = 0 \tag{1.1}$$

accompanied by the boundary condition

$$\Psi(\pm\infty) = 0. \tag{1.2}$$

We seek an expansion of the eigenvalue $E(P)$ as a series in powers of $1/P$. To leading order we simply set $P = \infty$. This gives the differential equation

$$\left[-\frac{d^2}{dx^2} - E(\infty)\right]\Psi_0(x) = 0 \tag{1.3}$$

subject to the square-well boundary condition

$$\Psi(\pm 1) = 0. \tag{1.4}$$

Note that the limit $P \to \infty$ is a singular limit because, as P becomes infinite, the boundary conditions undergo an abrupt change that reflects the confinement of the wave function to a square-well potential. The eigenvalues of the Schrödinger-equation problem (1.3)–(1.4) are

$$E_n(\infty) = \tfrac{1}{4}\pi^2(n+1)^2, \quad n = 0,1,2,..., \tag{1.5}$$

where n labels the energy level.

The result in (1.5) is the leading term of a series in powers of $1/P$. However, because the perturbation series in powers of $1/P$ is a singular perturbation series, it is not easy to guess the form that such a perturbation series takes. We will see that the eigenvalues have series expansions in powers of $1/P$ and that the coefficient of P^{-k} is a polynomial $Q_n[\ln(2P)]$ of degree k in $\ln(2P)$. Specifically, we will show that

$$E_n(P) = \frac{1}{4}\pi^2(n+1)^2 + \sum_{k=1}^{\infty} P^{-k} Q_k[\ln(P)]. \tag{1.6}$$

The series in (1.6) exhibits some remarkable features. First, the quantity $\ln(2P)$ always appears in the combination $\gamma - \ln(2P)$, where γ is Euler's constant, $\gamma = -\Gamma'(1) = 0.557\,215\,664\,901...$. Second, we can completely factor out the dependence of the series on the term $\gamma - \ln(2P)$ by writing (1.6) in the form

$$E_n(P) = \tfrac{1}{4}\pi^2(n+1)^2(2P)^{-2/(P+1)} \times \Gamma\left(\frac{P}{P+1}\right)^2\left(\sum_{k=0}^{\infty} A_k(n) P^{-k}\right)^2, \tag{1.7}$$

 0022-2488/90/112579-07$03.00

where the coefficients $A_k(n)$ are numbers that can be expressed in terms of the Riemann zeta function:

$$\begin{aligned} A_0(n) &= 1, \\ A_1(n) &= 0, \\ A_2(n) &= -1 - \tfrac{1}{2}\zeta(2), \\ A_3(n) &= \tfrac{3}{2} + \zeta(2) - \tfrac{1}{4}\zeta(3) - \tfrac{1}{3}\left[\tfrac{1}{4}\pi^2(n+1)^2\right]\zeta(3), \\ A_4(n) &= -\tfrac{4}{3} - \zeta(2) + \tfrac{3}{4}\zeta(3) + \tfrac{1}{16}\zeta(4) \\ &\quad + \left[\tfrac{1}{4}\pi^2(n+1)^2\right]\left[\zeta(3) + \tfrac{3}{4}\zeta(4)\right]. \end{aligned} \tag{1.8}$$

The most efficient way to extract accurate numerical predictions from (1.7) is to convert the series $\Sigma_k A_k(n) P^{-k}$ in (1.7) to a Padé approximant. We find that the diagonal series of Padé approximants $P_1^0, P_1^1, P_2^1, P_2^2, \ldots$ gives the best numerical results [in our case, the vanishing of $A_1(n)$ prevents us from constructing the first two terms P_1^0 and P_1^1 in the diagonal Padé sequence].

To demonstrate the accuracy of the $1/P$ expansion in (1.7) we have calculated numerically the ground-state energy $E_0(P)$ for 60 values of P in the range $\frac{1}{2} \leqslant P \leqslant 3500$. In Table I, we compare the numerical results for $E_0(P)$ with the series in (1.6) for some values of P. Observe that the relative error vanishes with increasing P like P^{-5}. In Figs. 1–4 we plot the relative error in various Padé approximants to the $1/P$ series in (1.7) as functions of P. The derivation of the series in (1.6) using the method of matched asymptotic expansions is given in Sec. II.

The accuracy of the $1/P$ expansion suggests that one should explore the connection between the $1/P$ expansion in this paper and the δ expansion in Ref. 1. The connection between these two expansions is easy to establish: Let K be the highest power of $1/P$ in the sum in (1.7). Setting $P = 1 + \delta$ in (1.7) and expanding the result as a series in powers of δ gives for each value of K an expression for the coefficients of each power of δ. As K increases we might expect that the coefficients of each power of δ stabilize and become equal to the coefficients in Ref. 1. If this stabilization actually occurs, it must do so in very high order. The coefficient of δ^0 is 1 for the ground-state energy. The first five approximations to this number obtained as described above are

$$\begin{aligned} &1 \qquad (K=0), \\ &1 \qquad (K=1), \\ &0.834411 \qquad (K=2), \\ &1.381265 \qquad (K=3), \\ &1.348792 \qquad (K=4), \end{aligned} \tag{1.9}$$

which, if they approach 1 do so rather slowly. From these results we are inclined to believe that the δ expansion and the $1/P$ expansion do not have a large common region of validity.

TABLE I. A comparison of the exact values of $E_0(P)$ obtained numerically with the predicted values obtained from (1.7) by converting the series $\Sigma_k A_k(0) P^{-k}$ to a (2,2)-Padé for selected values of P. Observe that the relative error decreases like P^{-5} as P increases. However, beyond $P = 200$ we no longer list the error because the computed eigenvalues are only correct to 11 decimal places.

P	$E_0(P)_{\text{exact}}$	$E_0(P)_{\text{predicted}}$	Relative error
2	1.060 362 090 5	1.121 861 636 6	5.8%
4	1.225 820 114 1	1.231 689 374 9	0.48%
5	1.298 843 700 6	1.301 373 512 8	0.19%
10	1.560 508 342 9	1.560 657 204 6	0.0095%
50	2.105 213 774 0	2.105 213 866 4	4.3×10^{-6}%
200	2.337 875 108 9	2.337 875 109 0	3.4×10^{-9}%
500	2.405 807 979 2	2.405 807 979 2	···
1500	2.443 094 773 6	2.443 094 773 6	···
3500	2.455 762 241 3	2.455 762 241 3	···

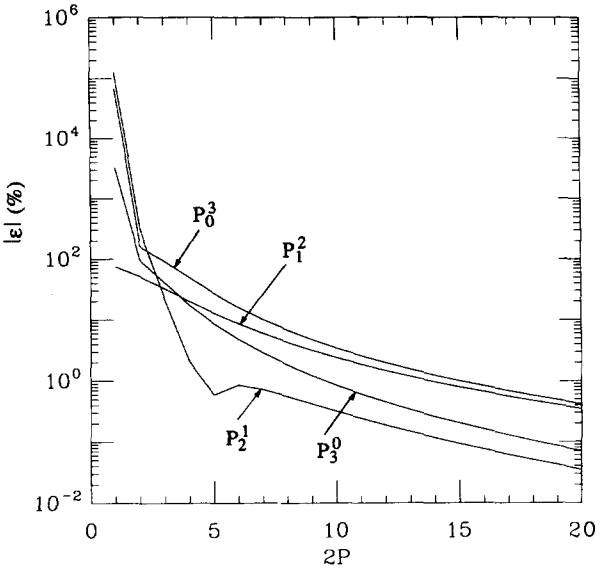


FIG. 1. A plot of the absolute value of the relative error between the exact value of $E_0(P)$ and the predictions for $E_0(P)$ obtained from (1.7) by forming the (3,0), (2,1), (1,2), and (0,3) Padé approximants to the series $\Sigma_{k=0}^{3} A_k(0) P^{-k}$ for $1 \leqslant 2P \leqslant 20$.

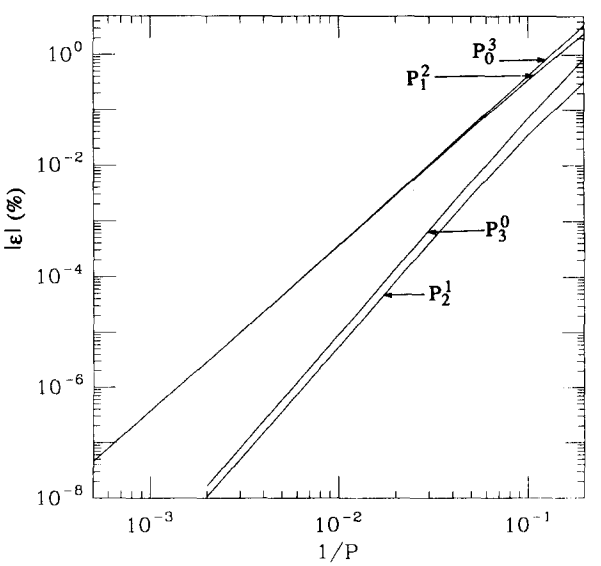


FIG. 2. Same as Fig. 1 for $0.0025 \leqslant 1/(2P) \leqslant 0.1$. The straight-line behavior of the curves in the graph imply that the error is of order P^{-4}.

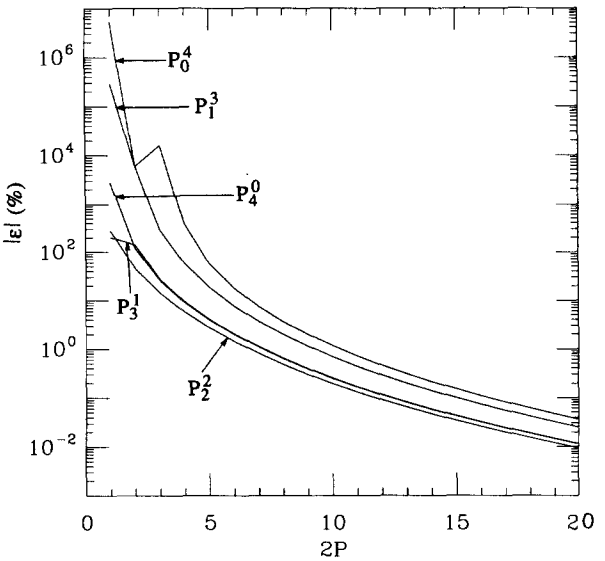


FIG. 3. A plot of the absolute value of the relative error between the exact value of $E(P)$ and the predictions for $E(P)$ obtained from (1.7) by forming the (4,0), (3,1), (2,2), (1,3), and (0,4) Padé approximants to the series $\Sigma_{k=0}^{4} A_k(0) P^{-k}$ for $1 \leqslant 2P \leqslant 20$.

II. DERIVATION OF THE $1/P$ SERIES USING THE METHOD OF MATCHED ASYMPTOTIC EXPANSIONS

In this section we show how to solve the eigenvalue differential equation in (1.1)–(1.2) for large P using the method of matched asymptotic expansions. Let us summarize the analysis briefly: We decompose the domain $0 \leqslant x < \infty$ into three regions. In region 1, where $0 \leqslant x < 1$, we can neglect the exponentially small term x^{2P} in the differential equation. Region 2 consists of the neighborhood of $x = 1$. We will specify the size of this region carefully later on. In region 3, where $x > 1$, we neglect the term E because it is small compared to x^{2P}, which is exponentially large.

In region 1 the differential equation is trivial because it is a constant-coefficient equation. Even-parity eigenvalues are determined by the requirement that the derivative of the wave function vanish at $x = 0$, and odd-parity eigenvalues are determined by requiring that the wave function vanish at $x = 0$. The boundary condition that the wave function vanish at $x = \infty$ imposed in region 3. We will solve the differential equation in each of the three regions, impose the boundary conditions at $x = 0$ and $x = \infty$, and require that the wave function satisfy asymptotic matching conditions at the boundaries of regions 1 and 2 and regions 2 and 3. This matching constraint determines the eigenvalues. Specifically, if we carry out an asymptotic match to k th order in powers of $1/P$, we determine the eigenvalues $E_n(P)$ correct to P^{-k}.

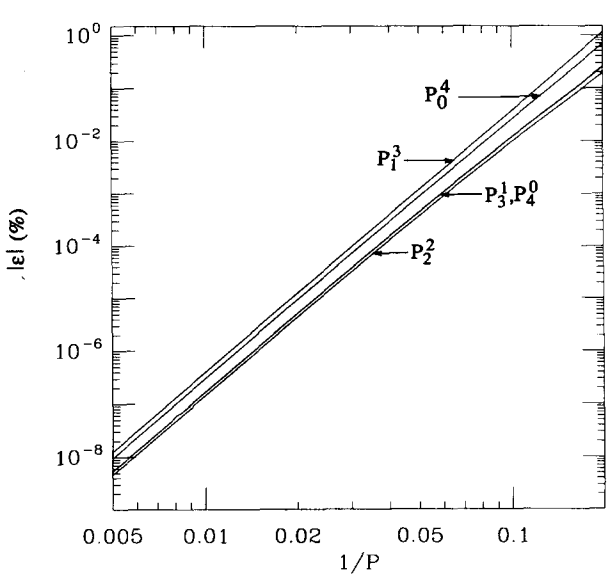


FIG. 4. Same as Fig. 3 for $0.0025 \leqslant 1/(2P) \leqslant 0.1$. The straight-line behavior of the curves in the graph imply that the error is of order P^{-5}.

We begin our analysis by introducing a new independent variable:

$$t = E^{-1/(2P)} x, \quad \Psi(x) = y(t). \tag{2.1}$$

Then (1.1) reads

$$y''(t) + (\pi^2/4) f^2 (1 - t^{2P}) y(t) = 0, \tag{2.2}$$

where

$$f = \frac{2}{\pi} E^{1/2 + 1/2P} = f_0 + f_1 \frac{1}{P} + f_2 \frac{1}{P^2} + \cdots. \tag{2.3}$$

A. Analysis of region 1

Since region 1 consists of those t for which t^{2P} is exponentially small compared with 1, then $y^{(1)}$, the wave function in region 1, satisfies

$$y^{(1)\prime\prime}(t) + \tfrac{1}{4}\pi^2 f^2 y^{(1)}(t) = 0. \tag{2.4}$$

The solution to this equation whose derivative vanishes at the origin is

$$y^{(1)} t = A \cos[(\pi/2) f t], \tag{2.5a}$$

and the solution that vanishes at the origin is

$$y^{(1)} t = A \sin[(\pi/2) f t], \tag{2.5b}$$

where A is a constant to be determined by asymptotic matching. Even-parity eigenvalues will come from matching to (2.5a) and odd-parity eigenvalues will come from matching to (2.5b).

B. Analysis of region 3

Region 3 consists of those t for which t^{2P} is exponentially large compared with 1. Thus $y^{(3)}$, the wave function in region 3, satisfies

$$y^{(3)\prime\prime}(t) - \tfrac{1}{4}\pi^2 f^2 t^{2P} y^{(3)}(t) = 0. \tag{2.6}$$

A simple transformation converts this equation to a modified Bessel equation. The solution to (2.6) satisfying the boundary condition $y(\infty) = 0$, is

$$y^{(3)}(t) = C t^{1/2} K_{1/2(P+1)}\{[\pi f/2(P+1)] t^{1+P}\}, \tag{2.7}$$

where C is a constant to be determined by asymptotic matching.

C. Analysis of region 2

Region 2 is the neighborhood of $t=1$. A convenient variable for the treatment of region 2 is

$$\delta = t - 1. \tag{2.8}$$

In terms of δ (2.2) becomes

$$y^{(2)\prime\prime}(\delta) + \tfrac{1}{4}\pi^2 f^2[1 - e^{2P\ln(1+\delta)}]y^{(2)}(\delta) = 0. \tag{2.9}$$

We take $\delta \ll 1$ and expand $\ln(1+\delta)$ in (2.9):

$$y^{(2)\prime\prime}(\delta) + \tfrac{1}{4}\pi^2 f^2[1 - e^{2P\delta}e^{-P\delta^2}e^{2P\delta^3/3}e^{-P\delta^4/2}\cdots] \times y^{(2)}(\delta) = 0. \tag{2.10}$$

Region 2 consists of those δ that are small compared with 1. However, in order that overlap regions exist between regions 1 and 2 and regions 2 and 3 where we will perform the asymptotic matching, it is necessary that δ must not be too small or else t^{2P} will not be exponentially small in region 1 and not be exponentially large in region 3. As we will see, it is sufficient to take

$$1/P \ll \delta \ll 1/\sqrt{P} \tag{2.11a}$$

to obtain the leading-order (first-order) asymptotic match,

$$1/P \ll \delta \ll 1/P^{3/4} \tag{2.11b}$$

to obtain the second-order asymptotic match,

$$1/P \ll \delta \ll 1/P^{5/6} \tag{2.11c}$$

to obtain the third-order asymptotic match, and so on. As we calculate to successively higher orders in powers of $1/P$, these asymptotic inequalities provide a self-consistent description of the extent of the matching (overlap) regions. Observe that as the order of perturbation theory increases, the size of the overlap regions shrinks. This is a well-known and necessary feature of all calculations involving matched asymptotic expansions.[3]

To obtain solutions to (2.10) to any order in $1/P$ of the form

$$y^{(2)} = y_0^{(2)} + \frac{1}{P}y_1^{(2)} + \frac{1}{P^2}y_2^{(2)} + \frac{1}{P^3}y_3^{(2)} + \cdots, \tag{2.12}$$

we make the key change of variables

$$b = \pi f_0/2P, \quad s = be^{P\delta}, \tag{2.13}$$

where f_0 is the leading term in the expansion of f in (2.3) and b is a small parameter of order $1/P$. With a little algebra it is easy to obtain

$$\exp\left\{2P\sum_{k=1}^{\infty}\frac{(-1)^{k-1}}{k}\delta^k\right\} = \left[\frac{s}{b}\right]^2\sum_{m=0}^{\infty}\frac{2^m\ln(s/b)^m}{m!} \times\left\{\sum_{k=1}^{\infty}\frac{1}{k}\left[-\frac{1}{P}\ln\left[\frac{s}{b}\right]\right]^k\left[\frac{k}{k+1}\right]\right\}^m = \left[\frac{s}{b}\right]^2\sum_{n=0}^{\infty}\left[\frac{1}{P}\right]^n\beta_n\left[\ln\left[\frac{s}{b}\right]\right],$$

with

$$\beta_n\left[\ln\left(\frac{s}{b}\right)\right] = (-1)^n\sum_{m=0}^{n}\sigma_{n,m}\ln\left(\frac{s}{b}\right)^{n+m}$$

and

$$\sigma_{n,m} = 2^m\sum_{\substack{\sum a_i = m \\ \sum i a_i = n}}\frac{1}{a_1!a_2!\cdots a_n!2^{a_1}3^{a_2}\cdots(n+1)^{a_n}} \tag{2.14}$$

and

$$\left(\frac{f}{f_0}\right)^2 = \left[\sum_{i=0}^{\infty}\left(\frac{1}{P}\right)^i\frac{f_i}{f_0}\right]^2 = \sum_{i=0}^{\infty}\left(\frac{1}{P}\right)^i\left\{\sum_{j=0}^{i}\frac{f_{i-j}}{f_0}\frac{f_j}{f_0}\right\}. \tag{2.15}$$

Then (2.10) for all $n = 0,1,2,\ldots$ reads

$$s\frac{d}{ds}s\frac{d}{ds}y_n^{(2)}(s) - \sigma_{0,0}s^2y_n^{(2)}(s) = -\sum_{k=0}^{n-2}\frac{1}{4}\pi^2y_k^{(2)}(s)\sum_{j=0}^{n-2-k}f_jf_{n-2-k-j} + \sum_{k=0}^{n-1}s^2y_k^{(2)}(s)\sum_{i=0}^{n-k}(-1)^i\left\{\sum_{j=0}^{n-k-i}\frac{f_jf_{n-k-i-j}}{f_0^2}\right\} \times\sum_{m=0}^{i}\sigma_{n,m}\ln\left(\frac{s}{b}\right)^{i+m}, \tag{2.16}$$

where $\sigma_{0,0} = 1$. We will see later that

$$y_0^{(2)}(s) = 0. \tag{2.17}$$

To first order in $1/P$, (2.16) is a homogeneous modified Bessel equation of order 0:

$$s^2y_1^{(2)\prime\prime}(s) + sy_1^{(2)\prime}(s) - s^2y_1^{(2)}(s) = 0. \tag{2.18}$$

Its general solution is

$$y_1^{(2)}(s) = K_0(s) + B_1I_0(s), \tag{2.19}$$

where we have exercised our freedom to choose the overall constant in the solution to a homogeneous linear equation by setting the coefficient of the K_0 function equal to 1. This choice will determine the multiplicative constants of the solutions in regions 1 and 3 when we perform the asymptotic matching.

By virtue of the asymptotic inequalities in (2.11), which now have the form

$$1 \ll \ln(s/b) \ll P^{1/2} \text{ (leading order)}, \tag{2.20a}$$

$$1 \ll \ln(s/b) \ll P^{1/4} \text{ (second order)}, \tag{2.20b}$$

$$1 \ll \ln(s/b) \ll P^{1/6} \text{ (third order)}, \tag{2.20c}$$

and so on, we see that to second order in powers of $1/P$, (2.16) now reads

$$s^2y_2^{(2)\prime\prime}(s) + sy_2^{(2)\prime}(s) - s^2y_2^{(2)}(s) = s^2y_1^{(2)}(s)\left[2\frac{f_1}{f_0} - 2\ln^2\left(\frac{s}{b}\right)\right]. \tag{2.21}$$

Unlike the leading-order equation in (2.18) this linear differential equation is inhomogeneous. Using the method of

reduction of order we can write down a formal solution to (2.21):

$$y_2^{(2)}(s) = y_1^{(2)}(s)\int_b^s dx \frac{1}{xK_0(x)^2}\int_x^\infty dz\, zK_0(z)y_1^{(2)}(z) \times\left[2\ln^2\left(\frac{z}{b}\right) - 2\frac{f_1}{f_0}\right] + B_2 I_0(s). \tag{2.22}$$

To third order in powers of $1/P$ (2.16) gives

$$s^2 y_3^{(2)\prime\prime}(s) + s y_3^{(2)\prime}(s) - s^2 y_3^{(2)}(s) = s^2 y_1^{(2)}(s)\left[\left(\frac{f_1}{f_0}\right)^2 + 2\frac{f_2}{f_0} - 4\frac{f_1}{f_0}\ln^2\left(\frac{s}{b}\right) + \frac{8}{3}\ln^3\left(\frac{s}{b}\right) + \ln^4\left(\frac{s}{b}\right)\right] + \frac{1}{4}\pi^2 f_0^2 y_1^{(2)}(s) + s^2 y_2^{(2)}(s)\left[2\frac{f_1}{f_0} - 2\ln^2\left(\frac{s}{b}\right)\right], \tag{2.23}$$

whose formal solution is

$$y_3^{(2)}(s) = y_1^{(2)}(s)\int_b^s dx\frac{1}{xK_0(x)^2}\int_x^\infty dz\, zK_0(z)y_1^{(2)}(z) \times\left[-\left(\frac{f_1}{f_0}\right)^2 - 2\frac{f_2}{f_0} + 4\frac{f_1}{f_0}\ln^2\left(\frac{s}{b}\right) - \frac{8}{3}\ln^3\left(\frac{s}{b}\right) - \ln^4\left(\frac{s}{b}\right)\right] - y_1^{(2)}(s)\int_b^s dx\frac{1}{xK_0(x)^2}\int_x^\infty dz\frac{1}{4}\pi^2 f_0^2\frac{1}{z}K_0(z)y_1^{(2)}(z) + y_1^{(2)}(s)\int_b^s dx\frac{1}{xK_0(x)^2}\int_x^\infty dz\, zK_0(z)y_2^{(2)}(z)\left[2\ln^2\left(\frac{z}{b}\right) - 2\frac{f_1}{f_0}\right] + B_3 I_0(s). \tag{2.24}$$

In every case we choose the limits of integration such that the first term vanishes for $s = b$, the center of region 2, and for $s = \infty$. This choice will simplify subsequent calculations.

This expansion process can be carried out to any order in powers of $1/P$. Note that in (2.23)–(2.24), the $[\pi f_0(2P)]^2$ term in (2.16) contributes for the first time.

D. Matching of regions 2 and 3

The overlap of region 2 and region 3 consists of all $\delta > 0$ satisfying (2.11a). The positivity of δ implies that the arguments of $y^{(2)}(s)$ and $y^{(3)}(s)$ are exponentially large. Thus, the solutions must be asymptotically matched for large arguments of the relevant modified Bessel functions. Besides determining the constant C in (2.7), this match gives the crucial result that to every order in $1/P$ the exponentially growing contributions to $y_1^{(2)}(s), y_2^{(2)}(s), y_3^{(2)}(s),\ldots$ coming from $I_0(s)$ must be eliminated. Thus $B_1 = 0, B_2 = 0,\ldots$.

Since, as we will see, $y_0^{(2)} = 0$, the leading-order term in C must be of order $1/P$. In the overlap region we have

$$y^{(3)}(s) \sim C\sqrt{\frac{\pi}{2s}}e^{-s}\left[1 + O\left(\frac{1}{P}\right) + O\left(\frac{1}{s}\right)\right], \tag{2.25}$$

and in the same overlap region we have

$$y^{(2)}(s) \sim \frac{1}{P}y_1^{(2)}(s) + O\left(\frac{1}{P^2}\right) \tag{2.25}$$

$$\sim \frac{1}{P}\sqrt{\frac{\pi}{2s}}e^{-s}\left(1 + O\left(\frac{1}{s}\right)\right) + B_1\frac{1}{P}\frac{1}{\sqrt{2\pi s}} \times e^{s}\left(1 + O\left(\frac{1}{s}\right)\right) + O\left(\frac{1}{P^2}\right). \tag{2.26}$$

Aside from the crucial result that $B_k = 0$ for any order k in $1/P$, the asymptotic match between solutions in regions 2 and 3 provides no further information. From this knowledge we can now determine the precise form of $y_2^{(2)}(s)$ in (2.22):

$$y_2^{(2)}(s) = K_0(s)\left[a_2 + \frac{1}{2}\ln\left(\frac{s}{b}\right)\right] + sK_1(s)\left[1 - \frac{f_1}{f_0} - \ln\left(\frac{s}{b}\right) + \frac{1}{2}\ln^2\left(\frac{s}{b}\right)\right],$$

with

$$a_2 = (f_1/f_0 - 1)bK_1(b)/K_0(b), \tag{2.27}$$

where we have evaluated all of the indicated integrals. Similarly, we can simplify the expressions for $y_3^{(2)}(s)$ and $y_4^{(2)}(s)$:

$$y_3^{(2)}(s) = \left\{K_0(s)\left[a_3 + \frac{1}{2}a_2\ln\left(\frac{s}{b}\right) - \frac{1}{8}\ln^2\left(\frac{s}{b}\right)\right] + sK_1(s)\left\{\frac{1}{2}\frac{f_1^2}{f_0^2} - \frac{1}{2} - \frac{f_2}{f_0} - \left(\frac{f_1}{f_0} - 1\right)a_2 + \left[\frac{1}{2}\left(1 - \frac{f_1}{f_0}\right) - a_2\right]\ln\left(\frac{s}{b}\right) + \frac{1}{2}a_2\ln^2\left(\frac{s}{b}\right) - \frac{1}{12}\ln^3\left(\frac{s}{b}\right)\right\} + s^2K_0(s)\left\{\frac{1}{2}\left(\frac{f_1}{f_0} - 1\right)^2 + \left(\frac{f_1}{f_0} - 1\right)\ln\left(\frac{s}{b}\right) + \left(1 - \frac{1}{2}\frac{f_1}{f_0}\right)\ln^2\left(\frac{s}{b}\right) - \frac{1}{2}\ln^3\left(\frac{s}{b}\right) + \frac{1}{8}\ln^4\left(\frac{s}{b}\right)\right\} + \frac{1}{4}(\pi^2 f_0^2 - 1)\left[I_0(s)\int_s^\infty dx\frac{1}{x}K_0^2(x) + K_0(s)\int_b^s dx\frac{1}{x}K_0(x)I_0(x)\right], \tag{2.28}$$

with

$$a_3=\left[\left(\frac{f_1}{f_0}-1\right)a_2+\frac{1}{2}+\frac{f_2}{f_0}-\frac{1}{2}\frac{f_1^2}{f_0^2}\right]\frac{bK_1(b)}{K_0(b)}-\frac{1}{2}\left(\frac{f_1}{f_0}-1\right)^2b^2-\frac{1}{4}(\pi^2f_0^2-1)\frac{I_0(b)}{K_0(b)}\int_b^\infty dx\frac{1}{x}K_0^2(x), \quad (2.29)$$

$$\begin{aligned}
y_4^{(2)}(s)=&K_0(s)\left\{a_4+\left[\frac{1}{4}\left(\frac{f_1}{f_0}-1\right)+\frac{1}{2}a_3\right]\ln\left(\frac{s}{b}\right)+\frac{1}{8}(1-a_2)\ln^2\left(\frac{s}{b}\right)+\frac{1}{48}\ln^3\left(\frac{s}{b}\right)\right\}\\
&+sK_1(s)\left\{\frac{1}{3}-\frac{1}{3}\frac{f_1^3}{f_0^3}+\frac{f_1f_2}{f_0^2}-\frac{f_3}{f_0}+a_2\left[\frac{1}{2}\left(\frac{f_1^2}{f_0^2}-1\right)-\frac{f_2}{f_0}\right]+a_3\left(1-\frac{f_1}{f_0}\right)\right.\\
&+\left[\frac{1}{4}\left(\frac{f_1^2}{f_0^2}-1\right)-\frac{1}{2}\frac{f_2}{f_0}-\frac{1}{2}\left(\frac{f_1}{f_0}-1\right)a_2-a_3\right]\ln\left(\frac{s}{b}\right)\\
&\left.+\left[\frac{1}{8}\left(\frac{f_1}{f_0}-1\right)+\frac{1}{2}a_3\right]\ln^2\left(\frac{s}{b}\right)+\left(\frac{1}{24}-\frac{1}{12}a_2\right)\ln^3\left(\frac{s}{b}\right)+\frac{1}{48}\ln^4\left(\frac{s}{b}\right)\right\}\\
&+s^2K_0(s)\left\{-\frac{5}{6}+\frac{3}{2}\frac{f_1}{f_0}-\frac{1}{2}\frac{f_1^2}{f_0^2}-\frac{1}{6}\frac{f_1^3}{f_0^3}-\frac{f_2}{f_0}+\frac{f_1f_2}{f_0^2}+\frac{1}{2}\left(\frac{f_1}{f_2}-1\right)^2a_2\right.\\
&+\left[\frac{7}{4}-\frac{5}{2}\frac{f_1}{f_0}+\frac{3}{4}\frac{f_1^2}{f_0^2}+\frac{f_2}{f_0}+\left(\frac{f_1}{f_0}-1\right)a_2\right]\ln\left(\frac{s}{b}\right)\\
&+\left[-\frac{7}{4}+2\frac{f_1}{f_0}-\frac{1}{4}\frac{f_1^2}{f_0^2}-\frac{1}{2}\frac{f_2}{f_0}+\left(1-\frac{1}{2}\frac{f_1}{f_0}\right)a_2\right]\ln^2\left(\frac{s}{b}\right)+\left(1-\frac{11}{12}\frac{f_1}{f_0}-\frac{1}{2}a_2\right)\ln^3\left(\frac{s}{b}\right)\\
&\left.+\left(-\frac{5}{12}+\frac{1}{4}\frac{f_1}{f_0}+\frac{1}{8}a_2\right)\ln^4\left(\frac{s}{b}\right)+\frac{7}{48}\ln^5\left(\frac{s}{b}\right)-\frac{1}{24}\ln^6\left(\frac{s}{b}\right)\right\}\\
&+s^3K_1(s)\left\{-\frac{1}{6}\left(\frac{f_1}{f_0}-1\right)^3-\frac{1}{2}\left(\frac{f_1}{f_0}-1\right)^2\ln\left(\frac{s}{b}\right)\right.\\
&+\left(\frac{3}{4}-\frac{f_1}{f_0}+\frac{1}{4}\frac{f_1^2}{f_0^2}\right)\ln^2\left(\frac{s}{b}\right)+\left(\frac{1}{2}\frac{f_1}{f_0}-\frac{2}{3}\right)\ln^3\left(\frac{s}{b}\right)\\
&\left.+\left(\frac{3}{8}-\frac{1}{8}\frac{f_1}{f_0}\right)\ln^4\left(\frac{s}{b}\right)-\frac{1}{8}\ln^5\left(\frac{s}{b}\right)+\frac{1}{48}\ln^6\left(\frac{s}{b}\right)\right\}\\
&+\frac{1}{2}\left(\pi^2f_0f_1+\frac{1}{2}-a_2\right)\left[I_0(s)\int_s^\infty dx\frac{1}{x}K_0^2(x)+K_0(s)\int_b^s dx\frac{1}{x}K_0(x)I_0(x)\right]\\
&+\frac{1}{4}(\pi^2f_0^2+1)\left\{I_0(s)\int_s^\infty dx\frac{1}{x}K_0^2(x)\left[a_2+\frac{1}{2}\ln\left(\frac{x}{b}\right)\right]\right.\\
&+I_0(s)\int_s^\infty dx\,K_0(x)K_1(x)\left[1-\frac{f_1}{f_0}-\ln\left(\frac{x}{b}\right)+\frac{1}{2}\ln^2\left(\frac{x}{b}\right)\right]+K_0(s)\int_b^s dx\frac{1}{x}K_0(x)I_0(x)\left[a_2+\frac{1}{2}\ln\left(\frac{x}{b}\right)\right]\\
&\left.+K_0(s)\int_b^s dx\,I_0(x)K_1(x)\left[1-\frac{f_1}{f_0}-\ln\left(\frac{x}{b}\right)+\frac{1}{2}\ln^2\left(\frac{x}{b}\right)\right]\right\}, \quad (2.30)
\end{aligned}$$

with

$$\begin{aligned}
a_4=&-\left\{\frac{1}{3}-\frac{1}{3}\frac{f_1^3}{f_0^3}+\frac{f_1f_2}{f_0^2}-\frac{f_3}{f_0}+a_2\left[\frac{1}{2}\left(\frac{f_1^2}{f_0^2}-1\right)-\frac{f_2}{f_0}\right]-a_3\left(\frac{f_1}{f_0}-1\right)\right\}\frac{bK_1(b)}{K_0(b)}\\
&+\left\{\frac{5}{6}-\frac{3}{2}\frac{f_1}{f_0}+\frac{1}{2}\frac{f_1^2}{f_0^2}+\frac{1}{6}\frac{f_1^3}{f_0^3}+\frac{f_2}{f_0}-\frac{f_1f_2}{f_0^2}-\frac{1}{2}\left(\frac{f_1}{f_2}-1\right)^2a_2\right\}b^2\\
&+\frac{1}{6}\left(\frac{f_1}{f_0}-1\right)^3\frac{b^3K_1(b)}{K_0(b)}-\frac{1}{2}\left(\pi^2f_0f_1+\frac{1}{2}-a_2\right)\frac{I_0(b)}{K_0(b)}\int_b^\infty dx\frac{1}{x}K_0^2(x)\\
&-\frac{1}{4}(\pi^2f_0^2+1)\frac{I_0(b)}{K_0(b)}\left\{\int_b^\infty dx\frac{1}{x}K_0^2(x)\left[a_2+\frac{1}{2}\ln\left(\frac{x}{b}\right)\right]\right.\\
&\left.+\int_b^\infty dx\,K_0(x)K_1(x)\left[1-\frac{f_1}{f_0}-\ln\left(\frac{x}{b}\right)+\frac{1}{2}\ln^2\left(\frac{x}{b}\right)\right]\right\}. \quad (2.31)
\end{aligned}$$

In principle, it is straightforward to obtain $y_5^{(2)}, y_{6,\ldots}^{(2)}$ in a similar fashion. Ultimately, however, one encounters difficult integrals of the form

$$\int_s^\infty dx \frac{1}{x} K_0^2(x) \int_b^x dz \frac{1}{z} K_0(z) I_0(z), \tag{2.32}$$

which make it difficult to express the expansion in terms of b or s.

E. Matching of regions 1 and 2

The constant A in (2.5) is a series in powers of $1/P$ beginning with $(1/P)^0$. To obtain an asymptotic match here we replace the variable t in (2.5) by $1 + 1/P \ln(s/b)$ and expand the result as a series in powers of $1/P$. Using the fact that s is exponentially small ($\delta < 0$) and b is of order $1/P$, we expand $y_1^{(2)}(s) = K_0(s)$ and (2.27)–(2.29). Demanding that the expansion of (2.5) agrees order-by-order with the expansion of $y^{(2)}(s)$ in the overlap region gives a sequence of relations for the coefficient A and the eigenvalue E in terms of f. To leading order the condition on E is

$$\cos[(\pi/2) f_0] = 0 \tag{2.33a}$$

or

$$\sin[(\pi/2) f_0] = 0, \tag{2.33b}$$

for the infinite square well, confirming that

$$y_0^{(2)}(s) = 0, \tag{2.34}$$

which gives a discrete spectrum

$$f_0 = n + 1, \quad n = 0,1,2,\ldots. \tag{2.35}$$

To higher order, the process of matching establishes equations between coefficients of powers of $\ln(x/b)$. Solving these equations iteratively gives (1.6). In particular, we find that

$$\begin{aligned}
Q_1 &= \tfrac{1}{4}\pi^2(n+1)^2(2v), \\
Q_2 &= \tfrac{1}{4}\pi^2(n+1)^2(2v^2 - 2v - 2), \\
Q_3 &= \frac{1}{4}\pi^2(n+1)^2\left\{\frac{4}{3}v^3 - 4v^2 - 2v + 3 + \frac{1}{3}\zeta(3) \right. \\
&\quad \left. - \left[\frac{1}{4}\pi^2(n+1)^2\right]\frac{2}{3}\zeta(3)\right\}, \\
Q_4 &= \frac{1}{4}\pi^2(n+1)^2\left\{\frac{2}{3}v^4 - 4v^3 + 2v^2 + \left[8 + \frac{1}{3}\zeta(3)\right]v \right. \\
&\quad - \frac{5}{3} - \frac{1}{2}\zeta(3) + \left[\frac{1}{4}\pi^2(n+1)^2\right] \\
&\quad \left. \times\left[-\frac{4}{3}v\zeta(3) + 2\zeta(3) + \frac{3}{2}\zeta(4)\right]\right\},
\end{aligned} \tag{2.36}$$

$v = \gamma - \ln(2P)$.

ACKNOWLEDGMENT

We thank the U. S. Department of Energy for partial financial support.

[1] C. M. Bender, K. A. Milton, M. Moshe, S. S. Pinsky, and L. M. Simmons, Jr., Phys. Rev. Lett. **58**, 2615 (1987); C. M. Bender, K. A. Milton, M. Moshe, S. S. Pinsky, and L. M. Simmons, Jr., Phys. Rev. D **37**, 1472 (1988).

[2] We can express a ϕ^{2P} quantum field theory as a functional integral

$$Z[J] = \int D\phi \exp\left\{-\int dx\left[\frac{1}{2}(\partial\phi)^2 + \frac{1}{2}m^2\phi^2 + g\phi^{2P} + J\phi\right]\right\}.$$

Interpreting this functional integral as a finite product of Riemann integrals on lattice gives

$$Z[J] = \int_{-\infty}^{\infty}\left[\prod_n d\phi_n\right]\exp\left\{-a^d\sum_n\left[\frac{1}{2}(\partial\phi_n)^2 + \frac{1}{2}m^2\phi_n^2 + g\phi_n^{2P} + J_n\phi_n\right]\right\}.$$

Now, as $P \to \infty$ we have

$$Z[J] = \int_{-1}^{+1}\left[\prod_n d\phi_n\right]\exp\left\{-a^d\sum_n\left[\frac{1}{2}(\partial\phi_n)^2 + \frac{1}{2}m^2\phi_n^2 + J_n\phi_n\right]\right\}$$

and we see that the field at each lattice point n is restricted to lie in the range $-1 \leqslant \phi_n \leqslant 1$. Thus we have a *free* field theory confined to an infinite-dimensional square well.

[3] *Advanced Math Methods for Scientists and Engineers*, C. M. Bender and S. A. Orszag (McGraw-Hill, New York, 1978), Sec. 7.4.

On the quantization of the modified Pöschl–Teller potential

Yihren Wu
Department of Mathematics, Hofstra University, Hempstead, New York 11550

(Received 1 March 1990; accepted for publication 20 June 1990)

It is shown that the connection one-form on the prequantization line bundle that generates the correct energy spectrum for the modified Pöschl–Teller potential consists of the canonical one-form on the corresponding cotangent bundle, together with an adjustment term that arises from the Berry's connection.

I. INTRODUCTION

In studying the role of the connection one-form on the prequantization line bundle in the scheme of geometric quantization,[1] McKenna and Wan[2] have shown that inequivalent choices of one-form will result in inequivalent Bohr–Sommerfeld conditions. Here, one-forms ω and ω' on a symplectic manifold M with fixed symplectic form $\Omega = d\omega = d\omega'$ are considered inequivalent if $\omega - \omega'$ is nonzero in $H^1_{\text{deRham}}(M)$. Thus, in the case where $H^1_{\text{deRham}}(M) \neq 0$, i.e., if M is multiply connected, the energy spectrum of a Hamiltonian will depend on the choice of one-form ω. Among the examples in Ref. 2, they have shown that in quantizing the bounded state Pöschl–Teller problem with Hamiltonian $H(p,q)$ defined on the phase space $\mathbb{R}^2 \setminus 0$ with the usual symplectic two-form $dp\,dq$:

$$H(p,q) = \tfrac{1}{2}(p^2 - \lambda(\lambda-1)/\cosh^2 q). \tag{1}$$

The correct spectrum $E_n = -\frac{1}{2}(\lambda - 1 - n)^2$ will be obtained by choosing the connection one-form:

$$\omega = p\,dq + (\lambda - 1 - \{\lambda(\lambda-1)\}^{1/2})d\phi, \tag{2}$$

where ϕ is the angle in the action-angle variable.

It was discovered by Berry[3] that a phase shift will result in the eigenfunctions to a parameter-dependent Hamiltonian as the parameters traverse adiabatically along a closed curve. Simon[4] has recognized that this phase arises from a connection (Berry's connection) on the solution line bundle over the parameter space. Recent developments on the dynamical meaning of Berry's phase[5–7] suggest that the curvature of this connection can be interpreted as a symplectic form on the parameter space. Moreover, if the parameter space is itself a symplectic manifold (the parameters being the "slow" dynamical variables), this form acts as an adjustment term[8,9] in the geometric quantization process, when the collective motion is separated out. The phase shift along a closed curve γ is computed by integrating the curvature over a cap surface of γ, assuming the parameter space is simply connected (so Stokes' theorem applies). Thus emphasis has been placed on the curvature form of the bundle. Only recently, an example of a flat Berry's connection was studied,[10] there the Berry's curvature is 0 since the parameter space is one dimensional (circle).

The purpose of this paper is to show that Berry's connection gives the extra term $(\lambda - 1 - \{\lambda(\lambda-1)\}^{1/2})d\phi$ in Eq. (2) by a separation of variables within an enlarged system. We consider the geodesic flow on the one-sheeted hyperboloid in $\mathbb{R}^3$, with natural SO(2,1) symmetry. This Hamiltonian becomes Eq. (1) via the reduction procedure of Marsden and Weinstein,[11] where we see that the terms $\lambda - 1$ and $\{\lambda(\lambda-1)\}^{1/2}$ are identified with the energy and angular momentum variables, i.e., the Casimir operators for the SO(2,1) and SO(2) groups. Our consideration here suggests that in the case where the variables are separable, so there is no adjustment term for the symplectic two-form on the parameter space, an adjustment by a (flat) connection one-form still persists.

In Sec. II we discuss the symmetry and reduction for the Pöschl–Teller problem. It is shown that the reduced phase space has a one-form relevent to the discrete series representation of SO(2,1). We must emphasize that our work does not imply that there is always a natural choice of connection one-form which gives the correct spectrum. Our choice is dictated by the discrete series representation of the symmetry group SO(2,1), which we discover in the enlarged system. In Sec. III we show that the Berry's phase gives the adjustment term in Eq. (2).

II. SYMMETRY AND REDUCTION

The Hamiltonian $H(p,q)$ in Eq. (1) can be recognized as the radial term in the geodesic flow problem on a one-sheeted hyperboloid in $\mathbb{R}^3$. Explicitly, let

$$\begin{aligned} y_1 &= r\cosh\alpha\cos\theta, \\ y_2 &= r\cosh\alpha\sin\theta, \\ y_3 &= r\sinh\alpha. \end{aligned} \tag{3}$$

Denote by $X = \{y \in \mathbb{R}^3 | [y,y] = r^2 = 1\}$ the one-sheeted hyperboloid, where $[\ ,\ ]$ has signature $+\ +\ -$, then X is conveniently parametrized by (α,θ) with $\alpha \in \mathbb{R}$ and $\theta \in S^1$. The group SO(2,1) acts on X transitively. Let T^*X denote the cotangent bundle on X, with canonical one-form $p_\alpha\,d\alpha + p_\theta\,d\theta$. We extend the group action to a Hamiltonian group action on T^*X with momentum mapping[12] $\mathbf{J}: T^*X \to \text{so}^*(2,1)$ given by

$$(p_\alpha, p_\theta, \alpha, \theta) \to \begin{pmatrix} 0 & \xi_{12} & \xi_{13} \\ -\xi_{12} & 0 & \xi_{23} \\ \xi_{13} & \xi_{23} & 0 \end{pmatrix}, \tag{4}$$

$$\xi_{12} = p_\theta, \quad \xi_{13} = p_\alpha\cos\theta - p_\theta(\text{sh}\,\alpha/\text{ch}\,\alpha)\sin\theta,$$

$$\xi_{23} = p_\alpha\sin\theta + p_\theta(\text{sh}\,\alpha/\text{ch}\,\alpha)\cos\theta.$$

The Casimir operator is

$$\xi_{13}^2 + \xi_{23}^2 - \xi_{12}^2 = p_\alpha^2 - p_\theta^2/\text{ch}^2\,\alpha = 2H(p_\alpha, p_\theta, \alpha, \theta). \tag{5}$$

By setting $p = p_\alpha$, $q = \alpha$, $p_\theta = \{\lambda(\lambda-1)\}^{1/2}$, we recover the Hamiltonian for the modified Pöschl–Teller potential of McKenna *et al.*

 0022-2488/90/112586-02$03.00

Using the standard reduction procedure,[11] denote by $\mathcal{O}_\mu$ the coadjoint orbit of

$$\begin{pmatrix} 0 & \mu & 0 \\ -\mu & 0 & 0 \\ 0 & 0 & 0 \end{pmatrix},$$

in so*(2,1), one checks that

$$\mathcal{O}_\mu = \{\xi_{12}^2 - \xi_{13}^2 - \xi_{23}^2 = \mu^2\}, \tag{6}$$

can be viewed as the upper sheet of a two-sheeted hyperboloid, of "radius" μ. The reduced phase space of the constant energy surface $\mathbf{J}^{-1}(\mathcal{O}_\mu) \subset T^*X$ is SO(2,1)/SO(2), which is equivalent to the Poincare disk D with canonical symplectic structure induced by the Kähler potential. The reduction is given by the map

$$\rho : \mathbf{J}^{-1}(\mathcal{O}_\mu) \to D,$$

$$z = [1/(p_\theta + \mu)][p_\alpha + ip_\theta(\text{sh}\,\alpha/\text{ch}\,\alpha)]e^{i\theta}, \tag{7}$$

with one-form on D and its curvature given by

$$\omega_D = i\mu[z\,d\bar{z}/(1 - z\bar{z})],$$

$$\Omega_D = i\mu[dz\,d\bar{z}/(1 - z\bar{z})^2]. \tag{8}$$

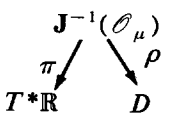
Any other choice of one-forms is equivalent since D is simply connected. One checks that our choice is natural in that the quantization of D using the Bargmann–Fock polarization[1] gives the discrete series representation of SO(2,1). The Hamiltonian vector field for the angular momentum on D is given by

$$i\left(z\frac{\partial}{\partial z} - \bar{z}\frac{\partial}{\partial \bar{z}}\right).$$

With the symplectic form Ω_D, we calculate the Hamiltonian function to be $\mu[z\bar{z}/(1 - z\bar{z})]$ using the relation $\Omega_D \lrcorner \mathcal{H}_f = -df$. This Hamiltonian function pulls back via ρ to $\frac{1}{2}(p_\theta - \mu)$. This action rotates t in $z = re^{it}$.

III. CALCULATION OF THE BERRY'S CONNECTION

We have natural projection $\pi : \mathbf{J}^{-1}(\mathcal{O}_\mu) \to T^*\mathbb{R}$ with $\pi(p_\alpha, p_\theta, \alpha, \theta) = (p_\alpha, \alpha)$ where the image space is the phase space considered by McKenna *et al.* (set $p = p_\alpha$ and $q = \alpha$) in their treatment of Pöschl–Teller potential. We let $T^*\mathbb{R}$ be the parameter space in the Berry's phase consideration. We have the diagram



The Berry's connection ω_B is given by

$$\omega_B = \text{av}\,\rho^* i\mu[z d_A \bar{z}/(1 - z\bar{z})], \tag{9}$$

where av denotes the average over the orbit along

$$i\left(z\frac{\partial}{\partial z} - \bar{z}\frac{\partial}{\partial \bar{z}}\right),$$

that is the same as over t in $z = re^{it}$; d_A is the differential with respect to the p_α, α coordinates.

According to Ref. 8, the effective one-form on $T^*\mathbb{R}$ reads:

$$p_\alpha\, d\alpha + \omega_B. \tag{10}$$

We will show that in terms of the action-angle coordinates introduced in Ref. 2,

$$\rho^* i\mu[z d_A \bar{z}/(1 - z\bar{z})] = (\mu - p_\theta)d\phi + df. \tag{11}$$

One then realizes that, with $\mu = (\lambda - 1)$ and $p_\theta = \{\lambda(\lambda - 1\}^{1/2}$, ω_B is equivalent to the correction term in Eq. (2). The last factor df, being a total differential, averages out to zero.

Proceeding with the calculation, one shows easily that

$$\omega_D = (p_\theta - \mu)dt, \tag{12}$$

where

$$t = \arg z = \theta + \tan^{-1}(\text{sh}\,\alpha\; p_\theta/\text{ch}\,\alpha\; p_\alpha).$$

It remains to compute $(p_\theta - \mu)d_A t$. We have

$$(p_\theta - \mu)d_A t = \frac{p_\theta}{p_\theta + \mu}\left\{\frac{\text{sh}\,\alpha}{\text{ch}\,\alpha}dp_\alpha - \frac{p_\alpha}{\text{ch}^2\,\alpha}\right\}d\alpha. \tag{13}$$

Using the action-angle coordinates of Ref. 2,

$$\sinh\alpha = (p_\theta^2 - \mu^2)^{1,2}\sin\phi/\mu,$$

$$p_\alpha = \frac{(p_\theta^2 - \mu^2)^{1/2}\mu\cos\phi}{\{\mu + (p_\theta^2 - \mu^2)^{1/2}\sin^2\phi\}^{1/2}}, \tag{14}$$

we have

$$(p_\theta - \mu)d_A t = (\mu - p_\theta)(p_\theta/\mu\,\text{ch}^2\,\alpha)d\phi. \tag{15}$$

Since

$$p_\theta/\mu\;\text{ch}^2\,\alpha = \mu\; p_\theta/(\mu^2\cos^2\phi + p_\theta^2\sin^2\phi). \tag{16}$$

Average over ϕ yields 1; therefore, we have Fourier series expansion on $\alpha(\mu,\phi)$:

$$\frac{p_\theta}{\mu\;\text{ch}^2\,\alpha} = 1 + \sum_{n=1}^{\infty} c_n(\mu, p_\theta)\cos n\phi. \tag{17}$$

Combining (15) and (17),

$$(p_\theta - \mu)d_A t = (\mu - p_\theta)d\phi + \sum_{n=1}^{\infty} k_n(\mu, p_\theta)\cos n\phi\, d\phi, \tag{18}$$

the last term is a total differential; p_θ is treated as a constant $\{\lambda(\lambda - 1)\}^{1/2}$. We finally obtain

$$\omega_B = \text{av}(p_\theta - \mu)d_A t = (\mu - p_\theta)d\phi, \tag{19}$$

which is the correction term in Ref. 2.

[1] J. Sniatycki, *Geometric Quantization and Quantum Mechanics* (Springer, New York, 1980).
[2] I. H. McKenna and K. K. Wan, J. Math. Phys. **25**, 1798 (1984).
[3] M. V. Berry, Proc. R. Soc. London Ser. A **392**, 45 (1984).
[4] B. Simon, Phys. Rev. Lett. **51**, 2167 (1983).
[5] M. V. Berry and J. H. Hanny, J. Phys. A **21**, L325 (1988).
[6] H. Kuratsuji and S. Iida, Phys. Rev. Lett. **56**, 1003 (1986).
[7] E. Gozzi and W. Thacker, Phys. Rev. D **35**, 2398 (1987).
[8] Y. Wu, J. Phys. A **22**, L117 (1989).
[9] Y. Wu, J. Math. Phys. **31**, 294 (1990).
[10] J. Zak, Phys. Rev. Lett. **62**, 2747 (1989).
[11] J. Marsden and A. Weinstein, Rep. Math. Phys. **5**, 121 (1974).
[12] V. Guillemin and S. Sternberg, *Symplectic Techniques in Physics* (Cambridge, U.P., Cambridge, England, 1984).

Accuracy-weighted variational principle for degenerate continuum states

Michael Danos
National Institute of Standards and Technology, Gaithersburg, Maryland 20899

(Received 13 December 1988; accepted for publication 27 June 1990)

A variational principle for continuum states is given that permits numerical solution by the Ritz method. It allows one to maximize the accuracy of the solution in preselected regions of space, and also allows the selection of that solution from the perhaps infinitely degenerate solution set that is needed in the particular application.

I. INTRODUCTION

In the treatment of the nuclear effects on the muon-catalyzed deuterium–tritium (dt) fusion we were confronted by the need to have a highly accurate representation of the three-body dtμ Coulomb wave function around $r_{dt} \simeq 5$ fm, while the muon Bohr radius is ~ 250 fm (Ref. 1). This need arises when one attempts to describe the nuclear interactions by means of the R-matrix method for the nuclear two-body subsystem, which demands as input data the value and normal derivative of the wave function at the nuclear channel radius. There exists at this time a very accurate R matrix for the ^{5}He system, covering the energy region of interest, i.e., the energy around the dt threshold.[2] As a result of the nuclear interaction, the bound states of the dtμ system are coupled to the three-body $n\alpha\mu$ (neutron–alpha–muon) system which, at that energy, is 17.6 MeV in the continuum. Hence, the dtμ states, which in the absence of the nuclear interactions, would have been bound become continuum states that, at the same time, acquire the (continuous) infinite degeneracy of the three-body $n\alpha\mu$ states. Still, the bulk of the dtμ wave function lies around $r_{dt} \sim 250$ fm, while the wave function around the nuclear interaction range, $r_{dt} \sim a \sim 5$ fm is exponentially small. This way the region of interest contributes a very small fraction to the energy matrix elements and hence will be determined with a relatively low accuracy in a variational treatment based on the Hamiltonian.[3] In this brief report a variational principle is described that will allow one to overcome these difficulties and to directly obtain the desired solution. Since the customary method of proving a variational principle, which requires a self-adjoint operator, cannot be used here we shall prove the validity of the approach of this paper by direct calculation. This proof is the technical substance of the present report.

II. FORMALISM

We want to solve the Schrödinger equation

$$(H-E)\Psi=0 \tag{1}$$

for continuum states. We shall work with real wave functions.

The following remarks must be made at this point. (i) Equation (1), being a differential equation, must be obeyed locally. This is a much more stringent condition than the vanishing of a matrix element, for example $\langle\Psi|(H-E)|\Psi\rangle$. We will use the locality condition to construct our variational principle. (ii) In the form (1) it does not matter whether the support of the solution is the infinite space (as is the case for continuum solutions) or a compact space (discrete spectrum solutions). Furthermore, our integrations actually involve a compact space; they will be carried out in the region $r_{dt} > a$ in the dtμ region of the nine-dimensional configuration space. Still, one has to account for the fact that one deals with a continuum. This can be taken care of by any one of the available procedures. For reasons of convenience we shall use the Weyl eigendifferential method. It has the advantage of being "transparent," in the sense that the formulas are not changed, except that all matrix elements, e.g., $\langle\psi|\psi\rangle$, are fully defined, even when integrating over the complete infinite space of the continuum solutions. Without further ado, we return to the problem at hand.

Being in the continuum, i.e., above the threshold, (1) has a solution for each energy E. Hence, one has to begin by specifying the energy, say $E = E_0$, at which one wants to solve (1). When using a variational approach, one requires an operator that has a definite sign and has an extremum at the sought solution. This is not the case with the operator of (1); it is, however, true for its square. Hence, the variational problem

$$\delta M=\delta[\langle\Psi|(H-E_0)^2|\Psi\rangle]/\langle\Psi|\Psi\rangle=0 \tag{2}$$

has the appearance that it will converge toward the solutions of (1) at the energy $E = E_0$. Namely, any solution of (1) with $E = E_0$ will yield the value zero for the positive semidefinite functional M of (2), for any region of integration.

The variational principle (2) has, however, an inherent difficulty. Namely, Eq. (1) is a second-order partial differential equation and it has two independent solutions. On the other hand, $(H - E)^2\psi = 0$ is a fourth-order partial differential equation, and has four independent solutions. Even though it is factorizable into $(H-E)(H-E)$, this only means that the two solutions of Eq. (1) (the regular and the irregular) are also solutions of $(H-E)^2$. The two other solutions thus are "hidden." To find them we observe that the solution χ_0 of the inhomogeneous equation

$$(H-E_0)\,\tilde{\chi}_0=\varphi_0, \tag{3a}$$

$$\chi_0=\tilde{\chi}_0-\varphi_0\langle\varphi_0|\tilde{\chi}_0\rangle, \tag{3b}$$

 0022-2488/90/112588-04$03.00

where φ_0 is a solution of (1) with $E = E_0$, fulfills the iterated equation

$$(H-E_0)^2\chi_0=0, \tag{4}$$

but not Eq. (1). Hence, the set χ_0 represents the hidden solutions. Since the solutions φ_E of the equation

$$(H-E)\varphi_E=0 \tag{5}$$

form a complete set the functions χ_0 can be formally expanded as

$$\chi_0(x)\simeq\chi_0(x)=\int dE\ f(E)\varphi_E(x). \tag{6}$$

This expansion converges uniformly except at the singular points of H (e.g., at $r=0$ and $r=\infty$ for a Coulomb problem); hence, the designation for the function $\chi_0(x)$ in (6). Thus, at the points of nonuniform convergence,

$$(H-E_0)\chi_0(x)\neq\varphi_0(x). \tag{7}$$

We will not need the expansion (6). It is only important in that it shows that, if not eliminated, χ_0 in effect introduces solutions with $E\neq E_0$ into the results of the variational wave function, in particular, in numerical work where the integration is performed only over an incomplete (compact) region of space. Therefore, we recognize that the variational principle (2) yields only a necessary, but not a sufficient condition for obtaining a solution of (1) since it cannot distinguish between the solutions φ and χ. We will return to the construction of a functional that will provide the needed sufficient condition. However, we first introduce the accuracy weighting since we will need it when discussing the full variational functional, which we shall denote by $F=\langle|K|\rangle$.

In numerical calculations one is always confronted by the problems associated with the errors contained in the results. As explained in the Introduction, in our problem we require a high accuracy in the value and derivative of the wave function at $r_{dt}\simeq 5$ fm, which, by the way, is one of the boundaries of the domain of integration, implied in the $\langle|\ |\rangle$ notation. The difficulty arises from the fact that this is precisely the region of configuration space (in our case an odd-shaped three-dimensional space, specified by a nonorthogonal coordinate system) where the wave function has very small values. Hence, the variational principle based on the Hamiltonian is very insensitive to errors in the wave function in that region of space. This drawback can be neutralized by modifying the variational functional according to

$$\langle\Psi|K|\Psi\rangle\rightarrow\langle\tilde{\Psi}|K|\Psi\rangle, \tag{8a}$$

$$\tilde{\Psi}=\Psi W, \tag{8b}$$

where $W\geqslant 0$ is a *suitable* (see below) weight function (in our case it could be r_{dt}^{-n} with n a positive integer). We see that (8) is permitted since the operator K is a local operator as long as it is based on the Hamiltonian (1). Hence, if K is chosen such that

$$K\Psi=0, \tag{9}$$

if Ψ fulfills (1), the modified matrix element (8) vanishes also, whatever the weight function W. Of course, in an actual calculation one never deals with an exact solution. In that case the variational procedure will tend to diminish the errors wherever W is large, while allowing errors to increase in the region where W is small.

Note that the introduction of W has no effect on the efficiency, or on any other aspects, of the calculation. In particular, it matters not in the least that the form WK may not be Hermitean. The factor W is to be used strictly as a weight factor, and must not be used in the order KW.

We now return to the variational principle, and analyze the form (2). The variational function has the general form

$$\Psi(x)=\int a(E)\varphi_E(x)dE+\int b(E)\chi_E(x)dE, \tag{10}$$

with φ and χ as defined above and where the desired solution is

$$a(E)=\delta(E-E_0), \tag{11a}$$

$$b(E)=0. \tag{11b}$$

[Actually, (11a) is a Kronecker, and not a Dirac, δ since we use Weyl's method.] Now, in view of (3), we have *locally*

$$(H-E_0)^2\chi_E=2(E-E_0)\varphi_E+(E-E_0)^2\chi_E \tag{12}$$

and

$$(H-E_0)^2\varphi_E=(E-E_0)^2\varphi_E. \tag{13}$$

We now investigate the quantity I:

$$I=\int dx\ \tilde{\Psi}^*(x)(H-E_0)^2\Psi(x)\equiv\langle\tilde{\Psi}|(H-E_0)^2|\Psi\rangle, \tag{14}$$

where the integration is over an arbitrary compact or the full infinite space. Inserting (10) and using (12) and (13) we have

$$\begin{aligned}
I&=\int dx\int dE\,dE'(a(E')\varphi_{E'}(x)+b(E')\chi_{E'}(x))^*W(x)(H-E_0)^2(a(E)\varphi_E(x)+b(E)\chi_E(x))\\
&=\int dx\int dE\,dE'(a(E')\varphi_{E'}(x)+b(E')\chi_{E'}(x))^*W(x)[b(E)\{2(E-E_0)\varphi_E(x)+(E-E_0)^2\chi_E\}\\
&\quad+a(E)(E-E_0)^2\varphi_E]\\
&=\int dE\,dE'[a(E')b(E)\{2(E-E_0)\langle\varphi_{E'}|W|\varphi_E\rangle+((E-E_0)^2+(E'-E_0)^2)\langle\varphi_{E'}|W|\chi_E\rangle\}
\end{aligned}$$

$$+\tfrac{1}{2}b(E')b(E)((E-E_0)^2+(E'-E_0)^2)\langle\chi_{E'}|W|\chi_E\rangle+\tfrac{1}{2}a(E')a(E)((E-E_0)^2+(E'-E_0)^2)\langle\varphi_{E'}|W|\varphi_E\rangle]. \tag{15}$$

We reemphasize the local character of the integrand of (15) which is a consequence of the local character of (12) and (13). We have introduced in (15) the notation

$$\int dx\,\varphi_E^*(x)W(x)\varphi_E(x)\equiv\langle\varphi_E|W|\varphi_E\rangle, \tag{16}$$

etc. Having used (12) and (13) the integrand of (15) is simply an ordinary function (as a consequence of Weyl's method no generalized functions arise). Hence, no subtleties are associated with the process of integration, Eq. (16). We emphasize here that as can be seen by inspection from (14), I is a positive semidefinite quantity since the integrand is *locally* a positive semidefinite quantity. Owing to the presence of the cross term in (12), this fact is not immediately evident in the last form of (15).

Coming back to our actual computational problem we note that since in numerical work the overlaps never extend over the complete space, the overlap integrals do not reflect the orthogonality relations, even for the case $W(x)=1$, and in general none of the overlap integrals vanish. Hence, (15) must be carefully discussed.

To begin with, assume that the starting function Ψ is reasonable, i.e., that it is close to fulfilling (11); more specifically, assume that over the energy interval where $a(E)$ is not negligibly small the overlap $\langle\varphi_E|W|\varphi_E\rangle$ and also the overlap $\langle\chi_E|W|\chi_E\rangle$ retain their positive sign, valid for $E'=E=E_0$. In that case the third and the fourth terms will have manifestly positive semidefinite integrands. The overall contribution of all terms can still only be positive in view of the positive semidefinite character of I. However, the condition (11b), i.e., the elimination of the contribution of χ_E from our solution, specifically the case $b(E)=b_0\delta(E-E_0)$, is not ensured since all contributions (positive and negative) vanish for $E=E_0$. Still, for $E\neq E_0$ the variational principle (2) can be used to eliminate all $a(E)$, $b(E)$ from the variational function Ψ. One then is left with

$$a(E)=a_0\delta(E-E_0) \tag{17a}$$

and

$$b(E)=b_0\delta(E-E_0). \tag{17b}$$

To eliminate the latter, one requires another variational principle, i.e., one which projects to φ_0 but not to χ_0. This then is provided by the functional

$$J=\langle\tilde{\Psi}|(H-E_0)|\Psi\rangle^2. \tag{18}$$

Taking into account (17a) and (17b) we find

$$J=[a_0b_0\langle\tilde{\varphi}_0|W|\varphi_0\rangle+b_0^2\langle\tilde{\chi}_0|W|\varphi_0\rangle]^2, \tag{19a}$$

which then enforces also $b_0=0$, by evaluating the matrix elements in (19a) with two different weight functions W_i such that

$$\frac{\langle\varphi_0|W_1|\varphi_0\rangle}{\langle\varphi_0|W_2|\varphi_0\rangle}\neq\frac{\langle\chi_0|W_1|\varphi_0\rangle}{\langle\chi_0|W_2|\varphi_0\rangle}. \tag{19b}$$

Now a remark concerning the abovementioned question of the suitability of a weight function. Since, in general, WK is not a Hermitean operator the point that $H-E$ is a local operator has to be used. More particularly, the overlap $\langle W\psi|K|\psi\rangle$ can achieve the value zero not only as a consequence of the desired case $K\psi=0$ but also by cancellations between positive and negative contributions of the integral. In particular, an "unsuitable" weight factor W by this mechanism could make J, Eq. (18) vanish, without achieving a solution of (1). Even though the functional I would guard against it, such a W could lead to instabilities in the calculation. The "suitability" of a chosen W must be checked by numerical tests. This completes the description of the variational principle. We now turn to the last problem, viz., the resolution of the degeneracy of the states.

Since the number of independent continuum solutions equals the number of open channels, which, depending on the system, may be finite or infinite, in general, one needs a prescription on how to select the solutions of interest without having to compute the full set of solutions. This prescription must be supplied by the physics of the problem. The prescription then may be expressed as an auxiliary condition, a constraint, specified by an operator Ω, which can be added by Lagrange's procedure to the variational problem. (In our case the fusion reaction proceeds through a unique doorway state that decays electromagnetically to the fusion state. The prescription here thus is:[4] maximize the square of the electromagnetic transition amplitude from the doorway state.) The constraint functional then can be written as

$$B=-\langle\Psi|\Omega|\Psi\rangle/\langle\Psi|\Psi\rangle=\text{Min}, \tag{20}$$

where Ω is the operator that splits the degeneracy in the desired fashion; it has to be constructed to have a positive definite matrix element. (In our case

$$\Omega=|H_{\text{int}}|\Psi_d\rangle\langle\Psi_d|H_{\text{int}}|, \tag{21}$$

where H_{int} is the electromagnetic transition operator, and Ψ_d is the wave function of the doorway state.[4] Inserting in (20) the exact solution one sees that for our case it equals minus the negative of the transition probability, and hence at the minimum it maximizes the transition probability, i.e., it picks out that linear combination of degenerate solutions at $E=E_0$, through which the reaction proceeds.)

In general, the operator Ω will be a function of E; it thus will pull the computed minimum of the full variational problem toward the rising side of its own absolute value. To check this tendency one can give the individual terms different weights, say L and $M>0$. Hence, we write for the final expression

$$L\left\{\frac{\langle\tilde{\Psi}|(H-E_0)^2|\Psi\rangle}{\langle\tilde{\Psi}|\Psi\rangle}+\frac{\langle\tilde{\tilde{\Psi}}|(H-E_0)|\Psi\rangle^2}{\langle\tilde{\Psi}|\Psi\rangle^2}\right\}$$

$$-M\frac{\langle\Psi|\Omega|\Psi\rangle}{\langle\Psi|\Psi\rangle}=\text{Min.} \tag{22}$$

The tilde and double-tilde notation indicates that the weight function in the two terms may be different; the degeneracy-breaking term has $W=1$.

Performing the calculation for different ratios M/L, maintaining, e.g., $L+M=2$, one can check for the stability of the obtained results. (The actual solution formally is given by the limes $M\to 0$.) This test will be important if B, Eq. (20) has a steep energy dependence. At any rate, one always can use (1) to compute the errors in the solution in the region of interest since (1) is a local equation, i.e., it must be fulfilled at every point in position space. This way one can check whether the achieved accuracy fulfills the needs at hand.

ACKNOWLEDGMENTS

I would like to thank L. C. Biedenharn, S. Haywood, H. Monkhorst, J. Rafelski, A. Stahlhofen, and K. Szalewicz for many useful discussions.

[1] See, e.g., *Proceedings of the Muon Catalyzed Fusion Workshop*, Sanibel Island, AIP Conference Proceedings, Vol. 181, edited by S. E. Jones, J. Rafelski, and H. Monkhorst (AIP, New York, 1989).

[2] G. M. Hale, N. Jarmie, and R. E. Brown, Phys. Rev. Lett. **59**, 763 (1987).

[3] For a review of the different available methods see, e.g., R. K. Nesbitt, *Variational Methods in Electron-Atom Scattering Theory* (Plenum, New York, 1980).

[4] A full discussion of this case including the proof of the uniqueness of the solution is contained in M. Danos, A. Stahlhofen, and L. C. Biedenharn, Ann. Phys. (NY) **192**, 158 (1989).

Physical constraints on the coefficients of Fourier expansions in cylindrical coordinates

H. Ralph Lewis[a)]
Los Alamos National Laboratory, Los Alamos, New Mexico 87545

Paul M. Bellan
California Institute of Technology, Pasadena, California 91125

(Received 12 March 1990; accepted for publication 20 June 1990)

It is demonstrated that (i) the postulate of infinite differentiability in Cartesian coordinates and (ii) the physical assumption of regularity on the axis of a cylindrical coordinate system provide significant simplifying constraints on the coefficients of Fourier expansions in *cylindrical coordinates*. These constraints are *independent* of any governing equations. The simplification can provide considerable practical benefit for the analysis (especially numerical) of actual physical problems. Of equal importance, these constraints demonstrate that if **A** is *any arbitrary* physical vector, then the *only* finite Fourier terms of A_r and A_θ are those with $m = 1$ symmetry. In the Appendix, it is further shown that postulate (i) may be inferred from a more primitive assumption, namely, the arbitrariness of the location of the cylindrical axis of the coordinate system.

I. INTRODUCTION

In mathematical physics, changing from one set of independent variables to a different, but equivalent, set is a matter of convenience only, since all physical results must be independent of the choice of coordinate system. Frequently, the choice of independent variables is motivated by some symmetry of the problem. For example, if there is a line source, or if boundary conditions are specified on a circular cylinder, then cylindrical coordinates are likely most appropriate. However, the choice of a particular set of independent variables (e.g., the r, θ, z of cylindrical coordinates) might inadvertently introduce mathematically allowable but physically unrealistic terms—e.g., singularities at the axis. These nonphysical terms must be eliminated by the imposition of *physical constraints* on the mathematical solutions.

Let us briefly review two traditional methods by which these constraints are imposed.

(1) Analytic method: First, a set of equations relevant to the problem at hand is derived; second, general *mathematical* solutions to these equations are found; third, some subset of these solutions is discarded as being nonphysical, and finally, boundary conditions are used to determine a suitable combination of the remaining physically allowable solutions to describe the specific problem.

(2) Numerical method: Again, a set of equations relevant to the problem is derived; second, the space in which the problem is to be solved is quantized into grid points; third, the equations are put into discrete form so as to establish a numerical algorithm; fourth, a numerical solution is developed using the algorithm. Mathematically allowable but nonphysical solutions are eliminated by constraints determined in an *ad hoc* manner. For example, a simplified but local analytic solution might be developed in the neighborhood of the axis, and then grafted onto the numerical solution outside this region.

A well-known example of the analytical method is where one Fourier analyzes in z and θ linear partial differential equations so as to obtain an ordinary differential equation in r for the Fourier coefficients. Physical constraints are imposed in order that solutions of the equations be regular in the region of interest. The most obvious example of this method is that involving Bessel's equation: Although both the J_m and Y_m Bessel functions satisfy Bessel's equation, for physical problems including the axis, one rejects the Y_m solution because the Y_m solution is singular there.

The point we wish to make in this paper is that *the physical constraints are more fundamental than the differential equations and can be determined without reference to differential or any other equations.* In this paper we restrict our attention to cylindrical coordinates. However, similar considerations apply to all variable transformations, those of a nongeometrical nature as well as those of a geometrical nature.

There are three reasons why it is important for one to be able to determine the constraints without reference to any governing equations. First, even if governing equations are known and could, in principle, be used to derive the constraints (e.g., by rejecting singular solutions), it is generally more convenient to be able to assume from the outset the form required of each quantity. In numerical computation of physical problems especially, singular nonphysical terms are a serious difficulty since small finite differencing or truncation errors easily excite the unwanted terms. Because these nonphysical terms are divergent, they mask the desired physical solution and render the computation useless. It is clearly essential in these situations to be able to invoke constraints that eliminate the unwanted divergences. (Often in contemporary work, if a problem is sufficiently complicated, numerical analysts resort to the messy stratagem of solving

[a)] Temporary address: U.S. Department of Energy, Office of Energy Research, Office of Fusion Energy, Washington, DC 20545.

 0022-2488/90/112592-05$03.00

the problem using a less appropriate coordinate system solely to avoid difficulties with these nonphysical singularities—e.g., Cartesian coordinates are often used for problems that have cylindrical symmetry simply to avoid singularities at the axis. Clearly, it would be preferable to be able to invoke appropriate physical constraints and so use the more natural cylindrical coordinates.)

Second, although the governing equations might be known, they may be so complicated that it is not feasible to use them to provide constraints. Indeed, the considerations in this paper originated from a numerical computation problem in plasma physics where the governing equations were integrodifferential equations in a cylindrical coordinate system. Because these equations were nonlocal, it was not possible to determine the behavior of solutions at the axis and so guarantee regularity.

Third, it may be that the governing equations are unknown or even that a solution to the governing equations is not sought. The results of this paper have been applied[1] to reduce in a most substantial and significant manner the complexity of the theoretical description of a magnetohydrodynamic dynamo problem where the precise form of the governing equations was not well established.

II. SYMMETRY CONSTRAINT

Before we launch into our discussion of *regularity,* it is worthwhile to derive a purely mathematical *symmetry* constraint on Fourier coefficients. Later, we will combine this result with the regularity analysis to determine the least restrictive physically permissible Fourier coefficients.

We consider the transformation between Cartesian coordinates (x,y,z) and cylindrical coordinates (r,θ,z) given by

$$x = r\cos\theta, \quad y = r\sin\theta. \tag{1}$$

In our considerations here and in the rest of this paper the z coordinate will never play a role; thus we hold it constant and henceforth suppress it in the notation. The Cartesian coordinates of a point are not changed by replacing r by $-r$ and θ by $\theta + \pi$ in Eq. (1). This symmetry of the transformation requires that any point function of (r,θ), say $f(r,\theta)$, satisfy

$$f(r,\theta) = f(-r,\theta+\pi). \tag{2}$$

If we make a Fourier series representation of $f(r,\theta)$,

$$f(r,\theta) = \sum_{m=-\infty}^{\infty} a_m(r)e^{im\theta}, \tag{3}$$

we see that in order to satisfy Eq. (2), we must have

$$a_m(r) = (-1)^m a_m(-r). \tag{4}$$

In words, if m is even, then a_m is an even function of r, whereas if m is odd, then a_m is an odd function of r. This symmetry result also restricts the form of physically allowed linear differential equations for $a_m(r)$ to be such as not to not alter the parity of $a_m(r)$; all terms in a linear differential equation must be odd in r or else all even in r (if the equation is nonlinear, then this restriction need not apply).

III. REGULARITY CONSTRAINT: SCALARS

Let us assume that $\Psi(r,\theta)$ is a physical scalar, regular at $r=0$, and express Ψ as a Fourier series with respect to θ. Thus we write

$$\Psi(r,\theta) = \sum_{m=-\infty}^{\infty} a_m(r)\exp(im\theta) = \sum_{m=-\infty}^{\infty} \Psi_m(r,\theta). \tag{5}$$

We now show that $a_m(r)$ must have a very specific type of r dependence. We use the assumption that each coefficient $\Psi_m(r,\theta)$ is a regular (i.e., infinitely differentiable) function of (x,y) at $r=0$. The function $\exp(im\theta)$ is not a regular function of (x,y) at $r=0$. However, the function $[r\exp(\pm im\theta)]^{|m|} = (x \pm iy)^{|m|}$ is obviously a regular function of (x,y) because it is a polynominal in (x,y). We express Ψ_m as

$$\Psi_m = a_m(r)\exp(im\theta) = [a_m(r)/r^{|m|}][r\exp(\pm i\theta)]^{|m|}$$
$$= [a_m(r)/r^{|m|}](x \pm iy)^{|m|}, \tag{6}$$

where the $+$ sign is used if $m>0$ and the $-$ sign is used if $m<0$. Since $(x \pm iy)^{|m|}$ is regular, we must require that $a_m(r)/r^{|m|}$ be regular. For $a_m(r)/r^{|m|}$ to be nonsingular, we must require

$$a_m(r) \sim r^{|m|} \text{ as } r \to 0. \tag{7}$$

The symmetry constraint of Eq. (4) then gives

$$a_m(r) = r^{|m|} f_m(r^2), \tag{8}$$

where $f_m(r^2)$ is a regular function of r^2 and so has a Taylor expansion

$$f_m(r^2) = f_m^{(0)} + f_m^{(2)}r^2 + f_m^{(4)}r^4 + \cdots. \tag{9}$$

If odd powers of r were present in Eq. (9) then, because $r = \sqrt{x^2+y^2}$, $f_m(r^2)$ would not be a regular function of (x,y).

The analysis in this section was based on the postulate that all physical quantities when expressed in Cartesian coordinates are infinitely differentiable on the axis. It is shown in the Appendix that this postulate may be inferred from a more primitive assumption, namely, the assumption that the location of the cylindrical axis of a coordinate system for describing a physical system is arbitrary.

IV. REGULARITY CONSTRAINTS: VECTORS

Let $\mathbf{A}$ be a vector representing a physical quantity. We will again exploit the transformation from cylindrical to Cartesian coordinates to determine the functional form of cylindrical components of $\mathbf{A}$. We do this by writing

$$\mathbf{A} = A_r\hat{r} + A_\theta\hat{\theta} + A_z\hat{z}$$
$$= A_r(\cos\theta,\sin\theta,0)$$
$$+ A_\theta(-\sin\theta,\cos\theta,0) + A_z(0,0,1), \tag{10}$$

in which case

$$A_x = A_r\cos\theta - A_\theta\sin\theta. \tag{11}$$

Let us now expand A_r and A_θ in terms of their Fourier terms in θ:

$$A_r = \sum_{m=-\infty}^{\infty} A_{rm} \exp(im\theta),$$

$$A_\theta = \sum_{m=-\infty}^{\infty} A_{\theta m} \exp(im\theta). \tag{12}$$

Combining Eqs. (11) and (12), we obtain

$$A_x = \frac{1}{2} \sum_{m=-\infty}^{\infty} [(A_{rm} + iA_{\theta m}) e^{i(m+1)\theta} + (A_{rm} - iA_{\theta m}) e^{i(m-1)\theta}] \tag{13a}$$

or

$$A_x = \frac{1}{2} \sum_{m=1}^{\infty} [(A_{rm} + iA_{\theta m}) e^{i(|m|+1)\theta} + (A_{rm} - iA_{\theta m}) e^{i(|m|-1)\theta}] + \frac{1}{2} [(A_{r0} + iA_{\theta 0}) e^{i\theta} + (A_{r0} - iA_{\theta 0}) e^{-i\theta}]$$
$$+ \frac{1}{2} \sum_{m=-1}^{-\infty} [(A_{rm} + iA_{\theta m}) e^{-i(|m|-1)\theta} + (A_{rm} - iA_{\theta m}) e^{-i(|m|+1)\theta}]. \tag{13b}$$

We may rewrite Eq. (13b) as

$$A_x = \frac{1}{2} \sum_{m=1}^{\infty} \left[(A_{rm} + iA_{\theta m}) \frac{(x+iy)^{|m|+1}}{r^{|m|+1}} + (A_{rm} - iA_{\theta m}) \frac{(x+iy)^{|m|-1}}{r^{|m|-1}} \right]$$
$$+ \frac{1}{2} \left[(A_{r0} + iA_{\theta 0}) \frac{(x+iy)}{r} + (A_{r0} - iA_{\theta 0}) \frac{(x-iy)}{r} \right]$$
$$+ \frac{1}{2} \sum_{m=-1}^{-\infty} \left[(A_{rm} + iA_{\theta m}) \frac{(x-iy)^{|m|-1}}{r^{|m|-1}} + (A_{rm} - iA_{\theta m}) \frac{(x-iy)^{|m|+1}}{r^{|m|+1}} \right]. \tag{14}$$

We require each term in Eq. (14) to be regular. Only positive powers of $x \pm iy$ occur and these factors are all regular. Thus we must require the regularity of the remaining factors. We will consider $m > 0$, $m = 0$, and $m < 0$ separately.

Case (i), $m = 0$: Here, we require that $(A_{r0} \pm A_{\theta 0})/r$ be regular, so that we must have both $A_{r0} \sim r$ and $A_{\theta 0} \sim r$ as $r \to 0$.

Case (ii), $m > 0$: For regularity we must have, as $r \to 0$, both

$$A_{rm} + iA_{\theta m} \sim r^p, \text{ where } p \geqslant |m| + 1 \tag{15a}$$

and

$$A_{rm} - iA_{\theta m} \sim r^q, \text{ where } q \geqslant |m| - 1, \tag{15b}$$

since both these terms occur for $m > 0$. We could let $p = q = |m| + 1$, but this is not the least restrictive possibility. To obtain the least restrictive possibility, we satisfy Eq. (15b) by letting $q = |m| - 1$ with both $A_{rm} \sim r^{|m|-1}$ and $A_{\theta m} \sim r^{|m|-1}$ as $r \to 0$. However, Eq. (15a) will then be violated unless we set $A_{rm} + iA_{\theta m} = 0$ for terms of order $r^{|m|-1}$. Thus the least restrictive allowable form is

$$A_{rm} = \lambda_m r^{|m|-1} + r^{|m|+1} g_m(r^2),$$
$$A_{\theta m} = i\lambda_m r^{|m|-1} + r^{|m|+1} h_m(r^2), \tag{16}$$

where λ_m is a constant and $g_m(r^2)$ and $h_m(r^2)$ are regular functions of r^2 of the form given by Eq. (9). [We have used the symmetry constraint of Eq. (4) as we did earlier when treating scalars.]

Case (iii), $m < 0$: Here, examination of Eq. (14) shows that we must require as $r \to 0$

$$A_{rm} + iA_{\theta m} \sim r^q, \text{ where } q \geqslant |m| - 1 \tag{17a}$$

and

$$A_{rm} - iA_{\theta m} \sim r^p, \text{ where } p \geqslant |m| + 1, \tag{17b}$$

since both these terms occur for $m < 0$. Using the same type of argument as for the $m > 0$ case above we find here that we must have

$$A_{rm} = \lambda_m r^{|m|-1} + r^{|m|+1} g_m(r^2),$$
$$A_{\theta m} = -i\lambda_m r^{|m|-1} + r^{|m|+1} h_m(r^2). \tag{18}$$

Since A_z behaves like a scalar, it is described by Eq. (8). Combining the results derived above we find that the general form for a vector is, for $m \neq 0$,

$$A_{rm} = \lambda_m r^{|m|-1} + r^{|m|+1} g_m(r^2),$$
$$A_{\theta m} = i \operatorname{sgn}(m) \lambda_m r^{|m|-1} u_m(r^2) + r^{|m|+1} h_m(r^2),$$
$$A_{zm} = r^{|m|} f_m(r^2), \tag{19a}$$

while, for $m = 0$,

$$A_{r0} = r g_0(r^2), \quad A_{\theta 0} = r h_0(r^2), \quad A_{z0} = f_0(r^2). \tag{19b}$$

The peculiar form of the λ_m terms in Eq. (19a) can also be derived by requiring that the divergence and curl of the vector are always finite and then examining these quantities as $r \to 0$. For example, if we require

$$\nabla \cdot \mathbf{A} = \text{finite}, \tag{20}$$

then, as $r \to 0$, Eq. (20) becomes

$$\lim_{r \to 0} \left[\frac{\partial (rA_r)}{\partial r} + \frac{\partial A_\theta}{\partial \theta} \right] = 0. \tag{21}$$

If we assume that $A_r \sim \lambda r^{|m|-1}$, where $m \geqslant 1$, and that both A_r and $A_\theta \sim \exp(im\theta)$, then we obtain $A_\theta = i \operatorname{sgn}(m) A_r$ consistent with Eq. (19a). A similar observation occurs when we calculate $(\nabla \times \mathbf{A})_z$.

V. IMPORTANCE OF $m=1$ MODES

Equation (19a) has the very interesting physical consequence of showing that $|m| = 1$ Fourier terms form a special class, because only these modes can have A_r and A_θ finite at $r = 0$. These $|m| = 1$ modes are observed to be important in magnetohydrodynamic (MHD) instabilities present in tokamaks,[2] reversed field pinches,[1] and spheromaks.[3]

In many physically interesting problems, one is interested in quantities of the form

$$\hat{z}\cdot\mathbf{A}\times\mathbf{B} = A_r B_\theta - A_\theta B_r.$$

For example, in magnetohydrodynamics, Ohm's law with the Hall term included has the form

$$\mathbf{E} + \mathbf{U}\times\mathbf{B} - (1/ne)\mathbf{J}\times\mathbf{B} = \eta\mathbf{J}, \tag{22}$$

where **J**, **B**, **E**, and **U** are, respectively, the electric current density, the magnetic field, the electric field, and the velocity field; η is the resistivity, n is the charged particle density, and e is the charge on an electron. In dynamo theory[1,4,5] and MHD rf current drive schemes,[6,7] one searches for ways of driving a dc current on the rhs of Eq. (22) using only ac quantities on the lhs. If this dc current is to be directed in the z direction of a cylindrical coordinate system, then we take the z component of Eq. (22)

$$\langle \tilde{E}_z + \hat{z}\cdot\tilde{\mathbf{U}}\times\tilde{\mathbf{B}} - (1/ne)\hat{z}\cdot\tilde{\mathbf{J}}\times\tilde{\mathbf{B}}\rangle = \eta J_{z\mathrm{dc}}, \tag{23}$$

where $\langle\ \rangle$ denotes time average and tilde denotes ac quantity. Since only the $|m| = 1$ terms in the θ and z components of the ac vectors are finite at $r = 0$, we immediately know that only $|m| = 1$ terms of the ac fields offer the possibility of driving dc currents along the z axis. This removes from consideration the infinity of modes for which $|m| \neq 1$ and so provides an enormous simplification[1] of the problem, even before having to determine the relevant equations for **U**, **J**, **B**.

ACKNOWLEDGMENTS

The authors are indebted to J. L. Schwarzmeier, R. A. Nebel, and M. A. Schalit for many stimulating conversations.

This work was supported by the U.S. Department of Energy.

APPENDIX: INFERENCE OF INFINITE DIFFERENTIABILITY FROM A MORE PRIMITIVE ASSUMPTION

The analysis in the main body of this paper was based on the postulate that all physical quantities when expressed in Cartesian coordinates are infinitely differentiable on the axis. We show here that this postulate may be inferred from a more primitive assumption, namely, the assumption that the location of the cylindrical axis of a coordinate system describing a physical system is arbitrary.

The basic requirement is that the expression for a quantity in terms of a particular set of variables do not exhibit peculiarities which are nonphysical and only a result of the choice of variables. We consider a quantity represented by a nonsingular Fourier series with respect to the angle variable of a cylindrical coordinate system. Suppose on physical grounds that the quantity in question could be represented equally well by a Fourier series with respect to the angle variable of *any* other cylindrical coordinate system, whose cylindrical axis is parallel to the original axis, but is located anywhere in a neighborhood of the original axis. That is, we suppose that the location of the cylindrical coordinate axis is *irrelevant* to the physical problem; the location is chosen purely for convenience.

For studying the differentiability at $r = 0$ of Fourier series, it suffices to take the case where $f(r,\theta)$ and the coefficients $a_m(r)$ are scalars. This is because those quantities can represent an arbitrary Cartesian component of a tensor. Any Fourier series in the variable θ is manifestly infinitely differentiable with respect to θ because $\exp(im\theta)$ is infinitely differentiable. We now suppose that the quantity represented by $f(r,\theta)$ is nonsingular and could be equally well represented by a Fourier series with respect to the angle variable of any other cylindrical coordinate system whose z axis is parallel to the original z axis, but is located anywhere in a neighborhood of the original axis. This is a weak assumption. If the quantity of physical interest were singular, then either a related nonsingular function could be defined (for example, by subtracting the electric potential due to a line charge located on the cylindrical axis), or the singular quantity could be viewed as the limit of a sequence of nonsingular quantities. The results presented here would apply to each member of the sequence and, therefore, to the singular limit as well.

Consider two equivalent cylindrical coordinate systems, as illustrated in Fig. 1. The coordinates of the chosen cylindrical system are denoted by (r,θ,z) and the associated Cartesian coordinates are (x,y,z). The z axis is normal to the plane of the drawing and passes through the point labeled A. The coordinates of the other cylindrical system are (r_n,θ_n,z) and the associated Cartesian coordinates are (x_n,y_n,z). The axis of that system has coordinates

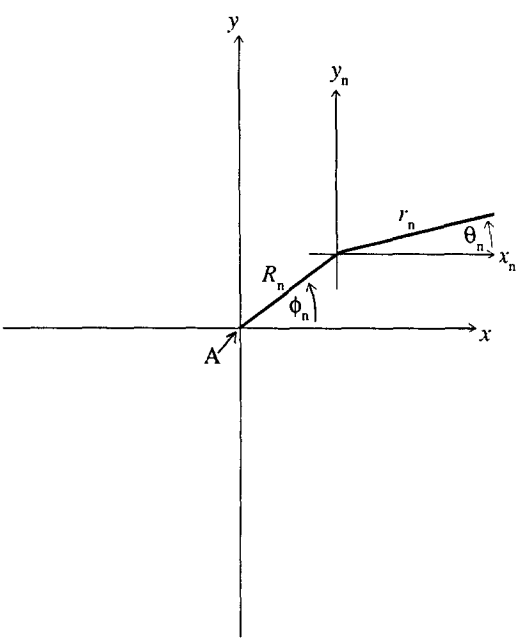


FIG. 1. Relation of the (x_n,y_n) coordinate system to the (x,y) coordinate system.

$$r = R_n, \quad \theta = \phi_n \tag{A1}$$

in the chosen cylindrical system; these coordinates are arbitrary. The coordinates (r_n,θ_n) of the axis A are

$$r_n = R_n, \quad \theta_n = \phi_n + \pi. \tag{A2}$$

Denote the function $f(r,\theta)$ expressed in terms of (r_n,θ_n) by $f_n(r_n,\theta_n)$ and expressed in terms of (x,y) by $F(x,y)$; by assumption, $f_n(r_n,\theta_n)$ possesses a Fourier series in the variable θ_n and is thus infinitely differentiable with respect to θ_n. The derivatives with respect to r_n and θ_n can be expressed in terms of derivatives with respect to x_n and y_n by repeated application of the formulas

$$\frac{\partial}{\partial r_n} = \cos\theta_n \frac{\partial}{\partial x_n} + \sin\theta_n \frac{\partial}{\partial y_n}, \tag{A3}$$

$$\frac{\partial}{\partial \theta_n} = -r_n \sin\theta_n \frac{\partial}{\partial x_n} + r_n \cos\theta_n \frac{\partial}{\partial y_n}. \tag{A4}$$

Since the (x_n,y_n) system is related to the (x,y) system by a simple translation, we have

$$\frac{\partial}{\partial x_n} = \frac{\partial}{\partial x}, \quad \frac{\partial}{\partial y_n} = \frac{\partial}{\partial y} \tag{A5}$$

and, in particular, we can write Eq. (A4) as

$$\frac{\partial}{\partial \theta_n} = -r_n \sin\theta_n \frac{\partial}{\partial x} + r_n \cos\theta_n \frac{\partial}{\partial y}. \tag{A6}$$

Using Eq. (A6), the derivative

$$\frac{\partial^p}{\partial \theta_n^p} f_n(r_n,\theta_n)$$

can be expressed in terms of derivatives with respect to x and y of $F(x,y)$; that expression will contain the $(p+1)$ derivatives of $F(x,y)$ of order p. By choosing $(p+1)$ different axes (R_n,ϕ_n) and evaluating the pth derivative with respect to θ_n for each of the $(p+1)$ functions $f_n(r_n,\theta_n)$ at the axis A, we obtain a linear system of $(p+1)$ equations for the $(p+1)$ derivatives of $F(x,y)$ of order p at the axis A. Solution of that system gives the derivatives of $F(x,y)$ of order p in terms of the derivatives of $F(x,y)$ of lower order and the derivatives

$$\frac{\partial^p}{\partial \theta_n^p} f_n(r_n,\theta_n),$$

all evaluated at the axis A. In this way, beginning with $p = 1$, we can evaluate all of the derivatives of $F(x,y)$ of order p for each successive value of p. The linear system is soluble for nearly any choice of $(p+1)$ axes. For example, for $p = 1$, the linear system is

$$-R_1 \sin(\phi_1 + \pi) F_x(0,0) + R_1 \cos(\phi_1 + \pi) F_y(0,0) = \frac{\partial}{\partial \theta_1} f_1(r_1,\theta_1)\Big|_{\substack{r_1 = R_1 \\ \theta_1 = \phi_1 + \pi}}, \tag{A7}$$

$$-R_2 \sin(\phi_2 + \pi) F_x(0,0) + R_2 \cos(\phi_2 + \pi) F_y(0,0) = \frac{\partial}{\partial \theta_2} f_2(r_2,\theta_2)\Big|_{\substack{r_2 = R_2 \\ \theta_2 = \phi_2 + \pi}}. \tag{A8}$$

The determinant of the coefficients is

$$R_1 R_2 \sin(\phi_2 - \phi_1), \tag{A9}$$

so that $F_x(0,0)$ and $F_y(0,0)$ can be determined from Eqs. (A7) and (A8) as long as $R_1 \neq 0$, $R_2 \neq 0$, and $\phi_1 \neq \phi_2$. By showing how to evaluate all of the derivatives of $F(x,y)$ at the axis A, we have shown that $F(x,y)$ is infinitely differentiable at the axis A solely as a result of the arbitrariness of the location of the cylindrical axis of the coordinate system and the postulate of a Fourier representation with respect to the angle variable in each of the possible cylindrical coordinate systems.

[1] E. J. Caramana, R. A. Nebel, and D. D. Schnack, Phys. Fluids **26**, 1305 (1983).
[2] M. N. Rosenbluth, D. A. Monticello, H. R. Strauss, and R. B. White, Phys. Fluids **19**, 1987 (1976).
[3] A. Janos, G. W. Hart, and M. Yamada, Phys. Rev. Lett. **55**, 2868 (1985).
[4] E. Hotta, M. Suzuki, H. Ohta, and I. Hayashi, Jpn. J. Appl. Phys. **24**, 110 (1985).
[5] M. J. Dutch, A. L. McCarthy, and R. G. Storer, Phys. Rev. Lett. **56**, 1563 (1986); M. J. Dutch and A. L. McCarthy, Plasma Phys. Controlled Fusion **28**, 695 (1986).
[6] T. Ohkawa, Comments Plasma Phys. Controlled Fusion **12**, 165 (1989).
[7] M. A. Schalit and P. M. Bellan, Phys. Fluids B **1**, 2533 (1989).

Bäcklund transformations for generalized nonlinear Schrödinger equations

J. G. Kingston and C. Sophocleous
Department of Mathematics, University of Nottingham, University Park, Nottingham NG7 2RD, England

(Received 18 January 1990; accepted for publication 30 May 1990)

A general class of Bäcklund transformations are considered for equations of the form $iz_y + z_{xx} + f(z,\bar{z}) = 0$, where $f(z,\bar{z})$ is a function of $z = x + iy$ and $\bar{z} = x - iy$. The nonlinear forms of this equation that admit such transformations are completely classified and shown to exist only when $f(z,\bar{z}) = z^2\bar{z}$ (the nonlinear Schrödinger equation), $z \ln \bar{z}$, $z \ln z$, $(z+\bar{z})^2$, or suitable combinations of these functions. The form $f(z,\bar{z}) = (z+\bar{z})^2$ leads to auto-Bäcklund transformations for the Boussinesq equation.

I. INTRODUCTION

McLaughlin and Scott[1] proved the elegant result that the only equations of the form $z_{xy} = F(z)$ admitting any members of a wide class of auto-Bäcklund transformations were those for which $F''(z) = \lambda F(z)$, λ constant. More recently, Nimmo and Crighton[2,3] obtained general results concerning the existence of Bäcklund transformations (BT's) for the cylindrical Korteweg–de Vries equation, $u_{xxx} + 6uu_x + u_y + m(u,x,y) = 0$, and the nonlinear parabolic equation, $u_t + u_{xx} + H(u_x,u,x,t) = 0$, respectively.

The present paper is in the spirit of the above work and explores the range of functions $f(z,\bar{z})$ for which

$$iz_y + z_{xx} + f(z,\bar{z}) = 0 \tag{1.1}$$

admits BT's of a given general class. Here z is complex so that (1.1) may also be regarded as two coupled real partial differential equations. The complex equation (1.1) is a generalization of the well-known nonlinear Schrödinger equation (NLSE), $iz_y + z_{xx} + z^2\bar{z} = 0$. Lamb[4] derived BT's for the NLSE using the method of Clairin.[5] In the process he made certain simplifying assumptions, suggesting that it might be advantageous to look at other cases, although it is seen later that this does not turn out to be the case. Indeed the more general equation (1.1) is here exhaustively investigated with the conclusion that f must be one of the forms $z^2\bar{z}$, $z \ln \bar{z}$, $z \ln z$, $(z+\bar{z})^2$, or suitable combinations of these functions. The logarithmic nonlinearity $z \ln \bar{z}$ occurs in wave equations derived from quantum models (see, e.g., Refs 6 and 7) and Steudl[8] has discussed invariant transformations of (1.1) when $f(z,\bar{z}) = z \ln z$. The BT's obtained for the case $f(z,\bar{z}) = (z+\bar{z})^2$ lead to auto-BT's for the Boussinesq equation in the form $u_{yy} + u_{xxxx} + (4u^2)_{xx} = 0$, $u(x,y)$ being the real part of $z(x,y)$. Fukushima *et al.*[9] study a nonlinear transmission line and modeled the circuit by a modified NLSE equivalent to $f(z,\bar{z}) = (z\bar{z})^{1/2}z$ in (1.1). They observe envelope solitons and obtain good agreement between experiment and theory. The results of this paper show that BT's of a wide class do not exist for this equation.

The calculations involved in making these deductions have been greatly facilitated by the computer algebraic manipulation package REDUCE.

In Sec. II the class of BT's considered is defined and in Sec. III a set of functional equations is derived, which are then investigated for three exclusive cases in Secs. IV–VI.

II. A CLASS OF BÄCKLUND TRANSFORMATIONS

We consider BT's of the form

$$p = \psi(z,\bar{z},z',\bar{z}',p',\bar{p}',q',\bar{q}'), \tag{2.1}$$

$$q = \phi(z,\bar{z},z',\bar{z}',p',\bar{p}',q',\bar{q}'), \tag{2.2}$$

which relate the two equations

$$iq + r + f(z,\bar{z}) = 0 \tag{2.3}$$

and

$$iq' + r' + f'(z',\bar{z}') = 0, \tag{2.4}$$

where, with the usual notation $p = z_x$, $q = z_y$, $r = z_{xx}$, $s = z_{xy}$, and $t = z_{yy}$, and similarly for p', q', r', s', and t' in terms of z'. If f and f' are the same functions the BT is an auto-Bäcklund transformation. As long as p and q occur explicitly (2.1) and (2.2) represent the most general pair of complex functional relations connecting the dependent variables z and z' and their derivatives. This is a very wide class of BT's. An example of a BT outside this class is given by Kingston and Rogers,[10] where the BT given above is essentially amalgamated with a hodograph-type transformation.

Equations (2.1)–(2.4) enable p, q, r, and r' (and their conjugates) to be expressed in terms of z, z', p', and q' (and their conjugates). We require this system of equations to be consistent in the sense that no new relations between these variables and s', t' and their conjugates are implied. Since p, q, and r are connected by

$$r = \frac{\partial p}{\partial x}, \quad \frac{\partial p}{\partial y} = \frac{\partial q}{\partial x}, \tag{2.5}$$

these relations must therefore be identities in z, z', p', q', s', and t' and their conjugates. It is these two identities that enable the BT's to be derived and that ultimately impose restrictions on the functional forms of f, f', ψ, and ϕ.

Since the main advantage of BT's is that they enable progress to be made in the study of nonlinear equations we will assume that (2.3) is nonlinear and exclude cases in which $f(z,\bar{z})$ is linear in both z and $\bar{z}$.

Note that if (2.1)–(2.5) can be satisfied for suitable functions f, f', ψ, and ϕ then under the linear transformation $z_1 = \alpha z + \beta$, $z_1' = \alpha' z' + \beta'$, where α, β, α', and β' are constants and α and α' are nonzero, the new transformed equations [(2.1)–(2.5)] take the same form and may be satisfied

 0022-2488/90/112597-06$03.00

for suitable functions f_1, f_1', ψ_1, and ϕ_1. For example, (2.3) becomes

$$iq_1 + r_1 + f_1(z_1, \bar{z}_1) = 0,$$

where $q_1 = \partial z_1/\partial y$, $r_1 = \partial^2 z_1/\partial x^2$, and where the function f_1 is defined by $f_1(z_1, \bar{z}_1) = \alpha f((z_1 - \beta)/\alpha, (\bar{z}_1 - \bar{\beta})/\bar{\alpha})$. This enables simplifications to be made in the subsequent analysis without loss of generality.

III. CONDITIONS ON THE BÄCKLUND TRANSFORMATIONS

The two identities implied by (2.5) may be represented by

$$E_1 \equiv 0, \quad E_2 \equiv 0, \tag{3.1}$$

where

$$E_1 = \frac{\partial p}{\partial x} - r, \tag{3.2}$$

$$E_2 = \frac{\partial p}{\partial y} - \frac{\partial q}{\partial x}. \tag{3.3}$$

Using the forms of p, q, r, and r' given by (2.1)–(2.4), we can successively calculate that

$$E_{2l'} = \psi_{q'}, \quad E_{2\bar{l}'} = \psi_{\bar{q}'}, \tag{3.4}$$

$$E_{2s'} = \psi_{p'} - \phi_{q'}, \quad E_{2\bar{s}'} = \psi_{\bar{p}'} - \phi_{\bar{q}'}, \tag{3.5}$$

$$E_{2q'q'} = 2i\psi_{p'p'}, \quad E_{2\bar{q}'\bar{q}'} = -2i\psi_{\bar{p}'\bar{p}'}. \tag{3.6}$$

All these expressions must also be identically zero. Equation (3.4) shows that ψ is independent of q' and $\bar{q}'$ while (3.6) gives that ψ is linear in p' and $\bar{p}'$. Together with (3.5) these deductions show that p and q are of the form

$$p = kp'\bar{p}' + lp' + m\bar{p}' + n, \tag{3.7}$$

$$q = kq'\bar{q}' + lq' + k\bar{q}'p' + m\bar{q}' + H, \tag{3.8}$$

where k, l, m, and n are functions of $z, z', \bar{z}$, and $\bar{z}'$ and H is a function of $z, z', \bar{z}, \bar{z}', p'$, and $\bar{p}'$.

Calculation of $E_{1\bar{q}'}$ [$= 2i(kp' + m)$] immediately gives $k = m = 0$. It now follows that

$$E_{1\bar{p}'} - E_{2\bar{q}'} = 2iH_{\bar{p}'}, \quad E_{2q'p'} = iH_{p'p'},$$

so that

$$H = \tau p' + \chi, \tag{3.9}$$

where τ and χ are functions of $z, z', \bar{z}$, and $\bar{z}'$.

Now E_2 is a polynomial in $p', \bar{p}', q'$, and $\bar{q}'$ and E_1 is a polynomial in p' and $\bar{p}'$. All coefficients must vanish and this leads directly to the 11 conditions:

$$\bar{l}l_{\bar{z}} + l_{\bar{z}'} = 0, \tag{3.10}$$

$$ll_z + l_{z'} = 0, \tag{3.11}$$

$$nl_z + \bar{n}l_{\bar{z}} = 0, \tag{3.12}$$

$$n_{\bar{z}}\bar{l} + n_{\bar{z}'} = 0, \tag{3.13}$$

$$n_z l + n_{z'} = -i\tau, \tag{3.14}$$

$$-l_z\tau + \tau_z l + \tau_{z'} = 0, \tag{3.15}$$

$$-l_{\bar{z}}\bar{\tau} + \tau_{\bar{z}}\bar{l} + \tau_{\bar{z}'} = 0, \tag{3.16}$$

$$i\chi + nn_z + \bar{n}n_{\bar{z}} + f - lf' = 0, \tag{3.17}$$

$$-n_{\bar{z}}\bar{\tau} + \chi_{\bar{z}}\bar{l} + \chi_{\bar{z}'} = 0, \tag{3.18}$$

$$-l_z\chi - l_{\bar{z}}\bar{\chi} - n_z\tau + \tau_z n + \tau_{\bar{z}}\bar{n} + \chi_z l + \chi_{z'} = 0, \tag{3.19}$$

$$-f'\tau - n_z\chi - n_{\bar{z}}\bar{\chi} + \chi_z n + \chi_{\bar{z}}\bar{n} = 0. \tag{3.20}$$

The BT is now of the form

$$p = lp' + n, \tag{3.21}$$

$$q = lq' + \tau p' + \chi. \tag{3.22}$$

The derivatives, $n_{z'}$ and $n_{\bar{z}'}$, given by (3.14) and (3.13), are compatible providing [use is made here of (3.10), its conjugate and (3.16)]

$$\bar{\tau}l_{\bar{z}} = 0. \tag{3.23}$$

Either $\tau = 0$ or $\tau \neq 0$. If $\tau \neq 0$ then $l_{\bar{z}} = 0$ and it is also clear from (3.14) that $n \neq 0$ so that (3.10)–(3.12) give all derivatives of l zero, showing that l is constant.

Note that in the constraints (3.10)–(3.20), when l is constant, the transformations $z \to lz$, $n \to ln$, $\tau \to l\tau$, $\chi \to l\chi$, $f \to lf$, and $f' \to f'$ makes l "disappear." That is, the same constraints are obtained as would have been obtained by setting $l = 1$. Hence we can take $l = 1$ without loss of generality ($l = 0$ leads nowhere) in the l constant case with the proviso that any result may be generalized by reversing the above transformation.

We may now divide the analysis into case 1: $\tau = 0, l = 1$; case 2: $\tau = 0$, l not a constant; and case 3: $\tau \neq 0, l = 1$.

IV. CASE 1: $\tau = 0, l = 1$

If the variables ξ and η are introduced,

$$\xi = z + z', \quad \eta = z - z', \tag{4.1}$$

Eqs. (3.13) and (3.14) show that n is independent of ξ and $\bar{\xi}$, that is, $n = n(\eta, \bar{\eta})$. Also, (3.18) and (3.19) give $\chi = \chi(\eta, \bar{\eta})$. Equation (3.17) now implies that $f(z, \bar{z}) - f'(z', \bar{z}')$ is independent of both ξ and $\bar{\xi}$. Taking partial derivatives with respect to ξ and $\bar{\xi}$ gives, respectively,

$$f_z - f'_{z'} = 0, \quad f_{\bar{z}} - f'_{\bar{z}'} = 0.$$

It follows, in particular, that $f(z, \bar{z})$ is linear in both z and $\bar{z}$ so that it is not possible for a nonlinear equation (2.3) to arise in this case.

V. CASE 2: $\tau = 0$, l NOT CONSTANT

From (3.17),

$$\chi = i(f - lf' + nn_z + \bar{n}n_{\bar{z}}), \tag{5.1}$$

and substituting for χ into (3.18) gives

$$\bar{l}f_{\bar{z}} - lf'_{\bar{z}'} = 0. \tag{5.2}$$

Also, (3.19) may be expressed (simplifying χ_z and $\chi_{z'}$) as

$$l_z\chi + l_{\bar{z}}\bar{\chi} = il(f_z - f'_{z'}). \tag{5.3}$$

Now define the operator T:

$$T \equiv l\frac{\partial}{\partial z} + \frac{\partial}{\partial z'}. \tag{5.4}$$

Since $Tl = T\bar{l} = 0$ from (3.11) and the conjugate of (3.10), operating on (5.2) with T gives

$$\bar{l}g = g', \tag{5.5}$$

where $g(z, \bar{z}) = f_{z\bar{z}}$ and $g'(z', \bar{z}') = f'_{z'\bar{z}'}$. Further, applying T and its conjugate $\bar{T} = \bar{l}(\partial/\partial\bar{z}) + \partial/\partial\bar{z}'$ to (5.5) gives, re-

spectively,

$$\bar{l}lg_z = g'_{z'}, \tag{5.6}$$

$$\bar{l}^2 g_{\bar{z}} = g'_{\bar{z}'}. \tag{5.7}$$

Since $f_{\bar{z}}$ and $f'_{\bar{z}'}$ do or do not vanish together [see (5.2)] and the same is true for $f_{z\bar{z}}$ and $f'_{z'\bar{z}'}$ [see (5.5)], the analysis may be divided into the three subcases: (a) $f_{z\bar{z}} \neq 0, f'_{z'\bar{z}'} \neq 0$; (b) $f_{z\bar{z}} = 0, f'_{z'\bar{z}'} = 0, f_{\bar{z}} \neq 0, f'_{\bar{z}'} \neq 0$; and (c) $f_{\bar{z}} = f'_{\bar{z}'} = 0$.

In what follows it will be seen that only the first and third of these cases lead to nonlinear forms of $f(z,\bar{z})$ and $f'(z',\bar{z}')$.

(a) $f_{z\bar{z}} \neq 0, f'_{z'\bar{z}'} \neq 0$. In this case $g \neq 0$ and $g' \neq 0$, and if $\bar{l}$ is eliminated between (5.5) and (5.7) and the variables separated, it is found that

$$g = 1/(\overline{\alpha z} + h(z)), \quad g' = 1/(\overline{\alpha z'} + h'(z')), \tag{5.8}$$

where α is constant and h,h' are arbitrary functions. Equation (5.6), using (5.5) and (5.8), may now be separated to give

$$h_z(\alpha z + \bar{h})/(\overline{\alpha z} + h) = h'_{z'}(\alpha z' + \bar{h}')/(\overline{\alpha z'} + h') = \text{const.} \tag{5.9}$$

It is now straightforward to deduce that h is linear in z and that h' is linear in z', so that $1/g$ is linear in z and $\bar{z}$ and $1/g'$ is linear in z' and $\bar{z}'$. Neither g nor g' can be constant unless both are constants [see (5.6)] and then (5.5) is contradicted since l is not constant. The existence of a BT and the form of the differential equations [(2.3) and (2.4)] are unaffected by linear transformations of z and z' (see Sec. II). Thus without loss of generality we may consider separately $g = 1/\bar{z}$ and $g = 1/(z + \mu\bar{z} + \nu)$, μ and ν being constant. However, the latter means that $f_{\bar{z}}$ will involve the "ln" function and remembering that $\bar{l} = g'/g$ it is clear that (5.2) cannot be satisfied. Hence $g = 1/\bar{z}$ and f must take the form

$$f(z,\bar{z}) = z \ln \bar{z} + j(\bar{z}) + k(z). \tag{5.10}$$

A similar argument gives

$$f'(z',\bar{z}') = z' \ln \bar{z}' + j'(\bar{z}') + k'(z') \tag{5.11}$$

and hence

$$l = z/z'. \tag{5.12}$$

Equation (5.2) now gives that j, j' are constants, which may be absorbed in k and k' (i.e., taken as zero). Also, from (3.12),

$$n = 0. \tag{5.13}$$

Finally, (3.19) gives

$$k(z) = \alpha z \ln z + \beta z, \quad k'(z') = \alpha z' \ln z' + \beta' z', \tag{5.14}$$

where β and β' are complex constants, and from (3.17),

$$\chi = i[z \ln(\bar{z}/\bar{z}') + \alpha z \ln(z/z') + z(\beta - \beta')]. \tag{5.15}$$

Equations (3.10)–(3.20) are now all satisfied. For this case, therefore, the partial differential equation (2.3) becomes

$$iz_y + z_{xx} + z \ln \bar{z} + \alpha z \ln z + \beta z = 0, \tag{5.16}$$

which transforms under the BT

$$z_x = (z/z')z'_x,$$

$$z_y = \left(\frac{z}{z'}\right)z'_y + i\left[z \ln\left(\frac{\bar{z}}{\bar{z}'}\right) + \alpha z \ln\left(\frac{z}{z'}\right) + z(\beta - \beta')\right] \tag{5.17}$$

to

$$iz'_y + z'_{xx} + z' \ln \bar{z}' + \alpha z' \ln z' + \beta' z' = 0. \tag{5.18}$$

Equation (5.17) can, in fact, be integrated to give

$$z = z'F(y), \tag{5.19}$$

where $F(y)$ satisfies the first-order ordinary differential equation

$$\frac{dF}{dy} = i(F \ln \bar{F} + \alpha F \ln F + (\beta - \beta')F). \tag{5.20}$$

(b) $f_{z\bar{z}} = f'_{z'\bar{z}'} = 0, f_{\bar{z}} \neq 0, f'_{\bar{z}'} \neq 0$. Here f and f' now take the forms

$$f = j(z) + \overline{k(z)}, \tag{5.21}$$

$$f' = j'(z') + \overline{k'(z')}, \tag{5.22}$$

and the conjugate of (5.2) becomes

$$lk_z = \bar{l}k'_{\bar{z}'}. \tag{5.23}$$

Eliminating l it follows that when $f_{\bar{z}} \neq 0$ and $f'_{\bar{z}'} \neq 0$, k_z, $k'_{\bar{z}'}$, and $\bar{l}/l$ are all constant. Equation (5.3) and its conjugate now show that j_z and $j'_{\bar{z}'}$ are also constant so that this case leads only to linear forms of f and f'.

(c) $f_{\bar{z}} = f'_{\bar{z}'} = 0$. In addition to $Tl = T\bar{l} = 0$, Eq. (3.14) and the conjugate of (3.13) give $Tn = T\bar{n} = 0$. It then follows that $Tl_z = (Tl)_z - l_z^2 = -l_z^2$ and similarly $Tl_{\bar{z}} = -l_z l_{\bar{z}}$, $Tn_z = -n_z l_z$, $Tn_{\bar{z}} = -n_z l_{\bar{z}}$, $T\bar{n}_z = -l_z \bar{n}_z$, and $T\bar{n}_{\bar{z}} = -l_{\bar{z}}\bar{n}_z$. Equation (3.17) now gives, using (3.12), $T\chi = il(f_z - f'_{z'})$ and $T\overline{\chi} = 0$. Applying T to (5.3) and making use of these results shows that

$$lf_{zz} - f'_{z'z'} = 0. \tag{5.24}$$

Note that $f_{zz} = 0$ can be exluded since (5.24), again, gives both f and f' linear, and substituting for l into (3.11) gives

$$f_{zzz}f'^2_{z'z'} = f'_{z'z'z}f^2_{zz}.$$

Separating the variables, solving for $f(z)$ and $f'(z')$, and making linear transformations to z and z', without loss of generality (see Sec. II), we find that $f(z) = \alpha z \ln z + \beta z$ and $f'(z') = \alpha' z' \ln z' + \beta' z'$, so that (5.24) gives $l = z/z'$. This is essentially a subset of the case 2(a) since in (5.16) it is possible to make scale changes in x and y to introduce a parameter coefficient of the $z \ln \bar{z}$ term, without otherwise changing the form of the equation. Setting this parameter to be zero gives the present case. The form of $F(y)$ in (5.20) may, however, now be obtained explicitly. Summing up,

$$iz_y + z_{xx} + \alpha z \ln z + \beta z = 0 \tag{5.25}$$

transforms to

$$iz'_y + z'_{xx} + \alpha z' \ln z' + \beta' z' = 0 \tag{5.26}$$

under the transformation (cf. Ref. 8)

$$z = z' \exp[\gamma e^{i\alpha y} - (\beta - \beta')/\alpha], \tag{5.27}$$

with γ constant.

VI. CASE 3: $\tau \neq 0, l = 1$

It is now convenient to introduce new independent variables ξ, η, such that

$$z = (\xi + \eta)/2, \quad z' = (\xi - \eta)/2. \tag{6.1}$$

Equations (3.10)–(3.12) are now satisfied. Equations (3.15) and (3.16) reduce to $\tau_\xi = \tau_{\bar{\xi}} = 0$, so that

$$\tau = \tau(\eta,\bar{\eta}). \tag{6.2}$$

Equations (3.13) and (3.14) reduce to

$$n_{\bar{\xi}} = 0, \quad n_{\xi} = -i\tau/2,$$

so that n is of the form

$$n = -i\tau\xi/2 + im(\eta,\bar{\eta}). \tag{6.3}$$

Equations (3.18) and (3.19) give expressions for $\chi_{\bar{\xi}}$ and χ_{ξ}, respectively, which may be integrated to give

$$\chi = \tfrac{1}{4}i[\,-\bar{\tau}\tau_{\bar{\eta}}\xi\bar{\xi} + 2\bar{\tau}m_{\bar{\eta}}\bar{\xi} + (\,-\tau^2 + 2\tau m_{\eta} - 2m\tau_{\eta} + 2\bar{m}\tau_{\bar{\eta}})\xi - 4A(\eta,\bar{\eta})\,], \tag{6.4}$$

$A(\eta,\bar{\eta})$ being an arbitrary function of η and $\bar{\eta}$.

It only remains now to satisfy Eqs. (3.17) and (3.20). That is, suitable functions $m(\eta,\bar{\eta})$, $A(\eta,\bar{\eta})$, $\tau(\eta,\bar{\eta})$, $f(z,\bar{z})$, and $f'(z',\bar{z}')$ must be found so that these last two equations become identities in ξ, $\bar{\xi}$, η, and $\bar{\eta}$.

The first of these two equations (3.17) gives

$$f(z,\bar{z}) - f'(z',\bar{z}') = \xi^2 P_1 + \xi\bar{\xi}P_2 + \xi P_3 + \bar{\xi}P_4 + P_5 = F(\xi,\bar{\xi},\eta,\bar{\eta}), \tag{6.5}$$

where

$$P_1 = \tfrac{1}{4}\tau\tau_{\eta}, \quad P_2 = -\tfrac{1}{2}\bar{\tau}\tau_{\bar{\eta}}, \quad P_3 = \bar{m}\tau_{\bar{\eta}} - m\tau_{\eta},$$
$$P_4 = \bar{\tau}m_{\bar{\eta}}, \quad P_5 = G(\eta,\bar{\eta}), \tag{6.6}$$

and where

$$G(\eta,\bar{\eta}) = mm_{\eta} - \bar{m}m_{\bar{\eta}} - m\tau/2 - A(\eta,\bar{\eta}). \tag{6.7}$$

The nature of the left-hand side of (6.5) imposes conditions on P_1, P_2, P_3, P_4, and P_5. Specifically, $F_{zz'} = F_{z\bar{z}'} = F_{\bar{z}z'} = F_{\bar{z}\bar{z}'} = 0$ or

$$F_{\xi\xi} - F_{\eta\eta} = F_{\bar{\xi}\bar{\xi}} - F_{\bar{\eta}\bar{\eta}} = F_{\eta\bar{\xi}} - F_{\bar{\eta}\xi} = F_{\xi\bar{\xi}} - F_{\eta\bar{\eta}} = 0. \tag{6.8}$$

Hence

$$P_1 = \tfrac{1}{4}\tau\tau_{\eta} = \alpha_1\eta + \beta_1\bar{\eta} + \gamma_1, \tag{6.9}$$

$$P_2 = -\tfrac{1}{2}\bar{\tau}\tau_{\bar{\eta}} = 2\beta_1\eta + \gamma_2, \tag{6.10}$$

$$P_3 = \bar{m}\tau_{\bar{\eta}} - m\tau_{\eta} = \alpha_3\eta + \beta_3\bar{\eta} + \gamma_3, \tag{6.11}$$

$$P_4 = \bar{\tau}m_{\bar{\eta}} = \beta_3\eta + \beta_4\bar{\eta} + \gamma_4, \tag{6.12}$$

$$P_5 = G = \bar{\eta}(\beta_1\eta^2 + \gamma_2\eta + \epsilon_1) + \tfrac{1}{3}\alpha_1\eta^3 + \gamma_1\eta^2 + \delta_2\eta + \epsilon_2, \tag{6.13}$$

where $\alpha_1, \alpha_3, \beta_1, \beta_3, \beta_4, \gamma_1, \gamma_2, \gamma_3, \gamma_4, \delta_2, \epsilon_1$, and ϵ_2 are constants. Note that once the functional form of $F(\xi,\bar{\xi},\eta,\bar{\eta})$ has been established, f and f' may be obtained directly from F, in terms of ξ, $\bar{\xi}$, η, and $\bar{\eta}$, by the substitutions

$$f = F\left(\frac{(\xi+\eta)}{2}, \frac{(\bar{\xi}+\bar{\eta})}{2}, \frac{(\xi+\eta)}{2}, \frac{\bar{\xi}+\bar{\eta}}{2}\right) + \epsilon_3 - \epsilon_2, \tag{6.14}$$

$$f' = -F\left(\frac{(\xi-\eta)}{2}, \frac{(\bar{\xi}-\bar{\eta})}{2}, \frac{(\eta-\xi)}{2}, \frac{(\bar{\eta}-\bar{\xi})}{2}\right) + \epsilon_3, \tag{6.15}$$

where $\epsilon_3 (= f(0,0))$ is an arbitrary constant and where ϵ_2 in (6.14) is the value of $F(0,0,0,0)$, obtained from (6.5) and (6.13).

The remaining equation (3.20) may be written in terms of ξ and η as

$$E_3 = 0, \tag{6.16}$$

where

$$E_3 = -\tau f' + n(\chi_{\xi} + \chi_{\eta}) + \bar{n}(\chi_{\bar{\xi}} + \chi_{\bar{\eta}}) - \chi(n_{\xi} + n_{\eta}) - \bar{\chi}(n_{\bar{\xi}} + n_{\bar{\eta}}) \tag{6.17}$$

and where τ, n, χ, and f' are given by (6.2), (6.3), (6.4), and (6.15), respectively. Direct calculation gives

$$E_{3\xi\xi\xi} = -\tau\alpha_1 \tag{6.18}$$

and

$$E_{3\xi\bar{\xi}\bar{\xi}} = -\beta_4(\tau + \bar{\tau})/2, \tag{6.19}$$

so that

$$\alpha_1 = 0 \tag{6.20}$$

and

$$\beta_4 = 0 \quad \text{or} \quad \bar{\tau} = -\tau. \tag{6.21}$$

With $\alpha_1 = 0$, (6.9) integrates to give

$$\tau^2 = 8(\beta_1\eta\bar{\eta} + \gamma_1\eta + \overline{p(\eta)}), \tag{6.22}$$

where $p(\eta)$ is an arbitrary function of η. Hence

$$\tau\tau_{\bar{\eta}} = 4(\beta_1\eta + \overline{p_{\eta}}). \tag{6.23}$$

Comparison of (6.23) and (6.10) and using $|\bar{\tau}/\tau| = 1$ shows that p must be linear. Taking $p(\eta) = \delta_1\eta + \delta_3$, (6.10) and (6.23) give

$$\tau(\beta_1\eta + \gamma_2/2) = -\bar{\tau}(\beta_1\eta + \bar{\delta}_1), \tag{6.24}$$

where

$$\tau^2 = 8(\beta_1\eta\bar{\eta} + \gamma_1\eta + \bar{\delta}_1\bar{\eta} + \bar{\delta}_3). \tag{6.25}$$

Squaring (6.24), substituting for τ^2 from (6.25), and equating coefficients of different powers of η and $\bar{\eta}$ gives that β_1 must be real. In addition, conditions arise that may be conveniently separated into five discrete cases. The corresponding forms of $\tau(\neq 0)$ are also given:

(a) $\beta_1 = 0$, $\gamma_2 = 0$, $\delta_1 = 0$, $\gamma_1 = 0$, τ constant;

(b) $\beta_1 = 0$, $\gamma_2 = 0$, $\delta_1 = 0$, $\gamma_1 \neq 0$,
$\tau^2 = 8(\gamma_1\eta + \bar{\delta}_3)$;

(c) $\beta_1 = 0$, $\gamma_1 \neq 0$, $\gamma_2 = 2\bar{\gamma}_1$, $\delta_1 = \gamma_1$,
$\bar{\delta}_3 = \delta_3$, $\bar{\tau} = -\tau$, $\tau^2 = 8(\gamma_1\eta + \bar{\gamma}_1\bar{\eta} + \bar{\delta}_3)$;

(d) $\beta_1 = 0$, $\gamma_1 \neq 0$, $\gamma_2 \neq 2\bar{\gamma}_1$, $\gamma_2\bar{\gamma}_2 = 4\gamma_1\bar{\gamma}_1$,
$\delta_1 = (1/4)\bar{\gamma}_2^2/\gamma_1$,
$$\bar{\delta}_3 = \frac{(1/4)\gamma_2^2\delta_3}{\bar{\gamma}_1^2}, \quad \bar{\tau} = -\frac{2\tau\bar{\gamma}_1}{\gamma_2},$$
$$\tau^2 = 8\left(\gamma_1\eta + \frac{(1/4)\gamma_2^2\bar{\eta}}{\bar{\gamma}_1} + \bar{\delta}_3\right);$$

(e) $\bar{\beta}_1 = \beta_1 \neq 0$, $\gamma_2 = 2\bar{\gamma}_1$, $\delta_1 = \gamma_1$, $\bar{\delta}_3 = \delta_3$,
$\bar{\tau} = -\tau$, $\tau^2 = 8(\beta_1\eta\bar{\eta} + \gamma_1\eta + \bar{\gamma}_1\bar{\eta} + \delta_3)$.

In cases (b) and (d), $\bar{\tau} \neq -\tau$, so that, from (6.21), $\beta_4 = 0$.

For each of these five cases (6.12) may be integrated directly to give an expression for $m(\eta,\bar{\eta})$ that will, of course, involve an arbitrary function of η. Substituting τ and m into (6.11) then leads to further constraints on the constants and

also specifies, in most cases, a more precise form of m. In summary, Eqs. (6.9)–(6.13) imply the following forms of τ, m, and G corresponding to cases (a)–(e) above:

(a) τ constant,

$$m = (\tfrac{1}{2}\beta_4\bar{\eta}^2 + \gamma_4\bar{\eta})/\bar{\tau} + V(\eta), \quad V \text{ arbitrary},$$

$$G = \bar{\eta}\epsilon_1 + \eta\delta_2 + \epsilon_2;$$

(b) $\tau^2 = 8(\gamma_1\eta + \bar{\delta}_3), \quad \gamma_1 \neq 0,$

$$m = -\tfrac{1}{4}\tau(\alpha_3\eta + \gamma_3)/\gamma_1,$$

$$G = \gamma_1\eta^2 + \delta_2\eta + \epsilon_1\bar{\eta} + \epsilon_2;$$

(c) $\tau^2 = 8(\gamma_1\eta + \bar{\gamma}_1\bar{\eta} + \delta_3), \quad \delta_3 \text{ real}, \quad \bar{\tau} = -\tau,$

$$m = -\tau(\gamma_1(\gamma_1\bar{\alpha}_3 + 2\bar{\gamma}_1\alpha_3)\eta + \bar{\gamma}_1(\gamma_1\bar{\alpha}_3 - \bar{\gamma}_1\alpha_3) \times\bar{\eta} - 2\delta_3(\gamma_1\bar{\alpha}_3 - \bar{\gamma}_1\alpha_3) + 3\gamma_1^2\gamma_4)/(12\gamma_1^2\bar{\gamma}_1),$$

$$G = 2\bar{\gamma}_1\eta\bar{\eta} + \gamma_1\eta^2 + \delta_2\eta + \epsilon_1\bar{\eta} + \epsilon_2;$$

(d) $\tau^2 = 8(\gamma_1\eta + \tfrac{1}{4}\gamma_2^2\bar{\eta}/\bar{\gamma}_1 + \bar{\delta}_3), \quad \bar{\tau} = -2\tau\bar{\gamma}_1/\gamma_2,$

$$m = -\tfrac{1}{4}\tau(\alpha_3\eta/\gamma_1 + 2\gamma_4/\gamma_2) + \epsilon_6,$$

$$G = \gamma_2\eta\bar{\eta} + \gamma_1\eta^2 + \delta_2\eta + \epsilon_1\bar{\eta} + \epsilon_2;$$

where

$$\bar{\epsilon}_6 = \frac{\epsilon_6\bar{\gamma}_2}{\gamma_2}, \quad \bar{\alpha}_3 = \frac{\alpha_3\bar{\gamma}_1}{\gamma_1}, \quad \gamma_2\bar{\gamma}_2 = 4\gamma_1\bar{\gamma}_1,$$

$$\bar{\delta}_3 = \frac{(1/4)\delta_3\gamma_2^2}{\bar{\gamma}_1^2},$$

$$\bar{\gamma}_3 = 2\gamma_3\bar{\gamma}_1/\gamma_2, \quad \bar{\gamma}_4 = -4\gamma_1\bar{\gamma}_1(2\gamma_1\gamma_4 - \gamma_2\gamma_3)/\gamma_2^3,$$

$$\gamma_1 \neq 0, \quad \gamma_2 \neq 2\bar{\gamma}_1;$$

(e) $\tau^2 = 8(\beta_1\eta\bar{\eta} + \gamma_1\eta + \bar{\gamma}_1\bar{\eta} + \delta_3), \quad \bar{\tau} = -\tau,$

$$m = -\tfrac{1}{4}\tau\bar{\alpha}_3/\beta_1 + \epsilon_6(\beta_1\eta + \bar{\gamma}_1),$$

$$G = \beta_1\eta^2\bar{\eta} + \gamma_1\eta^2 + 2\bar{\gamma}_1\eta\bar{\eta} + \delta_2\eta + \epsilon_1\bar{\eta} + \epsilon_2;$$

where β_1, δ_3, and ϵ_6 are all real, $\beta_1 \neq 0$.

In each case Eq. (6.5) enables F to be found in a form that is consistent with its representation $F = f(z,\bar{z}) - f'(z',\bar{z}')$. Here f and f' may be calculated using the substitutions given in (6.14) and (6.15). Of the original equations, (3.10)–(3.20), only (3.20), which was expressed earlier [see (6.16) and (6.17)] as $E_3 = 0$, remains to be satisfied.

Using the above results E_3 may be expressed in terms of ξ, $\bar{\xi}$, η, and $\bar{\eta}$ and must be identically zero. Of the five cases above, three [(a), (b), and (d)] lead to no nonlinear forms of $f(z,\bar{z})$. For case (b) direct calculation leads to

$$(E_3/\tau)_{\xi\bar{\xi}} = -\alpha_3,$$

giving $\alpha_3 = 0$, and then to

$$(\tau E_3)_{\xi\eta\eta} = -4\gamma_1^2,$$

leading to the contradiction $\gamma_1 = 0$.

For case (d),

$$(\tau E_3)_{\xi\bar{\xi}\eta} = -8\gamma_1\alpha_3,$$

so that, since $\gamma_1 \neq 0$, $\alpha_3 = 0$. Hence

$$(\tau E_3)_{\xi\eta\eta} = -4\bar{\gamma}_1^2,$$

giving, again, the contradiction $\gamma_1 = 0$.

In case (a),

$$E_{3\xi\bar{\xi}} = \tfrac{1}{2}\tau^2 V_{\eta\eta},$$

so that $V(\eta)$ is linear. Then

$$E_{3\eta\eta\eta\bar{\eta}} = 6\beta_4\bar{\beta}_4^2/(\tau^2\bar{\tau}),$$

giving $\beta_4 = 0$ and hence

$$F = \gamma_4\bar{\xi} + \delta_2\eta + \epsilon_1\bar{\eta} + \epsilon_2,$$

which leads only to linear forms for both $f(z,\bar{z})$ and $f'(z',\bar{z}')$.

Case (e) leads to the cubic Schrödinger equation. The identities $(\tau E_3)_{\xi\eta\bar{\eta}} = 0$, $(\tau E_3)_{\xi\eta\eta} = 0$, $(\tau E_3)_{\xi\eta} = 0$, and $(\tau E_3)_{\eta\bar{\eta}} = 0$ give in turn

$$\bar{\delta}_2 = \delta_2,$$

$$\epsilon_1 = \tfrac{1}{4}(\bar{\alpha}_3^2 + 4\bar{\gamma}_1^2)/\beta_1,$$

$$\epsilon_2 = \tfrac{1}{4}(4\beta_1\bar{\gamma}_1\delta_2 - 4\gamma_1\bar{\gamma}_1^2 + \gamma_1\bar{\alpha}_3^2)/\beta_1^2,$$

$$\epsilon_3 = \tfrac{1}{16}(8\beta_1\bar{\gamma}_1\delta_2 + 4\beta_1\bar{\alpha}_3\delta_2 - 8\gamma_1\bar{\gamma}_1^2 + 2\gamma_1\bar{\alpha}_3^2 + 4\bar{\gamma}_1^2\alpha_3 - \bar{\alpha}_3^2\alpha_3)/\beta_1^2.$$

E_3 is now identically zero.

The forms of $f(z,\bar{z})$ and $f'(z',\bar{z}')$ may be calculated from (6.14) and (6.15), and these may be simplified by replacing z by $z - \tfrac{1}{4}(2\bar{\gamma}_1 + \bar{\alpha}_3)/\beta_1$ and z' by $z' - \tfrac{1}{4}(-2\bar{\gamma}_1 + \bar{\alpha}_3)/\beta_1$. In addition, the constants involved may be simplified by setting

$$\delta_2 = \mu + \tfrac{1}{2}(4\gamma_1\bar{\gamma}_1 + \alpha_3\bar{\alpha}_3)/\beta_1,$$

$$\delta_3 = -\tfrac{1}{2}\beta_1\lambda_1 + \gamma_1\bar{\gamma}_1/\beta_1,$$

$$\epsilon_6 = \lambda_2/\beta_1,$$

where μ, λ_1, and λ_2 are all real constants. With these substitutions,

$$f(z,\bar{z}) = 4\beta_1 z^2\bar{z} + \mu z,$$

$$f'(z',\bar{z}') = 4\beta_1 z'^2\bar{z}' + \mu z'$$

and auto BT's for the equation

$$iz_y + z_{xx} + 4\beta_1 z^2\bar{z} + \mu z = 0 \tag{6.26}$$

are given by

$$p = p' + n, \quad q = q' + \tau p' + \chi, \tag{6.27}$$

where

$$\begin{aligned} \tau^2 &= 4\beta_1(2(z - z')(\bar{z} - \bar{z}') - \lambda_1), \\ n &= -\tfrac{1}{2}i\tau(z + z') + i\lambda_2(z - z'), \\ \chi &= i\lambda_2\tau z + i\mu(z - z') + i\lambda_1\beta_1(z + z') - i\lambda_2^2(z - z') \\ &\quad + 2i\beta_1(z - z')(z\bar{z}' - z'\bar{z} + 2z'\bar{z}'). \end{aligned} \tag{6.28}$$

This agrees with Lamb.[4]

The only remaining case is case (c). Here, the identity $(\tau E_3)_{\xi\eta\eta} = 0$ gives

$$\alpha_3 = \pm 2i\gamma_1/\sqrt{3}.$$

Considering $\alpha_3 = +2i\gamma_1/\sqrt{3}$, the identities $(\tau E_3)_{\xi\eta} = 0$, $(\tau E_3)_{\xi} = 0$, $(\tau E_3)_{\eta\eta} = 0$, $(\tau E_3)_{\eta\bar{\eta}} = 0$, and $(\tau E_3)_{\eta} = 0$ give in turn

$$\epsilon_1 = [\sqrt{3}\gamma_1\bar{\gamma}_1\bar{\delta}_2 - i(2\gamma_1^2\gamma_4 + \bar{\gamma}_1^2\bar{\gamma}_4)]/(\sqrt{3}\gamma_1^2),$$

$$\epsilon_2 = \tfrac{1}{4}(\gamma_1^4\gamma_4^2 - r_1)/(\gamma_1^3\bar{\gamma}_1^2), \quad r_1 \text{ real},$$

$$\bar{\delta}_2 = \delta_2,$$
$$\delta_2 = i(\gamma_1^2\gamma_4 - \bar{\gamma}_1^2\bar{\gamma}_4)/(\sqrt{3}\gamma_1\bar{\gamma}_1),$$
$$\epsilon_3 = \tfrac{1}{8}[r_1(-\sqrt{3}+i) + \gamma_1^4\gamma_4^2(\sqrt{3}+i) - 2i\bar{\gamma}_1^4\bar{\gamma}_4^2]/(\sqrt{3}\gamma_1^3\bar{\gamma}_1^2).$$

Now E_3 is identically zero.

Here $f(z,\bar{z})$, $f'(z',\bar{z}')$, n, and χ may now be calculated from (6.14) and (6.15). Their forms are considerably simplified by first writing

$$\gamma_4 = \bar{\gamma}_1\gamma_5(\sqrt{3}-i)/\gamma_1,$$
$$r_1 = -\gamma_1^2\bar{\gamma}_1^2(8\gamma_5\bar{\gamma}_5 + 3r_2), \quad r_2 \text{ real},$$
$$\delta_3 = \tfrac{1}{16}[2\sqrt{3}(\gamma_5+\bar{\gamma}_5) + 6i(\gamma_5-\bar{\gamma}_5) - r_5], \quad r_5 \text{ real},$$

i.e., replacing constants γ_4, r_1, and δ_3 by constants γ_5, r_2, and r_5 and then applying linear transformations to z and z' as follows:

$$z \to \tfrac{1}{8}(\sqrt{3}-i)(\sqrt{3}z - (\gamma_5+\bar{\gamma}_5))/\gamma_1,$$
$$z' \to -\frac{1}{8}[z'(3+i\sqrt{3}) + \gamma_5(\sqrt{3}-i) + 2\bar{\gamma}_5 i]/\gamma_1.$$

The two partial differential equations then simplify to

$$iz_y + z_{xx} + (z+\bar{z})^2 + r_2 = 0,$$
$$iz'_y + z'_{xx} + (z'+\bar{z}')^2 + r_2 = 0 \tag{6.29}$$

and the BT's relating them are

$$p = -\tfrac{1}{2}(1+i\sqrt{3})p' + n,$$
$$q = -\tfrac{1}{2}(1+i\sqrt{3})q' + \tau p' + \chi, \tag{6.30}$$

where

$$\tau^2 = \tfrac{1}{4}[8\sqrt{3}i(z+z') - (12-i4\sqrt{3})(\bar{z}+z') + r_5(1-i\sqrt{3})],$$
$$n = (\tau/36)[z(2\sqrt{3}-18i) - \bar{z}(2\sqrt{3}+6i) + z'(-8\sqrt{3}+12i) - \bar{z}'4\sqrt{3} - r_5(\sqrt{3}+i)],$$
$$\chi = [-12z^2(1+i3\sqrt{3}) + 24z\bar{z}(1+i\sqrt{3}) - 12\bar{z}^2(1-i\sqrt{3}) - 12z'^2(13-i\sqrt{3}) - 96z'\bar{z}'(2-i\sqrt{3}) - 12\bar{z}'^2(7-i3\sqrt{3}) - 12zz'(1-i5\sqrt{3}) - 12z\bar{z}'(5+i3\sqrt{3}) - 12\bar{z}z'(7+i\sqrt{3}) - 12\bar{z}z'(11+i\sqrt{3}) + r_5(-4z(6-i\sqrt{3}) - 4i\sqrt{3}\bar{z} + 2z'(9-i5\sqrt{3}) - 2\bar{z}'(3+i\sqrt{3})) + r_5^2(1+i\sqrt{3}) - 108r_2(1-i\sqrt{3})]/(72\sqrt{3}), \tag{6.31}$$

where r_5 is the Bäcklund parameter.

For the case $\alpha_3 = -2i\gamma_1/\sqrt{3}$ the above may be followed through in exactly the same way, arriving at the same partial differential equations. For the BT's, however, i should be replaced by $-i$ where it occurs explicitly in the expressions for p, q, τ, n, and χ above. We therefore have two different auto-BT's, each with an arbitrary parameter for the partial differential equation (6.29).

These BT's incorporate auto-BT's for the Boussinesq equation since if $z = u + iv$, Eq. (6.29) becomes

$$iu_y - v_y + u_{xx} + iv_{xx} + 4u^2 + r_2 = 0.$$

If this is separated into real and imaginary parts and v eliminated, u satisfies

$$u_{yy} + u_{xxxx} + (4u^2)_{xx} = 0,$$

which is a form of the well-known Boussinesq equation.

[1] D. W. McLaughlin and A. C. Scott, J. Math. Phys. **14**, 1817 (1973).
[2] J. J. C. Nimmo and D. G. Crighton, Phys. Lett. A **82**, 211 (1981).
[3] J. J. C. Nimmo and D. G. Crighton, Proc. R. Soc. London Ser. A **384**, 381 (1982).
[4] G. L. Lamb, J. Math. Phys. **15**, 2157 (1974).
[5] M. J. Clairin, Ann. Toulouse 2^e Sér. 5, 437 (1903).
[6] I. Bialynicki-Birula and J. Mycielski, Bull. Acad. Polon. Sci. **23**, 461 (1975).
[7] R. Brito, J. A. Cuesta, and A. F. Ranada, Phys. Lett. A **128**, 360 (1988).
[8] H. Steudel, Physica D **1**, 420 (1980).
[9] K. Fukushima, M. Wadati, and Y. Narahara, J. Phys. Soc. Jpn. **49**, 1593 (1980).
[10] J. G. Kingston and C. Rogers, Phys. Lett. A **92**, 261 (1982).

An alternate characterization of integrability

Ashok Das and Wen-Jui Huang
Department of Physics and Astronomy, University of Rochester, Rochester, New York 14627

(Received 16 April 1990; accepted for publication 20 June 1990)

This paper will show that the existence of at least three independent symplectic forms (related in a simple way) on the phase space of a dynamical system is a sufficient condition for the integrability of the system.

I. INTRODUCTION

The study of integrable models has been of much interest recently.[1] These are nonlinear Hamiltonian systems that can be solved exactly and have been studied both from the group theory point of view as well as from the geometrical point of view. Most of the integrable models possess at least two Poisson bracket structures.[2,3] In such a case, one can construct an appropriate Nijenhuis tensor[4] on the phase space and it is known from the studies in the geometric approach that the vanishing of the Nijenhuis tensor is a sufficient condition for the integrability of the system.[5-7]

In this paper we give an alternate but equivalent characterization of the integrability condition which may be more useful particularly in the study of continuum models. In particular, we show that the existence of at least three distinct symplectic forms (related in a simple manner) on the phase space of a dynamical system is a sufficient condition for integrability.

II. THEORY

We begin by deriving some of the consequences of the vanishing Nijenhuis tensor on a manifold. Let M be a finite-dimensional differentiable manifold and $S\in T^1_1 M$. In local coordinates S can be written as

$$S = S^{\nu}_{\mu}(x)\frac{\partial}{\partial x^{\nu}}\otimes dx^{\mu}, \quad \mu,\nu = 1,2,\ldots,\dim M. \tag{1}$$

Clearly, therefore, S maps a tangent space to itself. We can now define a second tensor field N_S known as the Nijenhuis tensor field associated with S which, for any vector field X on the manifold, satisfies

$$L_{X\cdot S}S - (L_X S)\cdot S = X\cdot N_S. \tag{2}$$

Here, L_X is the Lie derivative with respect to the vector field X and the $\cdot$ stands for the contraction of a contravariant index with the first covariant index. Thus in local coordinates

$$\begin{aligned} X\cdot S &= X^{\mu}S^{\nu}_{\mu}\frac{\partial}{\partial x^{\nu}}, \\ S\cdot S &= S^{\lambda}_{\mu}S^{\nu}_{\lambda}\frac{\partial}{\partial x^{\nu}}\otimes dx^{\mu}, \\ S^n &= (S^{n-1})\cdot S = S^{\lambda_1}_{\mu}S^{\lambda_2}_{\lambda_1}\cdots S^{\nu}_{\lambda_{n-1}}\frac{\partial}{\partial x^{\nu}}\otimes dx^{\mu}. \end{aligned} \tag{3}$$

It is clear that $N_S\in T^1_2 M$ and in local coordinates, it takes the form

$$N_S = N^{\mu}_{\alpha\beta}(x)\frac{\partial}{\partial x^{\mu}}\otimes dx^{\alpha}\wedge dx^{\beta}, \tag{4}$$

where the Nijenhuis tensor $N^{\mu}_{\alpha\beta}(x)$ has the explicit form

$$\begin{aligned} N^{\mu}_{\alpha\beta} &= -N^{\mu}_{\beta\alpha} \\ &= S^{\lambda}_{\alpha}\partial_{\lambda}S^{\mu}_{\beta} - S^{\lambda}_{\beta}\partial_{\lambda}S^{\mu}_{\alpha} - S^{\mu}_{\lambda}(\partial_{\alpha}S^{\lambda}_{\beta} - \partial_{\beta}S^{\lambda}_{\alpha}). \end{aligned} \tag{5}$$

Note that when the Nijenhuis tensor vanishes, we have from Eq. (2)

$$L_{X\cdot S}S = (L_X S)\cdot S, \tag{6}$$

for any vector field X on the manifold. It is now straightforward to show[8,9] that if $N_S = 0$,

$$L_{X\cdot S}S^n = (L_X S^n)\cdot S, \tag{7}$$

$$L_{X\cdot S^n}S = (L_X S)\cdot S^n, \tag{8}$$

$$L_{X\cdot S^m}S^n = (L_X S^n)\cdot S^m. \tag{9}$$

In particular, from Eq. (9) we see that

$$L_{X\cdot S^n}S^n = (L_X S^n)\cdot S^n. \tag{10}$$

Combining this with the definition of the Nijenhuis tensor in Eq. (2), we conclude that if $N_S = 0$, then

$$N_{S^n} = 0. \tag{11}$$

Let us next assume that the manifold M is a special symplectic manifold with two distinct symplectic forms:

$$\begin{aligned} f &= f_{\mu\nu}(x)dx^{\mu}\wedge dx^{\nu}, \\ F &= F_{\mu\nu}(x)dx^{\mu}\wedge dx^{\nu}. \end{aligned} \tag{12}$$

By definition, these two-forms are nondegenerate and closed. Thus

$$df = 0 = dF. \tag{13}$$

From the definition of the Lie derivative

$$L_X = d\cdot i_X + i_X\cdot d, \tag{14}$$

we then see that

$$\begin{aligned} L_X f &= d(i_X f), \\ L_X F &= d(i_X F). \end{aligned} \tag{15}$$

Note that because the two-forms are nondegenerate, we can define the Poisson bracket structures $f^{\mu\nu}(x)$ and $F^{\mu\nu}(x)$ which are second rank antisymmetric tensors on the manifold and satisfy

$$\begin{aligned} f^{\mu\lambda}f_{\lambda\nu} &= f_{\nu\lambda}f^{\lambda\mu} = \delta^{\mu}_{\nu}, \\ F^{\mu\lambda}F_{\lambda\nu} &= F_{\nu\lambda}F^{\lambda\mu} = \delta^{\mu}_{\nu}. \end{aligned} \tag{16}$$

This, in turn, allows us to define naturally a tensor field $S\in T^1_1 M$ as

 0022-2488/90/112603-03$03.00

$$S = S^{\nu}_{\mu}(x)\frac{\partial}{\partial x^{\nu}} \otimes dx^{\mu}$$

$$= F_{\mu\lambda}(x) f^{\lambda\nu}(x)\frac{\partial}{\partial x^{\nu}} \otimes dx^{\mu}. \tag{17}$$

Equivalently, we can think of the tensor field S as connecting the two distinct symplectic forms, namely,

$$F = S \cdot f. \tag{18}$$

It is, in fact, clear now that with the help of the tensor field S, we can define a series of two-forms as

$$F^{(n)} = S^{n} \cdot f, \quad n \geqslant 0. \tag{19}$$

Proposition: If $N_S = 0$, then the sequence of two-forms defined by

$$F^{(n)} = S^{n} \cdot f, \quad n \geqslant 0$$

are indeed symplectic forms.[10]

Proof: Note that the two-forms $F^{(n)}$ are automatically nondegenerate since both $S^{\nu}_{\mu}(x)$ and $f_{\mu\nu}(x)$ are invertible. Thus, to show that the sequence of two-forms $F^{(n)}$ are symplectic forms, we merely have to prove closure.

We note that

$$F^{(0)} = f,$$

$$F^{(1)} = F, \tag{20}$$

both of which are closed two-forms. Let us next assume that up to some $p < n$, the two-forms are closed. Thus we assume that

$$dF^{(p)} = 0, \quad p < n. \tag{21}$$

Equivalently,

$$L_X F^{(p)} = d(i_X F^{(p)}). \tag{22}$$

Let us next note that

$$i_{X\cdot S}F^{(p)} = (X\cdot S)^{\mu} F^{(p)}_{\mu\nu}\, dx^{\nu} = X^{\lambda} S^{\mu}_{\lambda} F^{(p)}_{\mu\nu}\, dx^{\nu}$$

$$= X \cdot S \cdot F^{(p)} = X \cdot F^{(p+1)} = i_X F^{(p+1)}. \tag{23}$$

Consequently, using Eq. (22) we obtain

$$L_{X\cdot S}F^{(p)} = d(i_{X\cdot S}F^{(p)}) = d(i_X F^{(p+1)}). \tag{24}$$

On the other hand, for $1 \leqslant p < n$,

$$\begin{aligned} L_{X\cdot S}F^{(p)} &= L_{X\cdot S}(S\cdot F^{(p-1)}) \\ &= (L_{X\cdot S}S)\cdot F^{(p-1)} + S\cdot L_{X\cdot S}F^{(p-1)} \\ &= (L_X S)\cdot S\cdot F^{(p-1)} + S\cdot d(i_{X\cdot S}F^{(p-1)}) \\ &= (L_X S)\cdot F^{(p)} + S\cdot d(i_X F^{(p)}) \\ &= (L_X S)\cdot F^{(p)} + S\cdot L_X F^{(p)} \\ &= L_X(S\cdot F^{(p)}) = L_X F^{(p+1)}. \end{aligned} \tag{25}$$

Here, we have used the relation in Eq. (6) which holds for $N_S = 0$. Comparing Eqs. (24) and (25), we conclude now that

$$L_X F^{(p+1)} = d(i_X F^{(p+1)}). \tag{26}$$

From the definition of L_X in Eq. (14), it now follows that

$$dF^{(p+1)} = 0, \tag{27}$$

This shows that if the two-forms $F^{(p-1)}$ and $F^{(p)}$ are closed, then so is $F^{(p+1)}$. Since $F^{(0)}$ and $F^{(1)}$ are known to be closed, it follows through induction that the sequence of two-forms $F^{(n)}$ are all closed and, therefore, are symplectic forms.

We are now ready to prove the main result of our paper.

Proposition: Let f denote a symplectic form on the manifold M. If there exists a tensor field $S \in T^1_1 M$ such that $S\cdot f$ and $S^2\cdot f$ are closed two-forms, then

$$N_S = 0.$$

Proof: By assumption, we have for any vector field X on the manifold,

$$\begin{aligned} &L_X f = d(i_X f), \\ &L_X(S\cdot f) = d(i_X(S\cdot f)), \\ &L_X(S^2\cdot f) = d(i_X(S^2\cdot f)). \end{aligned} \tag{28}$$

Using Eq. (24) we conclude that for $n = 0,1$

$$L_{X\cdot S}(S^{n}\cdot f) = L_X(S^{n+1}\cdot f). \tag{29}$$

Explicitly, we note that

$$\begin{aligned} &L_{X\cdot S} f = L_X(S\cdot f), \\ &L_{X\cdot S}(S\cdot f) = L_X(S^2\cdot f). \end{aligned} \tag{30}$$

Let us next note that

$$\begin{aligned} L_{X\cdot S}(S\cdot f) &= (L_{X\cdot S}S)\cdot f + S\cdot L_{X\cdot S} f \\ &= (L_{X\cdot S}S)\cdot f + S\cdot L_X(S\cdot f). \end{aligned} \tag{31}$$

On the other hand, we have [using Eq. (30)]

$$\begin{aligned} L_{X\cdot S}(S\cdot f) &= L_X(S^2\cdot f) \\ &= L_X(S\cdot S\cdot f) \\ &= (L_X S)\cdot S\cdot f + S\cdot L_X(S\cdot f). \end{aligned} \tag{32}$$

Comparing (31) and (32) we see that for any vector field X on the manifold we have

$$(L_{X\cdot S}S)\cdot f = (L_X S)\cdot S\cdot f. \tag{33}$$

Equivalently,

$$L_{X\cdot S}S = (L_X S)\cdot S.$$

We conclude from Eq. (2) then that, in this case,

$$N_S = 0. \tag{34}$$

Let us note in addition that an integrable system must have exactly $n(\dim M = 2n)$ conserved quantities. In the geometric approach this corresponds to the matrix S^{ν}_{μ} having exactly n nontrivial, doubly degenerate eigenvalues. Consequently, S^{ν}_{μ} would be nonsingular for such a system and hence both $S\cdot f$ and $S^2\cdot f$ would be symplectic forms on the manifold.

This now proves the main claim of the paper. Namely, if the phase space of a dynamical system admits three distinct symplectic forms related by powers of $S \in T^1_1 M$, then the Nijenhuis tensor associated with S vanishes. As it has been shown earlier, this is a sufficient condition for integrability. We now give two examples to bring out the usefulness of the above observation.

Toda lattice: For the Toda chain with N points, it is known[7] that two symplectic forms can be written in terms of $2N \times 2N$ matrices as

$$f_{\mu\nu} = \begin{pmatrix} 0 & -I \\ I & 0 \end{pmatrix} F_{\mu\nu} = \begin{pmatrix} A & -B \\ B & e \end{pmatrix}, \quad \mu,\nu = 1,2,...,2N, \tag{35}$$

where the $N \times N$ blocks take the form

$$\begin{aligned} A_{ij} &= \delta_{i+1,j} e^{-(Q_{i+1}-Q_i)} - \delta_{i,j+1} e^{-(Q_{j+1}-Q_j)}, \\ B_{ij} &= P_i \delta_{i,j}, \\ e_{ij} &= \epsilon(j-i), \quad i,j = 1,2,...,N, \end{aligned} \tag{36}$$

with

$$\epsilon(j-i) = \begin{cases} 1, & \text{if } j > i, \\ 0, & \text{if } j = i, \\ -1, & \text{if } j < i. \end{cases} \tag{37}$$

From this we can construct

$$S_\mu^\nu = F_{\mu\lambda} (f^{-1})^{\lambda\nu} = \begin{pmatrix} B & A \\ -e & B \end{pmatrix}. \tag{38}$$

A third two-form can now be defined as

$$G_{\mu\nu} = S_\mu^\lambda F_{\lambda\nu} = \begin{pmatrix} BA + AB & -B^2 + Ae \\ -eA + B^2 & Be + eB \end{pmatrix}. \tag{39}$$

In components,

$$\begin{aligned} G_{ij} &= \delta_{i+1,j}(P_i + P_{i+1}) e^{-(Q_{i+1}-Q_i)} \\ &\quad - \delta_{i,j+1}(P_j + P_{j+1}) e^{-(Q_{j+1}-Q_j)}, \\ G_{i,N+j} &= -G_{N+j,i} \\ &= -\delta_{ij} P_i^2 + \epsilon(j-i-1) e^{-(Q_{i+1}-Q_i)} \\ &\quad - \epsilon(j-i+1) e^{-(Q_i - Q_{i-1})} \\ G_{N+i,N+j} &= (P_i + P_j)\epsilon(j-i). \end{aligned} \tag{40}$$

It is now quite straightforward to see that

$$\partial_\mu G_{\nu\lambda} + \partial_\nu G_{\lambda\mu} + \partial_\lambda G_{\mu\nu} = 0. \tag{41}$$

According to the main result of the paper, this implies that the Nijenhuis torsion tensor associated with S_μ^ν vanishes[7] and that the system is integrable.

KdV: The dual Poisson bracket structures in this case are known to be[2]

$$\begin{aligned} f^{-1} &= D^3 + \tfrac{1}{3}(Du + uD), \\ F^{-1} &= D, \end{aligned} \tag{42}$$

where $u(x,t)$ is the dynamical variable and D represents the derivative operator with respect to the coordinate x. We can now construct

$$S = Ff^{-1} = D^2 + \tfrac{1}{3}u + \tfrac{1}{3}D^{-1}uD, \tag{43}$$

and, consequently, a third two-form associated with this system has the form

$$G = SF = D + \tfrac{1}{3}uD^{-1} + \tfrac{1}{3}D^{-1}u. \tag{44}$$

In the coordinate basis, this takes the form

$$\begin{aligned} G(x,y) &= \langle y|G|x\rangle \\ &= \frac{\partial}{\partial x}\delta(x-y) + \frac{1}{3}\epsilon(x-y)(u(x) + u(y)), \end{aligned} \tag{45}$$

where we have used

$$\langle y|D^{-1}|x\rangle = \epsilon(x-y) = -\epsilon(y-x) = (\theta(x-y) - \tfrac{1}{2}). \tag{46}$$

It is now obvious that

$$\frac{\delta G(x,y)}{\delta u(z)} + \frac{\delta G(y,z)}{\delta u(x)} + \frac{\delta G(z,x)}{\delta u(y)} = 0, \tag{47}$$

so that the two-form G is closed. It follows from the main result of the paper now that the Nijenhuis torsion tensor associated with the KdV system must vanish. The simplicity of this method should be contrasted with an explicit verification of the vanishing Nijenhuis torsion tensor for the present case.

ACKNOWLEDGMENT

This work was supported in part by the U. S. Department of Energy Contract No. DE-AC02-76ER13065.

[1] For reviews, see M. A. Olshansky and A. M. Perelomov, Phys. Rep. **71**, 315 (1981); A. C. Newell, *Solitons in Mathematical Physics* (SIAM, Philadelphia, 1985); L. D. Faddeev and L. A. Takhtajan, *Hamiltonian Methods in the Theory of Solitons* (Springer, New York, 1987); A. Das, *Integrable Models* (World Scientific, Singapore, 1989).
[2] F. Magri, J. Math. Phys. **19**, 1156 (1978).
[3] F. Magri, C. Morosi, and O. Ragnisco, Commun. Math. Phys. **99**, 115 (1985).
[4] A. Nijenhuis, Indag. Math. **13**, 200 (1951).
[5] F. Magri and C. Morosi, "A geometrical characterization of integrable Hamiltonian systems through the theory of Poisson–Nijenhuis manifolds," Univ. of Milano preprint, 1984.
[6] S. De Filippo, G. Vilasi, G. Marmo, and M. Salerno, Nuovo Cimento B **83**, 97 (1984).
[7] S. Okubo and A. Das, Phys. Lett. B **209**, 311 (1988); A. Das and S. Okubo, Ann. Phys. **190**, 215 (1989).
[8] C. Ferrario, G. LoVecchio, G. Marmo, and G. Morandi, Lett. Math. Phys. **9**, 140 (1985); P. Antonini, G. Marmo, and C. Rubano, Nuovo Cimento B **86**, 17 (1985).
[9] S. Okubo, J. Math. Phys. **30**, 834 (1989).
[10] S. Okubo, "Nijenhuis tensor, BRST cohomology and related algebras," Univ. of Rochester preprint, UR-1110 (1990).

Multi-Hamiltonian structure of equations of hydrodynamic type

H. Gümral
Istanbul Technical University, Physics Department, 80626 Maslak, Istanbul, Turkey

Y. Nutku
Bilkent University, Department of Mathematics, 06533 Bilkent, Ankara, Turkey

(Received 7 November 1989, accepted for publication 11 July 1990)

The discussion of the Hamiltonian structure of two-component equations of hydrodynamic type is completed by presenting the Hamiltonian operators for Euler's equation governing the motion of plane sound waves of finite amplitude and another quasilinear second-order wave equation. There exists a doubly infinite family of conserved Hamiltonians for the equations of gas dynamics that degenerate into one, namely, the Benney sequence, for shallow-water waves. Infinite sequences of conserved quantities for these equations are also presented. In the case of multicomponent equations of hydrodynamic type, it is shown, that Kodama's generalization of the shallow-water equations admits bi-Hamiltonian structure.

I. INTRODUCTION

Historically, equations of hydrodynamic type[1] first arose as quasilinear second-order wave equations in $1+1$ dimensions. Equations of Euler,[2] Poisson,[3] and Born–Infeld[4] are examples of wave equations that can be written as a pair of first-order equations of hydrodynamic type. The equations of gas dynamics

$$u_t + uu_x + v^{\gamma-2}v_x = 0, \quad v_t + vu_x + uv_x = 0, \tag{1}$$

which are obtained from an Eulerian description of motion, provide the prototype of such equations. Recently, Kodama[5] has discussed the reduction of the dispersionless Kadomtsev–Petviashvili (dKP) equation that is based on the compatibility of first-order equations:

$$u^i_y = A^i_j u^j_x,$$
$$u^i_t = B^i_j u^i_x \quad i = 1,2,\ldots,n, \tag{2}$$

which are equations of hydrodynamic type in $2+1$ dimensions.

The Hamiltonian structure of the equations of gas dynamics[6–8] which, in particular, include the shallow-water equations,[9,10] the Poisson,[7] and Born–Infeld[11] equations were discussed earlier. In this paper, we shall conclude the discussion of the Hamiltonian structure of these two-component systems by presenting two further examples. First, we have Euler's equation

$$\Psi_{tt} - (1+\Psi_x)^{-(1+\gamma)}\Psi_{xx} = 0, \tag{3}$$

governing the propagation of plane sound waves of finite amplitude and finally another quasilinear second-order wave equation

$$\Psi_{tt} - e^{\Psi_x}\Psi_{xx} = 0, \tag{4}$$

which is related to a system considered in Ref. 8. Euler's equation results from a Lagrangian, whereas Eqs. (1) are obtained from an Eulerian description of motion. However, we shall show that, even though the equations of motion look quite different in different representations, they have the same Hamiltonian operators in common provided $\gamma + \gamma' = 2$.

We shall conclude by presenting the bi-Hamiltonian structure of the three- and four-component generalizations of the shallow-water equations that was proposed by Kodama.[5] The derivation of the second Hamiltonian operator and proof of Jacobi's identities for multicomponent equations of hydrodynamic type consist of a generalization of the processes described in detail in Refs. 7 and 8. We shall assume familiarity with these papers and present only the new results.

II. GAS DYNAMICS HIERARCHY

The equations of gas dynamics were shown to admit quadri-Hamiltonian structure:[7,8]

$$\begin{aligned}\mathbf{u}_t &= -J_0\mathbf{E}(H^{\mathrm{E}}_1) = -J_1\mathbf{E}(H^{\mathrm{E}}_0)\\ &= -(1/\gamma)J_2\mathbf{E}(H^{\mathrm{E}}_{-1}) = -J_3\mathbf{E}(H^{V}), \quad \gamma\neq 0,\end{aligned} \tag{5}$$

where $u^1 = v, u^2 = u$ are components of the vector $\mathbf{u}$ and $\mathbf{E}$ denotes the Euler operator, or the variational derivative with respect to $\mathbf{u}$. The Hamiltonian operators J_i, $i = 0,1,2,3$ are skew-adjoint matrices of differential operators satisfying the Jacobi identities. The first three of these are first order and mutually compatible, whereas the last one is third order and incompatible with the rest. The bi-Hamiltonian structure is given by

$$J_0 = \sigma_1 D, \quad D = \frac{\partial}{\partial x}, \tag{6}$$

$$J_1 = \frac{1}{\gamma}\begin{pmatrix} vD + Dv & (\gamma-2)Du + uD) \\ Du + (\gamma-2)uD & v^{\gamma-2}D + Dv^{\gamma-2}\end{pmatrix}, \quad \gamma\neq 0, \tag{7}$$

where σ_1 is the Pauli matrix. By Magri's theorem,[12] the recursion operator[13]

$$R = J_1 J_0^{-1} \tag{8}$$

generates, in general, two infinite families of conserved quantities. (See Sec. V for a discussion of the exceptional case $\gamma = 2$.) The third Hamiltonian operator is given by

 0022-2488/90/112606-06$03.00

$$J_2 = \begin{pmatrix} uvD + Duv & \\ D\left[\frac{1}{2}u^2 + \frac{v^{\gamma-1}}{(\gamma-1)}\right] + \left[\frac{1}{2}(\gamma-2)u^2 + \frac{v^{\gamma-1}}{(\gamma-1)}\right]D & \end{pmatrix}$$

$$D\left[\frac{1}{2}(\gamma-2)u^2 + \frac{v^{\gamma-1}}{(\gamma-1)}\right] + \left[\frac{1}{2}u^2 + \frac{v^{\gamma-1}}{(\gamma-1)}\right]D \Bigg), \quad (9)$$

$$uv^{\gamma-2}D + Duv^{\gamma-2}$$

We note that for the case of shallow-water waves, $\gamma = 2$, this operator is trivially related to the earlier ones by

$$J_2 = R^2 J_0 \quad (10)$$

but for a generic γ, J_2 is a new nontrivial Hamiltonian operator. Finally,[14,8] we have the fourth Hamiltonian operator

$$J_3 = DU_x^{-1}DU_x^{-1}\sigma_1 D, \quad (11)$$

where

$$U = \begin{pmatrix} u & v \\ \frac{v^{\gamma-2}}{\gamma-2} & u \end{pmatrix} \quad (12)$$

and this gives rise to the first-order conserved density

$$H^V = v_x/(u_x^2 - v^{\gamma-3}v_x^2), \quad (13)$$

which is due to Verosky.[15]

Equations of polytropic gas dynamics admit the following two sequences of conserved densities:

$$\begin{aligned} &H^{\mathrm{E}}_{-1} = \gamma v, \quad H^{\mathrm{E}}_0 = uv, \\ &H^{\mathrm{E}}_1 = \tfrac{1}{2}u^2 v + v^\gamma/\gamma(\gamma-1), \quad \gamma \neq 0,1, \\ &H^{\mathrm{E}}_2 = \tfrac{1}{6}u^3 v + uv^\gamma/\gamma(\gamma-1), \\ &H^{\mathrm{E}}_3 = \frac{1}{24}u^4 v + \frac{u^2 v^\gamma}{2\gamma(\gamma-1)} + \frac{v^{2\gamma-1}}{2\gamma(\gamma-1)^2(2\gamma-1)}, \\ &\cdots \end{aligned} \quad (14)$$

$$\begin{aligned} &H^{\mathrm{L}}_0 = u, \\ &H^{\mathrm{L}}_1 = \tfrac{1}{2}(\gamma-2)u^2 + v^{\gamma-1}/(\gamma-1), \\ &H^{\mathrm{L}}_2 = \tfrac{1}{6}(\gamma-2)u^3 + uv^{\gamma-1}/(\gamma-1), \\ &H^{\mathrm{L}}_3 = \frac{1}{24}(\gamma-2)u^4 + \frac{1}{2}\frac{u^2 v^{\gamma-1}}{2(\gamma-1)} \\ &\qquad + \frac{v^{2(\gamma-1)}}{2(\gamma-1)^2(2\gamma-3)}, \\ &\cdots, \end{aligned} \quad (15)$$

with superscripts E and L denoting the Euler and Lagrange hierarchies, respectively. The recursion operator keeps the two hierarchies separate.

It is the third conserved quantity in the Euler hierarchy H^{E}_1 which acts as the Hamiltonian function for various physically interesting quasilinear wave equations. The exceptional case $\gamma = 1$ where

$$H^{\mathrm{E}}_1 = \tfrac{1}{2}u^2 v + v(\ln v - 1) \quad (16)$$

corresponds to Poisson's equation in nonlinear acoustics, while $\gamma = -1$, which is also known as Chaplygin gas, yields the Born–Infeld equation and classical shallow-water equations are obtained for $\gamma = 2$. The first nontrivial conserved quantity in the Lagrange hierarchy H^{L}_1 has not played a role in the equations that were discussed earlier.

III. EULER'S EQUATION

In 1757, Euler considered finite amplitude plane sound waves and obtained the quasilinear second-order partial differential equation (3) where $\Psi(x,t)$ is the displacement in a Lagrangian description of motion. This equation has been the subject of extensive classic investigations.[16] We shall now show that the second-order equation which results from the use of the Hamiltonian function H^{L}_1 in the Lagrange hierarchy is precisely Euler's equation.

Equation (3) is a member of a class of completely integrable nonlinear wave equations which admits infinitely many conservation laws.[17] Thus, through the introduction of a new potential Φ, Euler's equation can be realized as the integrability condition of a first-order system according to the formalism of Refs. 7, 17. Such a system is given by

$$\begin{aligned} &\Phi_t = -\gamma'^{-1}(1+\Psi_x)^{-\gamma'}, \quad \gamma' \neq 0, \\ &\Psi_t = \Phi_x, \end{aligned} \quad (17)$$

and using the velocity fields

$$\begin{aligned} &u = \Phi_x, \\ &v = 1 + \Psi_x, \end{aligned} \quad (18)$$

we find the evolution equations

$$\begin{aligned} &u_t = v^{-(1+\gamma')}v_x, \\ &v_t = u_x, \end{aligned} \quad (19)$$

which belong to the class of equations of hydrodynamic type. Equations (19) admit primary Hamiltonian structure

$$\mathbf{u}_t = -(1/\gamma')J_0 \mathbf{E}(H^L_1) \quad (20)$$

with Hamiltonian density

$$H^{\mathrm{L}}_1 = -\tfrac{1}{2}\gamma' u^2 + v^{1-\gamma'}/(1-\gamma'), \quad \gamma' \neq 0,1, \quad (21)$$

and comparison with Eq. (15) leads to the identification

$$\gamma + \gamma' = 2. \quad (22)$$

The choice of the velocity fields in Eqs. (18) and the identification (22) were designed to cast Euler's equation into a form whereby its bi-Hamiltonian structure is manifest. Thus

$$\mathbf{u}_t = -(1/\gamma')J_0 \mathbf{E}(H^L_1) = [\gamma'/(\gamma'-2)]J_1 \mathbf{E}(H^L_0), \quad (23)$$

where the Hamiltonian operators are identical to those given in Eqs. (6) and (7). It is not possible to extend the recursion relation beyond this because H^{L}_{-1} is not defined. On the other hand, using the recursion operator we can construct the next completely integrable equation:

$$\Psi_{tt} + \frac{1}{\gamma'(1+\Psi_x)^{\gamma'}}\Psi_{xt} + \left[\frac{1-3\gamma'}{\gamma'^2(1-\gamma')}\frac{1}{(1+\Psi_x)^{2\gamma'}} - \frac{2\Psi_t}{(1+\Psi_x)^{1+\gamma'}}\right]\Psi_{xx} = 0, \quad (24)$$

for which the first Hamiltonian function is H_2^L and we can therefore extend the recursion relation to include J_2. The third-order Hamiltonian operator J_3 exits for all of these equations. Quasilinear wave equations in the hierarchy of Euler's equation admit quadri-Hamiltonian structure. The first three Hamiltonian structures of Euler's equation were obtained by Kupershmidt[18] using a different approach.

IV. THE EQUATION $\Psi_{tt} = e^{\Psi_x}\Psi_{xx}$

In our discussion of the multi-Hamiltonian structure of Euler's equation we have introduced the first-order system (17) which is not defined for $\gamma' = 0$. On the other hand, the resulting equations of hydrodynamic type (19) are defined for all γ' but they are not bi-Hamiltonian in the exceptional case $\gamma' = 0$. This gap is filled by Eq. (4) which is obtained by an interchange of the roles of x and t in Euler's equation for $\gamma' = 0$.

We consider the first-order system

$$\Phi_t = e^{\Psi_x} - 1,$$
$$\Psi_t = \Phi_x, \tag{25}$$

the integrability conditions of which result in Eq. (4) for the potential Ψ, while Φ satisfies

$$\Phi_{xx} - (1 + \Phi_t)^{-1}\Phi_{tt} = 0, \tag{26}$$

which is precisely Euler's equation for $\gamma' = 0$ with the roles of x and t interchanged. Defining the velocity fields $u = \Phi_x$ and $v = \Psi_x$ we obtain the hydrodynamic system

$$u_t = e^v v_x, \quad v_t = u_x, \tag{27}$$

which are quadri-Hamiltonian equations of hydrodynamic type. The bi-Hamiltonian structure is defined by the operators $\hat{J}_0 = J_0$ and

$$\hat{J}_1 = \begin{pmatrix} e^v D + De^v & uD \\ Du & 2D \end{pmatrix}. \tag{28}$$

The third Hamiltonian operator

$$\hat{J}_2 = \begin{pmatrix} ue^v D + Due^v & (\frac{1}{2}u^2 + e^v)D + De^v \\ e^v D + D(e^v + \frac{1}{2}u^2) & uD + Du \end{pmatrix}, \tag{29}$$

which is trivially related to the first two was obtained earlier.[8] These operators are compatible. There is also a third-order Hamiltonian operator for Eqs. (27)

$$\hat{J}_3 = D\hat{U}_x^{-1}D\hat{U}_x^{-1}\sigma_1 D, \tag{30}$$

$$\hat{U} = \begin{pmatrix} u & v \\ e^v & u \end{pmatrix}, \tag{31}$$

which is incompatible with the rest. These equations admit the first-order conserved density

$$\hat{H} = v_x/(u_x^2 - e^v v_x^2), \tag{32}$$

which is analoguous to Verosky's result.

The recursion operator $\hat{R} = \hat{J}_1 J_0^{-1}$ generates infinitely many conserved quantities the first few of which are

$$\hat{H}_{-1}^E = v, \quad \hat{H}_0^L = u, \quad \hat{H}_1^L = \tfrac{1}{2}u^2 + e^v,$$
$$\hat{H}_2^L = \tfrac{1}{3}u^3 + 2ue^v, \quad \hat{H}_3^L = \tfrac{1}{4}u^4 + 3u^2e^v + \tfrac{3}{2}e^{2v},$$
$$\hat{H}_4^L = \tfrac{1}{5}u^5 + 4u^3e^v + 6ue^{2v}, \quad \ldots \tag{33}$$

Apart from the first one all the conserved quantities in this sequence are related to those in the Lagrangian sequence of generalized gas dynamics. The recursion operator $\hat{R}$ when applied to the conserved density $\hat{H}_{-1}^E$ in the Eulerian sequence generates the analog of the Lagrangian sequence. This situation is in agreement with the fact that the two sequences of conserved quantities for gas dynamics become identical[7] to the Benney sequence[19] when $\gamma = 2$, which corresponds to $\gamma' = 0$ for Euler's equation.

V. FURTHER CONSERVED QUANTITIES FOR SHALLOW-WATER EQUATIONS

Two-component equations of hydrodynamic type admit conserved quantities that satisfy a linear second-order pde in two variables.[17] From general theory we know that the solution of such an equation contains two arbitrary functions. Indeed, for a generic γ the two infinite families of conserved Hamiltonians of the Eulerian and the Lagrangian sequences form a complete set in terms of which we can express these arbitrary functions. However, it was already noted in Ref. 7 that for the case of shallow-water waves ($\gamma = 2$) these two sequences are no longer linearly independent and we have just seen above that the same phenomenon occurs for $\gamma' = 0$. This degeneration of the Eulerian and the Lagrangian sequences into one, namely, the Benney sequence, results in the loss of an arbitrary function. Thus for $\gamma = 2$, or $\gamma' = 0$ we are missing an infinite set of conserved quantities.

Possible end/starting elements of such a missing sequence of conserved quantities are the Casimirs C which satisfy

$$J_1 \mathbf{E}(C) = 0 \tag{34}$$

and for shallow-water equations we find that these distinguished functions are given by

$$C = \sqrt{u^2 - 4v} \tag{35}$$

and u itself. For the nontrivial Casimir (35) simple waves satisfying

$$u^2 = 4v \tag{36}$$

form a dividing line in the discussion which follows and appropriate restrictions must be imposed in order to insure that the arguments of the square roots are positive. The Benney sequence and the Casimirs can be obtained in various limits of Manin's generating function[20]

$$M(k) = \sqrt{(u+k)^2 - 4v} - (u+k) \tag{37}$$

for conserved quantities satisfied by the shallow-water equations. We have the following infinite sequences of conserved quantities which include the non-trivial Casimir

$$\cdots \to \frac{u}{\sqrt{u^2 - 4v}} \to \sqrt{u^2 - 4v} \to 0, \quad u^2 - 4v > 0, \tag{38}$$

which was pointed out in Ref. 21 and

$$0 \to \sqrt{4v - u^2}$$
$$\to 4v \sin^{-1}\left(\frac{u}{2\sqrt{v}}\right) + u\sqrt{4v - u^2} + \cdots, \quad u^2 - 4v < 0, \tag{39}$$

where the arrow indicates the sense of the recursion operator. Once again[8] $\partial/\partial u$ acts as the inverse of the recursion operator.

Simple waves subject to Eq. (36) also play an exceptional role in Miura transformations. For equations of hydrodynamic type Miura transformations reduce to coordinate transformations which bring the flat metric[1] defined by the second Hamiltonian operator of Eq. (7):

$$ds_1^2 = (v - \tfrac{1}{4}u^2)^{-1}(v\,du^2 - u\,du\,dv + dv^2) \tag{40}$$

into canonical form. That is,

$$ds_1^2 = d\tilde{u}^2 + \epsilon\, d\tilde{v}^2, \quad \epsilon = \pm 1 \tag{41}$$

and in terms of $\tilde{u}$ and $\tilde{v}$ the second Hamiltonian operator has constant coefficients which is the effect required of Miura transformations. This transformation is given by

$$\tilde{u} = u,$$

$$\tilde{v} = \begin{cases} \sqrt{u^2 - 4v}, & u^2 > 4v \Rightarrow \epsilon = +1, \\ \sqrt{4v - u^2}, & u^2 < 4v \Rightarrow \epsilon = -1 \end{cases}, \tag{42}$$

which results in either a Euclidean or a Lorentzian signature for the metric. Dubrovin and Novikov[1] have remarked that the only non-trivial invariant that a flat metric can admit is its signature. But this statement must be accompanied by the warning that it is not sufficient to know the form of the metric in one particular coordinate system in order to determine its signature. The metric may contain apparent singularities and this will put restrictions of the domain of coordinates for which it is going to be valid. The signature cannot be determined without reference to these domains. The flat metric defined by the second Hamiltonian operator for shallow-water waves is a case in point because the Jacobian of the Miura transformation vanishes when Eq. (36) is satisfied.

VI. KODAMA'S GENERALIZATION OF SHALLOW-WATER EQUATIONS

Kodama[5] has shown that the reduction of the dKP equation which results in the shallow-water equations leads to the following generalization for a three-component field

$$\begin{pmatrix} v \\ w \\ u \end{pmatrix}_t = \begin{pmatrix} u & w & v \\ 0 & u & w \\ 0 & 0 & u \end{pmatrix} \begin{pmatrix} v \\ w \\ u \end{pmatrix}_x \tag{43}$$

which are equations of hydrodynamic type. They admit an infinite sequence of conserved Hamiltonians

$$H_{-1} = 2v, \quad H_0 = uv + \tfrac{1}{2}w^2,$$
$$H_1 = \tfrac{1}{2}(u^2 v + v^2 + uw^2),$$
$$H_2 = \tfrac{1}{4}u^3 v + \tfrac{3}{4}uv^2 + \tfrac{3}{8}u^2 w^2 + \tfrac{3}{8}vw^2, \ldots, \tag{44}$$

analoguous to the Benney sequence for shallow-water waves. Kodama has further noted that his equations can be written in Hamiltonian form

$$\mathbf{u}_t = J_0 \mathbf{E}(H_1), \tag{45}$$

where

$$J_0 = \begin{pmatrix} 0 & 0 & D \\ 0 & D & 0 \\ D & 0 & 0 \end{pmatrix} \tag{46}$$

and $\mathbf{u} = (v,w,u)$ in that order. We shall now show that Kodama's equations admit bi-Hamiltonian structure.

We shall seek a second Hamiltonian operator for Eqs. (43) using a procedure similar to that presented in Ref. 7 for two component equations. Since the calculations are straightforward and lengthy we shall only present the results. The second Hamiltonian operator for Kodama's equations is given by

$$J_1 = \frac{1}{2}\begin{pmatrix} vD + Dv & wD + D\frac{w}{2} & uD \\ \frac{w}{2}D + Dw & \frac{1}{4}\left(2u - \frac{v^2}{w^2}\right)D + D\frac{1}{4}\left(2u - \frac{v^2}{w^2}\right) & \frac{v}{2w}D \\ Du & D\frac{v}{2w} & \frac{3}{2}D \end{pmatrix} \tag{47}$$

and it can be verified that this expression for J_1 satisfies the Jacobi identities and is compatible with J_0. The recursion operator for this bi-Hamiltonian system generates the infinite sequence of conserved Hamiltonians obtained by Kodama starting with v. But, in addition, there is an infinite sequence of conserved quantities starting with w:

$$H_0' = w, \quad H_1' = \frac{1}{2}uw + \frac{1}{4}\frac{v^2}{w},$$

$$H_2' = \frac{3}{8}u^2 w + \frac{3}{4}vw + \frac{3}{8}\frac{uv^2}{w} - \frac{1}{32}\frac{v^4}{w^3},$$

$$H_3' = \frac{5}{16}u^3 w + \frac{15}{8}uvw + \frac{5}{16}w^3 + \frac{15}{32}\frac{u^2 v^2}{w} - \frac{5}{64}\frac{uv^4}{w^3} + \frac{5}{16}\frac{v^3}{w} + \frac{1}{128}\frac{v^6}{w^5}, \tag{48}$$

$\cdots$.

Thus we have two infinite families of conserved Hamiltonians for the three-component Kodama equations, which are rational functions. However, we are missing an infinite family of conserved quantities.

As in our earlier discussion of the conservation laws for the classical shallow-water equations, the missing infinite sequence of conserved quantities must include the Casimir. For the Hamiltonian operator (47) we find that

$$C=(3v-u^2)^{1/2}\left(\frac{(1+\xi)^{1/2}-1}{(1+\xi)^{1/2}+1}\right)^{1/6},$$

$$\xi=\frac{16}{(27)^2}\frac{(3v-u^2)^3}{(w^2+\frac{4}{27}u^3-\frac{2}{3}uv)^2} \tag{49}$$

is the nontrivial Casimir. Once again u is also a Casimir.

Finally, we note that the four-component form of Kodama's equation

$$\begin{pmatrix} v \\ r \\ s \\ u \end{pmatrix}_t = \begin{pmatrix} u & s & r & v \\ 0 & u & s & r \\ 0 & 0 & u & s \\ 1 & 0 & 0 & u \end{pmatrix} \begin{pmatrix} v \\ r \\ s \\ u \end{pmatrix}_x \tag{50}$$

is also a bi-Hamiltonian system with

$$J_0=\begin{pmatrix} 0 & 0 & 0 & D \\ 0 & 0 & D & 0 \\ 0 & D & 0 & 0 \\ D & 0 & 0 & 0 \end{pmatrix} \tag{51}$$

and

$$J_1=\frac{1}{6}\begin{pmatrix} 6vD+3v_x & 5rD+2r_x & 4sD+s_x & 3uD \\ 5rD+3r_x & 4mD+2m_x & 3nD+n_x & 2pD \\ 4sD+3s_x & 3nD+2n_x & 2qD+q_x & (r/s)D \\ 3Du & 2Dp & D(r/s) & 4D \end{pmatrix}, \tag{52}$$

where

$$\begin{aligned} m&=s-\frac{1}{2}\frac{v^2}{s^2}+\frac{vr^2}{s^3}-\frac{1}{2}\frac{r^4}{s^4}, \\ p&=\frac{v}{s}-\frac{1}{2}\frac{r^2}{s^2}, \\ n&=u-\frac{vr}{s^2}+\frac{2}{3}\frac{r^3}{s^3}, \\ q&=v/s-r^2/s^2. \end{aligned} \tag{53}$$

The n-component Kodama equations are evidently also bi-Hamiltonian but the explicit expression for the second Hamiltonian operator is rather involved.

We note that in Kodama's equations (43) the limit $w\to 0$ is well defined and yields the classical shallow-water equations (1) with $\gamma=2$. However, we have found that this limit is not defined for the second Hamiltonian operator (47) of Kodama's equations. Thus it is not possible to obtain the Hamiltonian operator of Eq. (7) starting from Eq. (47). Similar remarks apply to the four-component Kodama equations (50) as well. This may seem surprising at first sight, indeed it was the source of failure of early easy guesses for the possible bi-Hamiltonian structure of Eqs. (43), but the nonexistence of the $w\to 0$ limit in Eq. (47) can be traced back to the following: There is a dimensional reason for the appearance of inverse powers of w in the second Hamiltonian operator J_1 for Kodama's equations. In the n-component Kodama equations the variables $\{u_i\}, i=1,2,\ldots,n$ carry the dimension

$$[u^i]=(2n-1-i)/(n-1) \tag{54}$$

and this results in the requirement that the i,k entry of the Hamiltonian operator must have the dimension

$$[J^{ik}]=(2n-i-k)/(n-1). \tag{55}$$

So we must start with an *Ansatz* for the entries of J^{ik} with the appropriate dimension (55) but if we were to exclude terms containing the inverse powers of some of the variables in this *Ansatz*, it can be verified that the Jacobi identities cannot be satisfied. The results presented in Eqs. (47) and (52) for the second Hamiltonian operators of Kodama's equations (43) and (50) satisfy these dimensional considerations and the Jacobi identities.

VII. CONCLUSION

We have shown that Euler's equation governing the propagation of plane sound waves of finite amplitude can be cast into the form of two-component equations of hydrodynamic type. The first Hamiltonian function of this system can be identified with the first conserved quantity in the Lagrangian sequence of gas dynamics. The quadri-Hamiltonian structure of gas dynamics can therefore be carried over to Euler's equation. We have further shown that the quasilinear second-order wave equation (4) also admits quadri-Hamiltonian structure.

The multi-Hamiltonian structure of equations of hydrodynamic type is a remarkably rich subject as the results reported in Refs. 7, 8, and 11 together with this paper will indicate. But so far we have mostly considered two-component equations and only begun to investigate the multicomponent case by exhibiting the bi-Hamiltonian structure of Kodama's equations generalizing the shallow-water equations. The Hamiltonian structure of multicomponent equations of hydrodynamic type which are obtained from different reductions of the dKP equation requires further investigation.

[1] B. A. Dubrovin and S. A. Novikov, Sov. Math. Dokl. **27**, 665 (1983).
[2] L. Euler, Hist. Acad. Berlin **11**, 449 (1757).
[3] S. D. Poisson, J. Ec. Polytech. **7**, 319 (1808).
[4] M. Born and L. Infeld, Proc. R. Soc. London Ser. A **144**, 425 (1934).
[5] Y. Kodama, Prog. Theor. Phys. Suppl. **94**, 184 (1988).
[6] Y. Nutku, J. Phys. A: Math. Gen. **16**, 4195 (1983).
[7] Y. Nutku, J. Math Phys. **28**, 2579 (1987).
[8] P. J. Olver and Y. Nutku, J. Math. Phys. **29**, 1610 (1988).
[9] J. Cavalcante and H. P. McKean, Physica D **4**, 253 (1982).
[10] B. A. Kupershmidt, Commun. Math. Phys. **99**, 51 (1985).

[11]M. Arik, F. Neyzi, Y. Nutku, P. Olver, and J. Verosky, J. Math. Phys. **30**, 1338 (1988).
[12]F. Magri, J. Math. Phys. **19**, 1356 (1978).
[13]P. J. Olver, *Applications of Lie Groups to Differential Equations*, Graduate Texts in Mathematics, Vol. 107 (Springer, New York, 1986).
[14]M. B. Sheftel, Func. Anal. Appl. **20**, 227 (1986).
[15]J. M. Verosky, J. Math. Phys. **25**, 884 (1984).
[16]Rayleigh, Proc. R. Soc. London Ser. A **84**, 247 (1910).
[17]Y. Nutku, J. Math. Phys. **26**, 1237 (1984).
[18]B. A. Kupershmidt, "Discrete Lax Equation and Differential-Difference Calculus," Societe Mathematique de France (1985).
[19]D. J. Benney, Stud. Appl. Math. **52**, 45 (1973).
[20]Yu. Manin, Itogi Nauki Tekh., Sov. Prob. Mat. **11**, 5 (1978).
[21]M. Antonowicz and A. P. Fordy, Physica D **28**, 345 (1987).

Multidimensional solitons and their spectral transforms

M. Boiti
Dipartimento di Fisica dell'Università, 73100 Lecce, Italy

J. P. Leon
Laboratoire de Physique Mathématique, U. S. T. L., 34060 Montpellier Cedex, France

F. Pempinelli
Dipartimento di Fisica dell'Università, 73100 Lecce, Italy

(Received 2 November 1988; accepted for publication 27 June 1990)

The soliton solution to the hierarchy of two-dimensional nonlinear evolution equations related to the Zakharov–Shabat spectral problem (including the Davey–Stewartson equation) are derived and studied. The solitons are localized two-dimensional structures traveling on straight lines at constant velocities. Their spectral transform is not uniquely defined and this point is discussed by giving two explicit different spectral transforms of the one-soliton solution and also by giving the general dependence of the spectral transform on the definition of the basic Jost-like solutions.

I. INTRODUCTION

Since the discovery that the inverse spectral transform can be formulated as a "$\bar\partial$ problem" a large class of nonlinear evolution equations in two spatial dimensions have been linearized.[1-9] These equations are generally integrodifferential equations and some of them are physically relevant equations. For the purpose of this paper, we consider as prototype the Davey–Stewartson (DS) equations[10]

$$iQ_t = -\tfrac{1}{2}\sigma_3(Q_{xx} + \alpha^2 Q_{yy}) + \sigma_3 Q^3 + [Q,A],$$
$$(\partial_x + \alpha\sigma_3\partial_y)A = \alpha(Q^2)_y, \quad \alpha^2 = \pm 1 \tag{1.1}$$

and the two-spatial dimensional analog of the modified KdV (2DMKdV) equation[11,12]

$$Q_t = \tfrac{1}{4}Q_{xxx} + \tfrac{3}{4}\alpha^2 Q_{xyy} - \tfrac{3}{2}Q^2 Q_x - \tfrac{3}{2}\sigma_3[Q,A]_x - \tfrac{3}{4}\alpha\{Q_y,A\} + \tfrac{3}{4}[Q,B],$$
$$(\partial_x + \alpha\sigma_3\partial_y)A = \alpha(Q^2)_y,$$
$$(\partial_x + \alpha\sigma_3\partial_y)B = -\alpha\sigma_3[Q_{xy},Q], \quad \alpha^2 = \pm 1. \tag{1.2}$$

The 2×2 matrix field Q is off-diagonal

$$Q = \begin{pmatrix} 0 & q(x,y,t) \\ r(x,y,t) & 0 \end{pmatrix}, \tag{1.3}$$

while the auxiliary 2×2 matrix fields A and B are diagonal.

The DS equations have been first derived by the method of multiple scales[10] for describing the propagation of two-dimensional nonlinear waves in dispersive media (strong nonlinearity and weak dispersion). It turns out that from both physical and mathematical point of view, the sign of α^2 is critical ($\alpha = 1$ is DSI and $\alpha = i$ is DSII).

The spectral transform of the DS and 2DMKdV equations is obtained from the spectral problem

$$(\partial_x + \alpha\sigma_3\partial_y + Q)\psi = 0, \tag{1.4}$$

which generalizes the Zakharov–Shabat spectral problem to two-spatial dimensions.

However, as far as we know, only recently it has been realized[11,12] that the second nonlinear equation in the hierarchy related to the Zakharov–Shabat two-dimensional problem can be written in the simple form (1.2), making evident that it is a two-dimensional analog of the mKdV equation. We hope that in this form it will be easier to find interesting physical applications.

Both equations, DS and 2DMKdV, admit the reduction

$$r = \epsilon\bar q, \quad \epsilon \in \mathbb{R}, \tag{1.5}$$

where the bar means complex conjugate.

In two previous papers,[11,12] we showed, together with Martina, that both equations in the hyperbolic case ($\alpha = 1$) admit two-dimensional soliton solutions (exponentially localized spatial structures moving with constant shape and velocity). Therefore, we claim that DS and 2DMKdV equations must play the same fundamental role in the theory of integratable systems in $2+1$ dimensions as in $1+1$ dimensions the nonlinear Schrödinger (NLS) equation and the modified KdV (mKdV) equation (to which they reduce when $Q_y \equiv 0$).

In this paper, we give the most general soliton solution for the DSI and the 2DMKdVI equations. The one-soliton solution is obtained by using the Bäcklund gauge transformation.[13] It is possible to build the two-soliton solution by means of the nonlinear superposition formula. This has been done in Refs. 11 and 12 for a particular choice of the parameters. Our purpose here is to focus on the soliton solution in the context of the spectral transform and we report the study of the general two-soliton solution to future work.

The Bäcklund gauge allows us to derive also the eigenfunction $\psi(x,y,k)$ of the corresponding spectral problem (1.4). Once fixed, the normalization of ψ according to the following asymptotic behavior:

$$\psi(x,y,k)e^{-ik(\sigma_3 x - y)} = \mathbb{1} + O(1/k), \quad \text{as } k\to\infty, \tag{1.6}$$

the spectral transform $R(k,l)$ of $Q(x,y)$ is defined as the measure of departure from analyticity of $\psi(x,y,k)$ as follows:

$$\frac{\partial\psi}{\partial\bar k}(x,y,k) = \int\int dl\wedge d\bar l\, \psi(x,y,l)R(k,l). \tag{1.7}$$

However, the normalization condition (1.6) does not fix the boundary conditions in the (x,y) plane. Any possible different choice of boundary conditions in the (x,y) plane

 0022-2488/90/112612-07$03.00

corresponds to different disjoint classes of spectral data $R(k,l)$. We shall prove, however, that any admissible choice of boundary conditions in the (x,y) plane still produce a $\bar\partial$ problem in the form (1.7). In particular, we show that, for the one soliton solution, if one chooses as boundary conditions for $\psi(x,y,k)$ those naturally suggested by the Bäcklund gauge, $\psi(x,y,k)$ has a simple pole in the complex k plane. On the contrary, if one chooses the boundary conditions proposed by Fokas and Ablowitz[14] the corresponding eigenfunction $\psi(x,y,k)$ has only a discontinuity on the real k axis (no discrete spectrum) also for the pure one-soliton solution. We shall compute this particular solution ψ (the one analytic in the upper complex k plane) and show that it has a continuation in the lower half plane where it has an *infinite set of simple poles* accounting for the presence of the one-soliton solution.

II. HYPERBOLIC SYSTEMS IN THE PLANE

Let us consider the hyperbolic 2×2 Zakharov–Shabat spectral problem in the plane

$$T_1 \equiv \partial_x + \sigma_3 \partial_y + Q, \quad Q = \begin{pmatrix} 0 & q(x,y) \\ r(x,y) & 0 \end{pmatrix},$$
$$T_1 \psi = 0. \tag{2.1}$$

The complex spectral parameter k is introduced by seeking matrix eigenfunctions $\psi(x,y,k)$ that are normalized as follows:

$$\psi(x,y,k) e^{-ik(\sigma_3 x - y)} = \mathbb{1} + O(1/k), \tag{2.2}$$

as $k \to \infty$ and are bounded in the (x,y) plane. For $Q(x,y) \to 0$ rapidly enough for large (x,y), the matrix

$$\nu(x,y,k) = \psi(x,y,k) e^{-ik(\sigma_3 x - y)} \tag{2.3}$$

satisfies the integral equation

$$\nu = \nu_0 - \int_{-\infty}^{x} dx' \exp[ik(x-x')\sigma_3] \begin{pmatrix} q\nu_{21} & q\nu_{22} \\ 0 & r\nu_{12} \end{pmatrix} (x', y-(x-x')\sigma_3)$$
$$\times \exp[-ik(x-x')\sigma_3] + \int_{x}^{\infty} dx' \exp[ik(x-x')\sigma_3] \begin{pmatrix} 0 & 0 \\ r\nu_{11} & 0 \end{pmatrix} (x', y-(x-x')\sigma_3) \exp[-ik(x-x')\sigma_3], \tag{2.4}$$

for $\operatorname{Im} k > 0$ and

$$\nu = \nu_0 - \int_{-\infty}^{x} dx' \exp[ik(x-x')\sigma_3] \begin{pmatrix} q\nu_{21} & 0 \\ r\nu_{11} & r\nu_{12} \end{pmatrix} (x', y-(x-x')\sigma_3)$$
$$\times \exp[-ik(x-x')\sigma_3] + \int_{x}^{\infty} dx' \exp[ik(x-x')\sigma_3] \begin{pmatrix} 0 & q\nu_{22} \\ 0 & 0 \end{pmatrix} (x', y-(x-x')\sigma_3) \exp[-ik(x-x')\sigma_3], \tag{2.5}$$

for $\operatorname{Im} k < 0$, where, for any matrix $M(x)$, $M(y-(x-x')\sigma_3)$ denotes the matrix obtained from $M(x)$ by evaluating the lth row at $y-(x-x')(\sigma_3)_{ll}$ and where

$$\nu_0 = \begin{pmatrix} \alpha_1(x-y,k) & 0 \\ 0 & \alpha_2(x+y,k) \end{pmatrix}, \tag{2.6}$$

with α_i arbitrary functions. The matrix ν_0 is subjected to the sole requirement that $\nu_0 \to \mathbb{1}$ as $k \to \infty$.

In the papers by Fokas and Ablowitz,[14] ν_0 has been chosen to be equal to 1. This choice seems to be the more convenient one for showing that the nonlinear evolution equations associated to (2.1) are linearizable. However, we show that this choice may not be convenient if one is interested in finding special solutions.

Once $\nu_0(x,y,k)$ chosen, the matrix spectral transform $R(k,l)$ is defined as the measure of departure from analyticity of $\psi(x,y,k)$ as follows:

$$\frac{\partial \psi}{\partial \bar k}(x,y,k) = \int\int dl \wedge d\bar l \, \psi(x,y,l) R(k,l). \tag{2.7}$$

If one supposes, as is generally admitted, that $R(k,l)$ determines uniquely $\psi(x,y,k)$ then the same Q possesses different spectral transforms $R(k,l)$ corresponding to different choices of ν_0.

In order to define the spectral transform of Q when ν_0 is not equal to $\mathbb{1}$, it is convenient to work with the eigenfunctions $\psi^{(\pm)}(x,y,k)$ solutions of the following integral equations ($k \in \mathbb{R}$):

$$\psi^{(+)}(x,y,k) = \nu_0 e^{ik[\sigma_3 x - y]} - \int_{-\infty}^{x} dx'$$
$$\times \begin{pmatrix} q\psi_{21}^{(+)} & q\psi_{22}^{(+)} \\ 0 & r\psi_{12}^{(+)} \end{pmatrix} (x', y-(x-x')\sigma_3)$$
$$+ \int_{x}^{\infty} dx' \begin{pmatrix} 0 & 0 \\ r\psi_{11}^{(+)} & 1 \end{pmatrix} (x', y-(x-x')\sigma_3)$$
$$\doteq \nu_0 e^{ik[\sigma_3 x - y]} + G_{xy}^{(+)} \psi^{(+)}(k), \tag{2.8}$$

$$\psi^{(-)}(x,y,k) = \nu_0 e^{ik[\sigma_3 x - y]} - \int_{-\infty}^{x} dx'$$
$$\times \begin{pmatrix} q\psi_{21}^{(-)} & 0 \\ r\psi_{11}^{(-)} & r\psi_{12}^{(-)} \end{pmatrix} (x', y-(x-x')\sigma_3)$$
$$+ \int_{x}^{\infty} dx' \begin{pmatrix} 0 & q\psi_{22}^{(-)} \\ 0 & 0 \end{pmatrix} (x', y-(x-x')\sigma_3)$$
$$\doteq \nu_0 e^{ik[\sigma_3 x - y]} + G_{xy}^{(-)} \psi^{(-)}(k). \tag{2.9}$$

The advantage of working with ψ instead of v is that the integral operators $G^{(\pm)}$ are k independent. Through a convenient factorization, the discontinuity of ψ on the real k axis obeys the following equation:

$$\{\psi^{(+)}(x,y,k) - \psi^{(-)}(x,y,k)\} = \int_{-\infty}^{+\infty} dm\, e^{im[\sigma_3 x - y]} S(k,m) + G_{xy}^{(-)}\{\psi^{(+)}(k) - \psi^{(-)}(k)\}, \tag{2.10}$$

where

$$2\pi S(k,m) = \int_{-\infty}^{+\infty} dx \int_{-\infty}^{+\infty} dy\, e^{-im[\sigma_3 x - y]} \times \begin{pmatrix} 0 & -q\psi_{22}^{(+)}(x,y,k) \\ r\psi_{11}^{(+)}(x,y,k) & 0 \end{pmatrix}. \tag{2.11}$$

To obtain (2.10), it is necessary to use the identity

$$2\pi M(x', y - (x - x')\sigma_3) \equiv \int_{-\infty}^{+\infty} dy' \int_{-\infty}^{+\infty} dm\, e^{-im[y - (x - x')\sigma_3 - y']} M(x',y'), \tag{2.12}$$

where M is a generic 2×2 matrix. The next step consists in multiplying (2.9), written for $k = m$, by some matrix $R_c(k,m)$ (the subscript c stands for "continuous") and then integrating over m. The resulting equation can be written

$$\int_{-\infty}^{+\infty} dm\, \psi^{(-)}(x,y,m) R_c(k,m) = \frac{i}{2} \int_{-\infty}^{+\infty} dm\, e^{im[\sigma_3 x - y]} S(k,m) + G_{xy}^{(-)} \int_{-\infty}^{+\infty} dm\, \psi^{(-)}(m) R_c(k,m), \tag{2.13}$$

if we choose $R_c(k,m)$ according to

$$\frac{i}{2} S(k,m) = \int_{-\infty}^{+\infty} dl\, N(l, m-l) R_c(k, m-l) \tag{2.14}$$

(note that we have commuted $G^{(-)}$ with the integral over m), where N is the Fourier transform of the arbitrary function v_0 described in (2.6), namely

$$v_0(x,y,k) = \int_{-\infty}^{+\infty} dl\, e^{il[\sigma_3 x - y]} N(l,k). \tag{2.15}$$

The final step consists in comparing (2.10) with (2.13) to conclude that, if the honogeneous integral equation has only the trivial solution, then

$$(i/2)[\psi^{(+)}(x,y,k) - \psi^{(-)}(x,y,k)] = \int_{-\infty}^{+\infty} dm\, \psi^{(-)}(x,y,m) R_c(k,m). \tag{2.16a}$$

For $v_0 = \mathbb{1}$, we obtain $N(l,k) = \delta(l)$ and then $R_c = S$ in agreement with.[14] For other choices of v_0, or equivalently other choices of $N(l,k)$, the direct spectral problem (construction of R_c from Q) requires the solution of (2.14) to obtain R_c from S.

Note that the spectral transform R defined in (1.7) is obtained from R_c by

$$R(k,l) = \delta(k_I)\delta(l_I + 0) R_c(k_R, l_R). \tag{2.16b}$$

The time evolution of Q (for a convenient choice of v_0) can be fixed by introducing the 2×2 matrix dispersion relation $\Omega(k,t)$ and by requiring that the spectral transform $R(k,l)$ evolves as

$$R_t(k,l,t) = R(k,l,t)\Omega(k,t) - \Omega(l,t)R(k,l,t). \tag{2.17}$$

A very general choice for Ω and the related evolution equations have been given in Ref. 15. Here, we are interested in the case

$$\Omega = \omega(k)\sigma_3, \tag{2.18}$$

with $\omega(k)$ a polynomial in k satisfying the additional condition

$$\bar{\omega}(k) = -\omega(\bar{k}). \tag{2.19}$$

This condition ensures that the linear approximation of the considered equations does not blow up at large time.

The nonlinear evolution equation satisfied by $Q(x,y,t)$ is obtained by seeking an auxiliary spectral operator (the second operator in the Lax pair) of the form

$$T_2 = \partial_t + V, \tag{2.20}$$

where $V = V(x,y,t,\partial_y)$ is a polynomial in the differential operator ∂_y. Here T_2 commutes with T_1:

$$[T_2, T_1] = 0, \tag{2.21}$$

and satisfies the equation

$$T_2\psi = \psi\Omega. \tag{2.22}$$

The two prototype equations, DSI and 2DMKdVI, given in the Introduction are obtained by choosing, respectively,

$$\omega(k) = ik^2 \tag{2.23}$$

and

$$\omega(k) = ik^3. \tag{2.24}$$

In the first case,

$$iV = \sigma_3\partial_y^2 + Q\partial_y - \tfrac{1}{2}\sigma_3 Q_x + \tfrac{1}{2}Q_y - \tfrac{1}{2}\sigma_3 Q^2 + A, \tag{2.25}$$

and, in the second case,

$$V = \sigma_3\partial_y^3 + Q\partial_y^2 + V_1\partial_y + V_2, \tag{2.26}$$

with

$$V_1 = -\tfrac{1}{2}\sigma_3 Q_x + Q_y - \tfrac{1}{2}\sigma_3 Q^2 + \tfrac{3}{2}A,$$
$$V_2 = \tfrac{1}{4}Q_{xx} + Q_{yy} - \tfrac{1}{4}\sigma_3 Q_{xy} - \tfrac{1}{2}Q^3 - \tfrac{3}{4}\sigma_3[Q,A] - \tfrac{1}{4}[Q,Q_x] - \tfrac{1}{2}\sigma_3 Q_y Q + \tfrac{3}{4}A_y + \tfrac{3}{4}\sigma_3[Q_y,Q] + \tfrac{3}{4}B. \tag{2.27}$$

III. BACKLUND GAUGES AND SOLITON SOLUTIONS

The simplest gauge transformation, $B = B(Q',Q)$, which generates the Bäcklund transform Q' of Q (see Ref. 13) is given by the equation

$$B(Q',Q) = a\partial_y - \tfrac{1}{2}\sigma_3(Q'a - aQ) - \tfrac{1}{2}\sigma_3 a \mathscr{J}(Q'^2 - Q^2) + b. \tag{3.1}$$

The operator $\mathscr{J} = (\partial_x + \sigma_3\partial_y)^{-1}$ is defined with the boundary requirement that, in a given direction, $\lim \mathscr{J} M = 0$ for any diagonal (well behaved at infinity)

matrix $M = M(x,y)$. Here, a,b are 2×2 constant diagonal matrices.

For $a = \left(\begin{smallmatrix}1&0\\0&0\end{smallmatrix}\right)$, $b = \left(\begin{smallmatrix}\lambda&0\\0&1\end{smallmatrix}\right)$ we get the so-called elementary Bäcklund gauge of first kind $B^{(\mathrm{I})}(Q',Q;\lambda)$ for $a = \left(\begin{smallmatrix}0&0\\0&1\end{smallmatrix}\right)$, $b = \left(\begin{smallmatrix}1&0\\0&\mu\end{smallmatrix}\right)$ the elementary Bäcklund gauge of second kind $B^{(\mathrm{II})}(Q',Q;\mu)$ and for $a = \left(\begin{smallmatrix}1&0\\0&1\end{smallmatrix}\right)$, $b = \left(\begin{smallmatrix}\lambda&0\\0&\mu\end{smallmatrix}\right)$ the Bäcklund gauge $B(Q',Q;\lambda,\mu)$ which generates the soliton.

We are interested in the solution Q that can be derived by comparing two elementary Bäcklund gauges and by starting from the zero solution. The main property to be used is the commutativity of Bäcklund gauges that is given in the specific case by the formula

$$B(Q,0;\lambda,\mu) = B^{(\mathrm{II})}(Q,Q^{(\mathrm{I})};\mu)B^{(\mathrm{I})}(Q^{(\mathrm{I})},0;\lambda) = B^{(\mathrm{I})}(Q,Q^{(\mathrm{II})};\lambda)B^{(\mathrm{II})}(Q^{(\mathrm{II})},0;\mu). \quad (3.2)$$

It results that[15]

$$Q^{(\mathrm{I})} = \begin{pmatrix}0&0\\r^{(\mathrm{I})}&0\end{pmatrix}, \quad Q^{(\mathrm{II})} = \begin{pmatrix}0&q^{(\mathrm{II})}\\0&0\end{pmatrix}, \quad Q = \begin{pmatrix}0&q\\r&0\end{pmatrix}, \quad (3.3)$$

where

$$r^{(\mathrm{I})} = \rho(x+y,t)e^{-\lambda(x-y)}, \quad q^{(\mathrm{II})} = \eta(x-y,t)e^{\mu(x+y)}, \quad (3.4)$$

$$q = \frac{q_y^{(\mathrm{II})} + (\lambda-\mu)q^{(\mathrm{II})}}{1+(1/4)r^{(\mathrm{I})}q^{(\mathrm{II})}}, \quad r = \frac{r_y^{(\mathrm{I})} - (\lambda-\mu)r^{(\mathrm{I})}}{1+(1/4)r^{(\mathrm{I})}q^{(\mathrm{II})}}, \quad (3.5)$$

with

$$\rho(x+y,t) = \int\int dl\wedge d\bar{l}\, e^{-il(x+y)}\tilde{\rho}(l,t),$$

$$\eta(x-y,t) = \int\int dl\wedge d\bar{l}\, e^{il(x-y)}\tilde{\eta}(l,t). \quad (3.6)$$

The time evolution is given by the equations

$$\tilde{\rho}(l,t) = \tilde{\rho}(l,0)\exp[\omega(l)t + \omega(-i\lambda)t],$$
$$\tilde{\eta}(l,t) = \tilde{\eta}(l,0)\exp[-\omega(l)t - \omega(-i\mu)t], \quad (3.7)$$

where $\tilde{\rho}(l,0)$ and $\tilde{\eta}(l,0)$ are arbitrary functions. The explicit form of $B(Q,0;\lambda,\mu)$ is

$$B(Q,0;\lambda,\mu) = 1\partial_y + B^{(1)}(Q,0;\lambda,\mu),$$
$$B^{(1)}(Q,0;\lambda\mu) = \begin{pmatrix}\lambda - \frac{1}{4}r^{(\mathrm{I})}q & -\frac{1}{2}q\\ \frac{1}{2}r & \mu - \frac{1}{4}q^{(\mathrm{II})}r\end{pmatrix}. \quad (3.8)$$

By applying the gauge operator $B(Q,0;\lambda,\mu)$ to $e^{ik(\sigma_3 x - y)}$, we get the following eigenfunction relative to Q in (3.3)

$$\psi(x,y,k) = \left\{1 - \frac{i}{4}\begin{pmatrix}\frac{r^{(\mathrm{I})}q}{k+i\lambda} & \frac{2q}{k+i\mu}\\ -\frac{2r}{k+i\lambda} & \frac{q^{(\mathrm{II})}r}{k+i\mu}\end{pmatrix}\right\} \times \exp[ik(\sigma_3 x - y)]. \quad (3.9)$$

The spectral transform of Q is computed from (2.7) and it results that

$$R(k,l) = i\pi\begin{pmatrix}0 & \delta(k+i\mu)\tilde{\eta}(l,t)(l+i\lambda)\\ -\delta(k+i\lambda)\tilde{\rho}(l,t)(l+i\mu) & 0\end{pmatrix}, \quad (3.10)$$

and, consequently, the solution Q is indeed related to discrete eigenvalues of the spectral problem (2.1).

In order to get a soliton solution from the general solution Q, we have to choose the arbitrary functions $\rho(x+y,0)$ and $\eta(x-y,0)$ in such a way that Q is localized in the space at any time. This can be easily done in the reduced case

$$r = \epsilon\bar{q}, \quad \epsilon\in\mathbb{R}. \quad (3.11)$$

If we introduce the following functions

$$\rho(x+y,0) = 2\exp[-\mu(x+y)]S(\sigma),$$
$$\eta(x-y,0) = 2\exp[\lambda(x-y)]T(\tau), \quad (3.12)$$

with

$$\sigma = (\mu+\bar{\mu})^{-1}\exp[(\mu+\bar{\mu})(x+y)],$$
$$\tau = [\epsilon/(\lambda+\bar{\lambda})]\exp[-(\lambda+\bar{\lambda})(x-y)], \quad (3.13)$$

the reduction condition (3.11) at $t=0$ becomes the following functional equation;

$$[1+\bar{S}(\sigma)\bar{T}(\tau)]\frac{dS}{d\sigma}(\sigma) = [1+S(\sigma)T(\tau)]\frac{d\bar{T}}{d\tau}(\tau). \quad (3.14)$$

It admits the following four different solutions:

$$\begin{cases}S = a\sigma + b, & b\bar{a} - \bar{b}a = 0,\\ T = \bar{a}\tau + c, & ca - \bar{c}\,\bar{a} = 0,\end{cases} \quad (3.15)$$

$$\begin{cases}S = a\sigma + b,\\ T = (a\bar{b} - \bar{a}b)^{-1}(\bar{a} + \bar{c}\exp[(\bar{a}b - a\bar{b})\tau]),\\ |c|^2 = |a|^2,\end{cases} \quad (3.16)$$

$$\begin{cases}S = (\bar{a}b - a\bar{b})^{-1}(a + c\exp[(a\bar{b} - \bar{a}b)\sigma]),\\ T = \bar{a}\tau + \bar{b},\\ |c|^2 = |a|^2,\end{cases} \quad (3.17)$$

$$\begin{cases}S = \frac{b}{a}\exp[-a\sigma], \quad a+\bar{a} = 0,\\ T = c|c|^2 a\exp[-(a|c|^2)^{-1}\tau],\\ |c|^2 = |b|^2,\end{cases} \quad (3.18)$$

with a,b, and c arbitrary complex constants subjected to the indicated conditions.

Only in the first case, with a convenient choice of the constants, one gets a localized solution.

Therefore, in the following, we choose

$$\tilde{\rho}(l,0) = \rho[\delta(l+i\mu) + \delta(l-i\bar{\mu})],$$
$$\tilde{\eta}(l,0) = \eta[\delta(l+i\lambda) + \delta(l-i\bar{\lambda})], \quad (3.19)$$

with η and ρ arbitrary complex constants.

Note that, in this case, both $R(k,l)$ and $\psi(l)$ are distributions in l and one has to take into account that only their product is unambiguously defined via Eq. (2.7).

We get for the one-soliton solution

$$q = 2\lambda_R \eta \exp[i\varphi]/D, \tag{3.20}$$
$$r = 2\mu_R \rho \exp[-i\varphi]/D,$$
$$D = \gamma \exp(\xi_I) + \gamma \exp(-\xi_I) + (1+\gamma)\exp(\xi_2) + \gamma \exp(-\xi_2),$$
$$\gamma = \tfrac{1}{4}\eta\rho,$$
$$\varphi = (\mu_I + \lambda_I)x + (\mu_I - \lambda_I)y - [\omega_I(-i\lambda) + \omega_I(-i\mu)]t,$$
$$\xi_1 = (\lambda_R + \mu_R)x - (\lambda_R - \mu_R)y - [\omega_R(-i\lambda) + \omega_R(-i\mu)]t,$$
$$\xi_2 = (\lambda_R - \mu_R)x - (\lambda_R + \mu_R)y - [\omega_R(-i\lambda) - \omega_R(-i\mu)]t.$$

For $\lambda_R \mu_R \neq 0$ and $\gamma(1+\gamma) > 0$, the above formula defines (up to the phase factor $\exp[\pm i\varphi]$ in the numerator) a two-dimensional bell-shaped solution, exponentially decreasing in all directions of the (x,y) plane, moving without deformation with velocity $\mathbf{v} = (v_x, v_y)$

$$v_x = (2\lambda_R\mu_R)^{-1}[\mu_R\omega_R(-i\lambda) + \lambda_R\omega_R(-i\mu)],$$
$$v_y = (2\lambda_R\mu_R)^{-1}[-\mu_R\omega_R(-i\lambda) + \lambda_R\omega_R(-i\mu)]. \tag{3.21}$$

In general, the soliton is the envelope of the plane wave $\exp[\pm i\varphi]$. The initial position of the soliton can be moved arbitrarily by the translation

$$x \to x - x_0, \quad y \to y - y_0, \quad x_0, y_0 \in \mathbb{R}. \tag{3.22}$$

It is worth noting that the eigenfunction $\nu = \psi \exp[-ik(\sigma_3 x - y)]$ corresponding to the soliton solution in (3.20) satisfies the integral equations (2.4) and (2.5) with the following values of α_1 and α_2 in ν_0

$$\alpha_1 = 1 - \frac{2i\lambda_R\gamma}{k+i\lambda}\frac{1}{\gamma + (\theta(\mu_R)+\gamma)\exp[\xi_1+\xi_2]},$$
$$\alpha_2 = 1 - \frac{2i\mu_R\gamma}{k+i\mu}\frac{1}{\gamma + (\theta(-\lambda_R)+\gamma)\exp[\xi_2-\xi_1]} \tag{3.23}$$

(θ is the step function).

The DSI soliton is obtained for $\omega(k) = ik^2$, while the 2DMKdVI soliton is obtained for $\omega(k) = ik^3$. In both cases with a special choice of the parameters λ and μ, the soliton can have zero velocity or zero phase factor. However, velocity and frequency cannot be both zero.

The auxiliary functions A for the DSI equation and A,B for the 2DMKdVI equation can be uniquely determined by using the so-called t component of the Bäcklund transformation, which can be written by using the Bäcklund gauge in (3.8) as follows:

$$T_2(Q)B(Q,0;\lambda,\mu) - B(Q,0;\lambda,\mu)T_2(0) = 0. \tag{3.24}$$

By equating to zero the coefficients of the powers of the differential operator ∂_y, one gets for the DSI equation

$$A = \tfrac{1}{2}(\partial_x - \sigma_3\partial_y)\partial_y \log Q^2, \tag{3.25}$$

and for the 2DMKdVI equation

$$A = \tfrac{1}{2}(\partial_x - \sigma_3\partial_y)\partial_y \log Q^2, \tag{3.26}$$
$$B = 0.$$

The auxiliary functions A can be obtained also by applying to Q^2 the following inverse operator $(\partial_x + \sigma_3\partial_y)^{-1}$:

$$(\partial_x + \partial_y)^{-1} f(x,y) = \int_{-\infty}^{x} dx' f(x', y - x + x') + 4\lambda_R\gamma/\{\gamma + (\theta(\mu_R)+\gamma)\exp[\xi_1+\xi_2]\},$$
$$(\partial_x - \partial_y)^{-1} f(x,y) = \int_{-\infty}^{x} dx' f(x', y + x - x') + 4\mu_R\gamma/\{\gamma + (\theta(-\lambda_R)+\gamma)\exp[\xi_2-\xi_1]\}. \tag{3.27}$$

and then by deriving with respect to y.

IV. SPECTRAL TRANSFORM OF THE SOLITON SOLUTION

The spectral transform $R(k,l,0)$ of Q for the choice ν_0 in (3.23) is computed in the previous section. It is very instructive to compute the eigenfunction, say $\psi^{(+)}$, solution of Eq. (2.8) for the choice $\nu_0 = \mathbb{1}$ of Fokas and Ablowitz.[14] The eigenfunction $\psi^{(+)}$ can be obtained by applying the Bäcklund gauge $B(Q,0;\lambda,\mu)$ in (3.8) computed at $t = 0$ to the following eigenfunction of the spectral problem (2.1) with $Q \equiv 0$

$$\begin{pmatrix} \beta_1(x-y) & 0 \\ 0 & \beta_2(x+y) \end{pmatrix} \exp[ik(\sigma_3 x - y)], \tag{4.1}$$

where the two arbitrary functions β_i have to be determined in such a way that the diagonal elements of the matrix $\nu = \psi \exp[-ik(\sigma_3 x - y)]$ satisfy the following requirements:

$\nu_{11} \to 1$ as $x \to -\infty$ for $x - y$ const,

$\nu_{22} \to 1$ as $x \to -\infty$ for $x + y$ const,

$\nu_{11}, \nu_{22} \to 1$ as $k \to \infty$.

Let us, first, consider the case $\lambda_R > 0$, $\mu_R > 0$. The functions β_i must satisfy the following ordinary differential equations

$$\lambda_R \tanh[\lambda_R(x-y) + \delta]\beta_1 - \beta_1' - i(k - \lambda_I)\beta_1 = 1,$$
$$-\mu_R \tanh[\mu_R(x+y)]\beta_2 + \beta_2' - i(k - \mu_I)\beta_2 = 1, \tag{4.2}$$

with the boundary conditions

$$\beta_1 \to 0 \text{ as } k_I(x-y) \to +\infty, \qquad \beta_2 \to 0 \text{ as } k_I(x+y) \to -\infty. \tag{4.3}$$

The constant phase δ is defined by

$$(1+\gamma)/\gamma = \exp(2\delta). \tag{4.4}$$

The solutions β_i can be written in terms of the hypergeometric function

$$\beta_1 = \frac{i}{k+i\lambda} \times F\left(1,1; -\frac{i}{2\lambda_R}(k+i\lambda) + 1; \frac{1}{1+\exp[\xi_1+\xi_2+2\delta]}\right),$$
$$\beta_2 = \frac{i}{k+i\mu} \times F\left(1,1; -\frac{i}{2\mu_R}(k+i\mu) + 1; \frac{1}{1+\exp[\xi_2-\xi_1]}\right), \tag{4.5}$$

or by using the following integral representations

$$\beta_1 = -\cosh[\lambda_R(x-y)+\delta] \times \int_{+k_I\infty}^{x-y} du \frac{\exp[-i(k-\lambda_I)(x-y-u)]}{\cosh[\lambda_R u+\delta]}$$

$$\beta_2 = \cosh[\mu_R(x+y)] \times \int_{-k_I\infty}^{x+y} du \frac{\exp[i(k-\mu_I)(x+y-u)]}{\cosh[\mu_R u]}. \quad (4.6)$$

Therefore, the eigenfunction $\psi^{(+)}$ relative to the soliton and satisfying the integral equations (2.4) and (2.5) with $\nu_0 = 1$ can be written as follows:

$$\psi^{(+)} = \left\{ 1 - \frac{1}{2}\sigma_3 Q \begin{pmatrix} \beta_1 & 0 \\ 0 & \beta_2 \end{pmatrix} + \frac{2}{D} \begin{pmatrix} -\dfrac{\lambda_R \beta_1}{\exp(-\xi_1)+\exp(\xi_2+2\delta)} & 0 \\ 0 & \dfrac{\mu_R \beta_2}{\exp(-\xi_1)+\exp(-\xi_2)} \end{pmatrix} \right\} \times \exp[ik(\sigma_3 x - y)]. \quad (4.7)$$

It has an infinite sequence of simple poles located in the lower half complex k plane and precisely for the first column of ψ at

$$\lambda_n = -i\lambda - 2i\lambda_R n, \quad n = 0,1,..., \quad (4.8)$$

and for the second column at

$$\mu_n = -i\mu - 2i\mu_R n, \quad n = 0,1,.... \quad (4.9)$$

For the other possible signs of λ_R and μ_R one can easily derive analogous formulas. In the general case, the poles are always located in lower-half k plane and precisely for the first column of ψ at

$$\lambda_n = \lambda_I - (2n+1)i|\lambda_R|, \quad n = 0,1,..., \quad (4.10)$$

and for the second column at

$$\mu_n = \mu_I - (2n+1)i|\mu_R| \quad n = 0,1,.... \quad (4.11)$$

A similar calculation gives $\psi^{(-)}$ which is analytic in the lower-half plane and has a continuation in the upper-half plane with an infinite set of simple poles.

In conclusion, choosing $\nu_0 = 1$ in (2.8) and (2.9) leads to a representation of the soliton solution in terms of a continuous spectrum and not in terms of a finite set of discrete eigenvalues (as in 1 + 1 dimension).

V. ELLIPTIC SYSTEMS IN THE PLANE

The elliptic nonlinear evolution equations DSII (Ref. 10) and 2DMKdVII (Refs. 11 and 12) can be derived from the corresponding ones DSI and 2DMKdVI via the change of variable

$$y \to -iy. \quad (5.1)$$

If one supposes that the possible soliton solutions can be obtained as the Bäcklund transform of the trivial zero solution, they have at $t = 0$ the form

$$q(x,y,0) = -\frac{(d/dz)[\exp(-\lambda z)\eta(z)]\exp[\mu\bar{z}+\lambda z]}{1+(1/4)\rho(\bar{z})\exp(\mu\bar{z})\eta(z)\exp(-\lambda z)},$$

$$r(x,y,0) = \frac{(d/d\bar{z})[\exp(\mu\bar{z})\rho(\bar{z})]\exp[-\mu\bar{z}-\lambda z]}{1+(1/4)\rho(\bar{z})\exp(\mu\bar{z})\eta(z)\exp(-\lambda z)},$$

$$z = x+iy, \quad (5.2)$$

with $\rho(\bar{z})$ and $\eta(z)$ some convenient analytic functions in $\bar{z} = x - iy$ and in $z = x + iy$, respectively. If $\exp(\mu\bar{z})\rho(\bar{z})$ and $\exp(-\lambda z)\eta(z)$ are rational functions, we get (for $\lambda = -\mu$) the lump solutions. If, on the contrary, $\exp(\mu\bar{z})\rho(\bar{z})$ and $\exp(-\lambda z)\eta(z)$ have an essential singularity located at infinity in order to get a regular solution the fields q and r cannot decay at infinity in all directions and the solution is not localized. We conclude that the previous result of Fokas and Ablowitz[14] about the existence of lumps in the elliptic case is confirmed by our analysis, which makes use of the Bäcklund gauge.

Note added in proof: Since the submission of this paper, some developments of the theory will be found in Refs. 16–18.

ACKNOWLEDGMENTS

This work was supported in part by MPI and INFN, Sezione di Lecce, Italy. The Laboratoire de Physique Mathematique is a Unité de Recherche Associée au CNRS UA 040768 and Recherche Cooperative sur Programme No. 080264.

[1] S. V. Manakov, Physica D **3**, 420 (1981).
[2] M. J. Ablowitz, D. Bar Yaakov, and A. S. Fokas, Stud. Appl. Math. **69**, 135 (1983).
[3] A. S. Fokas and M. J. Ablowitz, Stud. Appl. Math. **69**, 211 (1983).
[4] R. Beals and R. R. Coifman, Proc. Symp. Pure Math. **43**, 45 (1985).
[5] M. Boiti, J. J. -P. Léon, M. Manna, and F. Pempinelli, Inverse Prob. **2**, 271 (1986); **3**, 25 (1987).
[6] M. Boiti, J. J. -P. Léon, and F. Pempinelli, Inverse Prob. **3**, 37 (1987); **3**, 371 (1987).
[7] P. M. Santini and A. S. Fokas, Comm. Math. Phys. **115**, 375 (1988).
[8] A. S. Fokas and P. M. Santini, Comm. Math. Phys. **116**, 449 (1988).
[9] M. Boiti, J. J. -P. Léon, and F. Pempinelli, Studies Appl. Math. **78**, 1 (1988).
[10] A. Davey and K. Stewartson, Proc. R. Soc. London Ser. A **338**, 101 (1974).
[11] M. Boiti, J. J. -P. Léon, L. Martina, and F. Pempinelli, Phys. Lett. A **132**, 432 (1988).
[12] M. Boiti, J. J. -P. Léon, L. Martina, and F. Pempinelli, "Localized solitons in the plane," in *Nonlinear Evolution Equations, Integrability and Spectral Methods*, edited by A. Degasperis and A. Fordy (Manchester U. P., Manchester, England, 1990).
[13] M. Boiti, B. G. Konopelchenko, and F. Pempinelli, Inverse Prob. **1**, 33 (1985).
[14] A. S. Fokas, Phys. Rev. Lett. **51**, 3 (1983); A. S. Fokas and M. J.

Ablowitz, J. Math. Phys. **25**, 2494 (1984).
[15] M. Boiti, J. J. -P. Léon, L. Martina, and F. Pempinelli, J. Phys. A **21**, 3611 (1988).
[16] M. Boiti, J. Leon, and F. Pempinelli, "A new spectral transform for the Davey–Stewartson I equation," Phys. Lett. A **141**, 101 (1989).
[17] M. Boiti, J. Leon, and F. Pempinelli, "Spectral transform and orthogonality relations for the Kadomtsev–Petviashvili I equation," Phys. Lett. A **141**, 96 (1989).
[18] A. S. Fokas, and P. M. Santini, "Coherent structures in multidimensions," Phys. Rev. Lett. **63**, 1329 (1989).

Multiple scattering of elastic waves by a distribution of identical spheres

D. D. Phanord and N. E. Berger
Department of Mathematical Sciences, University of Alabama in Huntsville, Huntsville, Alabama 35859 and Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, Illinois 60680

(Received 12 December 1989; accepted for publication 11 July 1990)

Self-consistent integral equations for multiple scattering amplitudes and dispersion relations for coherent propagation of elastic waves, are obtained from general reciprocity, two space scatterer formalism and by averaging the functional equations relating the multiple and single scattering amplitudes of the spheres. Approximations for bulk wave propagation numbers are derived.

I. INTRODUCTION

We consider the multiple scattering of plane harmonic elastic waves by an unbounded distribution of identical spheres embedded in an infinitely extended host medium.

In these two media, solutions of the Navier's equation depend on time and space. Factoring out the harmonic time dependence, we consider only the solutions of the time-independent linearized equation of dynamic elasticity (reduced wave equations) with no body forces for two different values of the Lamé's constants and mass densities.

At the interface of the two regions, various boundary conditions are possible which depend on the elastic nature of the embedded spheres. The total field inside the distribution of the identical spheres must be regular. The total outside field that is the sum of the incident wave and the scatterred field must also be regular. In general, the scattered field that consists of longitudinal and transverse waves must satisfy the radiation conditions at infinity.[1]

Following the techniques of Twersky,[2] we derive a system of integral equations that permit us to express the far-field multiple scattering amplitudes in terms of the single responses of the individual scatterers in isolation.

The single scatterer in isolation has been treated by many authors. We mention in particular Ying and Truell;[3] Einspruch *et al.*[4], Einspruch and Truell,[5] and Pao and Mow.[6] These authors considered the case of the elastic inclusion, including the rigid sphere, the sperical cavity, and the fluid-filled cavity.

In order to represent the scattering amplitudes, we use techniques developed by Twersky[2] and extended to elasticity by Dassios[7] to obtain two pairs of vector single scattering amplitudes ${}_p\mathbf{g}_p$, ${}_p\mathbf{g}_s$ for a longitudinal or p incident wave and ${}_s\mathbf{g}_p$, ${}_s\mathbf{g}_s$ for a transverse or s incident wave.

The scattering amplitudes are given in terms of surface integrals that arise from applying the identities of Betti to elasticity.[1] Inside the surface integrals, are the net scattered waves ${}_p\mathbf{u}$, ${}_s\mathbf{u}$ corresponding to the single obstacle evaluated at its boundary or, the total outside fields ${}_p\psi$, ${}_s\psi$ evaluated at the boundary of the scatterer in isolation. These results are compared to Barratt.[8] In the surface integral of Barratt, only the longitudinal or transverse part of the scattered field appears.

For approximation purposes, we extend to elasticity the two space scatterer formalism of Twersky.[9-14] Because of mode conversion at the boundary of the single obstacle, it is necessary to consider excitation of the scatterer by longitudinal and transverse coherent waves and require the coherent waves to radiate into free space.

The multiple scattering case is divided into two parts. First, we consider a configuration of N-identical spheres and use the self-consistent approach of Twersky[11] to write the total outside field.

Using this field and two arbitrary solutions of the single scatterer in isolation, we derive four self-consistent integral equations for the scattering amplitudes ${}_p\mathbf{G}_{t:p}$, ${}_p\mathbf{G}_{t:s}$, and ${}_s\mathbf{G}_{t:p}$, ${}_s\mathbf{G}_{t:s}$ in terms of the known scattering amplitudes of the single object located by $\mathbf{r}_t$. Our self-consistent integral equations reduce to the acoustic results of Twersky[12,14] whenever the medium cannot sustain transverse waves.

Next, we consider the ensemble average over all configurations to obtain the total outside field as in Foldy[15] extended by Twersky.[12] From the ensemble average of the self-consistent integral equations, we derive dispersion relations for the bulk wave propagation numbers κ_p, κ_s. The dispersion relations contain the radiative functions ${}_\alpha u_{jp}$, ${}_\beta u_{js}$. Here, α, and β represent the sum of the responses due to either p or s incidence, respectively. To recast the dispersion relations into forms ready for applications, we introduce the associated radiative amplitudes ${}_\alpha F_p$, ${}_\beta F_s$ whose analogs are given in Twersky.[13,14]

From the associated dispersion relations, we give the Rayleigh approximation and the two space scatterer formalism approximation which in the sparce scatterer case reduces to the forward scattering results of Devaney.[16]

In general, we work with arbitrary scatterers and specialize to spherical geometry. We use boldface to indicate vectors. The hat on the top of a vector indicates that the vector has unit magnitude. The tilde on the top of a capital letter denotes a dyadic (second rank tensor). For brevity, we use (5:2) for Eq. (2) of Ref. 5.

II. DEFINITIONS AND REPRESENTATION

We consider the scattering of an incident plane harmonic elastic wave ϕ propagating in a direction $\hat{\mathbf{k}}$ by a spherical object embedded in an isotropic and homogeneous elastic host medium specified by Lamé's constants λ, μ and density ρ different from λ', μ', and ρ' of the obstacle.

The problem is to determine the total scattered field due

 0022-2488/90/112619-07$03.00

to the incident wave. The incident elastic wave can be either a longitudinal wave denoted as an p incident wave and given by ${}_p\phi$ or a transverse wave denoted as an incident wave and given by ${}_s\phi$. The spherical object can be either a perfect elastic sphere, a rigid sphere, or spherical cavity.

Let ν be the volume occupied by the scatterer and V be the volume outside of the scatterer. The surface of the spherical object is S, its radius is $\mathbf{a}$, and it is centered at $\mathbf{r}=\mathbf{0}$. The p incident wave is parallel to the propagation vector $\hat{\mathbf{k}}$ while the s incident wave lies along the polarization vector $\hat{\mathbf{e}}_s$ normal to the direction of propagation

$$\hat{\mathbf{e}}_s\cdot\hat{\mathbf{k}}=0. \tag{1}$$

The wave propagation numbers k_p, k_s satisfy

$$\omega=c_pk_p=c_sk_s, \tag{2}$$

where $c_p^2=(\lambda+2\mu)/\rho$, $c_s^2=\mu/\rho$ are the longitudinal and transverse phase velocities squared.

We suppress the time-harmonic dependence $e^{-i\omega t}$ and define the p and s incident waves as

$${}_p\phi(\mathbf{r})=\hat{\mathbf{k}}e^{ik_p\hat{\mathbf{k}}\cdot\mathbf{r}},\quad {}_s\phi(\mathbf{r})=\hat{\mathbf{e}}_se^{ik_s\hat{\mathbf{k}}\cdot\mathbf{r}}. \tag{3}$$

Hence, all quantities in the entire paper have at most a spatial dependence. A factor of $e^{-i\omega t}$ is assumed in the final answer.

In the region exterior to the scatterer, the total field due to an incoming x wave (here x stands for p or s wave)

$${}_x\psi={}_x\phi+{}_x\mathbf{u}, \tag{4}$$

in the absence of body forces satisfies the time-independent linearized equation of dynamic elasticity

$$[c_p^2\nabla(\nabla\cdot)-c_s^2\nabla\times(\nabla\times)+\omega^2]({}_x\psi)=\mathbf{0}. \tag{5}$$

In the expression of (4), the radiative function ${}_x\mathbf{u}$, which does not necessarily consist of p or s wave fields along, are given by (7:23) or (21:28)

$$\mathbf{u}(\mathbf{r})=\{\tilde{\mathbf{\Gamma}},{}_x\mathbf{u}\} \tag{6}$$

with

$$\{\tilde{\mathbf{\Gamma}},{}_x\mathbf{u}\}\equiv-\frac{1}{4\pi}\int_S[\tilde{\mathbf{\Gamma}}\cdot\mathbf{T}_{\mathbf{r}'_x}\mathbf{u}-{}_x\mathbf{u}\cdot\mathbf{T}_{\mathbf{r}'}\tilde{\mathbf{\Gamma}}]dS(\mathbf{r}'),$$

and

$$\mathbf{T}_{\mathbf{r}}\equiv2\mu\hat{\mathbf{n}}\cdot\nabla_{\mathbf{r}}+\lambda\hat{\mathbf{n}}\nabla_{\mathbf{r}}\cdot+\mu\hat{\mathbf{n}}\times(\nabla\times), \tag{7}$$

is the surface stress operator and $\hat{\mathbf{n}}$ is the exterior unit normal on S. The operator of (6) is a surface integral that arises in the identities of Betti for elasticity. Here $\tilde{\mathbf{\Gamma}}(\mathbf{r},\mathbf{r}')$ is the fundamental dyadic solution of

$$\mathbf{L}\{\tilde{\mathbf{\Gamma}}(\mathbf{r},\mathbf{r}')\}=-4\pi\delta(\mathbf{r}-\mathbf{r}')\,\tilde{\mathbf{I}}, \tag{8}$$

with

$$\mathbf{L}=c_s^2\nabla_r^2+(c_p^2-c_s^2)\nabla_{\mathbf{r}'}(\nabla_{\mathbf{r}'}\cdot)+\omega^2 \tag{9}$$

and $\tilde{\mathbf{I}}=\hat{\mathbf{r}}\hat{\mathbf{r}}+\hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}}+\hat{\boldsymbol{\varphi}}\hat{\boldsymbol{\varphi}}$ which is the identity dyadic and $\delta(\mathbf{r}-\mathbf{r}')$ represents the Dirac measure concentrated at $\mathbf{r}$.

The net scattered fields of the single obstacle due to an incoming x wave are

$${}_x\mathbf{u}(\mathbf{r})={}_x\mathbf{u}_p(\mathbf{r})+{}_x\mathbf{u}_s(\mathbf{r}). \tag{10}$$

Asymptotically, using (7:35), (7:36) and (21:2.20–2.22) for $r\to\infty$, we have from (10)

$${}_x\mathbf{u}(\mathbf{r})\sim\left[\frac{e^{ik_pr}}{ik_pr}\right]{}_x\mathbf{g}_p(\hat{\mathbf{r}},\hat{\mathbf{k}})+\left[\frac{e^{ik_sr}}{ik_sr}\right]{}_x\mathbf{g}_s(\hat{\mathbf{r}},\hat{\mathbf{k}}), \tag{11}$$

$$x\in\{p,s\},$$

where

$${}_x\mathbf{g}_p(\hat{\mathbf{r}},\hat{\mathbf{k}})=(ik_p/4\pi\rho c_p^2)\{\hat{\mathbf{r}}\hat{\mathbf{r}}e^{-ik_p\hat{\mathbf{r}}\cdot\mathbf{r}'},[{}_x\mathbf{u}(\mathbf{r}')]\}, \tag{12}$$

and

$${}_x\mathbf{g}_s(\hat{\mathbf{r}},\hat{\mathbf{k}})=(ik_s/4\pi\rho c_s^2)\{\tilde{\mathbf{I}}_*e^{-ik_s\hat{\mathbf{r}}\cdot\mathbf{r}'},[{}_x\mathbf{u}(\mathbf{r}')]\}, \tag{13}$$

with

$$\tilde{\mathbf{I}}_*=(\tilde{\mathbf{I}}-\hat{\mathbf{r}}\hat{\mathbf{r}}). \tag{14}$$

Since the p and s incident waves are nonsingular solutions of (5), we write using the third formula of Betti[1]

$$\{\tilde{\mathbf{\Gamma}}(\mathbf{r},\mathbf{r}'),[{}_x\phi(\mathbf{r}')]\}=0. \tag{15}$$

Combining (15) with (6) gives

$${}_x\mathbf{u}(\mathbf{r})=\{\tilde{\mathbf{\Gamma}}(\mathbf{r},\mathbf{r}'),[{}_x\psi(\mathbf{r}')]\}. \tag{16}$$

This gives us for (12), and (13)

$${}_x\mathbf{g}_p(\hat{\mathbf{r}},\hat{\mathbf{k}})=(ik_p/4\pi\rho c_p^2)\{\hat{\mathbf{r}}\hat{\mathbf{r}}e^{-ik_p\hat{\mathbf{r}}\cdot\mathbf{r}'},[{}_x\psi(\mathbf{r})]\}, \tag{17}$$

and

$${}_x\mathbf{g}_s(\hat{\mathbf{r}},\hat{\mathbf{k}})=(ik_s/4\pi\rho c_s^2)\{\tilde{\mathbf{I}}_*e^{-ik_s\hat{\mathbf{r}}\cdot\mathbf{r}'},[{}_x\psi(\mathbf{r}')]\}. \tag{18}$$

The representation given by Barratt[8] is different from (17) and (18). In the forms of Barratt, only the p or s part of the scattered wave appears. In his forms, the stress tensor operator is replaced by the normal derivatives.

Similar to Refs. 8, 17–21, we write the spectral representation of the scattered waves due a single object

$${}_x\mathbf{u}(\mathbf{r})=\frac{1}{2\pi}\int_c\{e^{i\mathbf{k}_{pc}\cdot\mathbf{r}}({}_x\mathbf{g}_p)+e^{i\mathbf{k}_{sc}\cdot\mathbf{r}}({}_x\mathbf{g}_s)\}d\Omega_c,$$

$${}_x\mathbf{g}_y={}_x\mathbf{g}_y(\hat{\mathbf{r}}_c,\hat{\mathbf{k}}),\quad x,y\in\{p,s\}.$$

In (19), $\mathbf{k}_{pc}=k_p\hat{\mathbf{r}}_c$ and $\mathbf{k}_{sc}=k_s\hat{\mathbf{r}}_c$. The unit vector $\hat{\mathbf{r}}_c=\hat{\mathbf{r}}_c(\theta_c,\varphi_c)=(\cos\theta_c\sin\varphi_c)\hat{\mathbf{i}}+(\sin\theta_c\sin\varphi_c)\hat{\mathbf{j}}+(\cos\varphi_c)\hat{\mathbf{k}}$, and c is the Sommerfeld's path.[19] The contour path c is chosen to insure that $\mathrm{Im}[\hat{\mathbf{r}}_c]\cdot(\mathbf{r}-\mathbf{r}')\geqslant0$ (see Refs. 8, 18–20).

III. TWO SPACE SCATTERER FORMALISM

Twersky in Refs. 9, 10, and 22 has considered the case of single scatterer excited by a coherent wave traveling in κ space but radiating into free space or k space. The single scatterer was embedded in a synthetic macroscopic medium associated with the exciting coherent field.

In Refs. 12–14, he used the two space scatterer formalism to approximate a statistically homogeneous ensemble of configurations of N, identical and aligned scatterers whose centers are uniformly distributed in a volume V by a single equivalent obstacle. We now extend these techniques to elasticity.

We consider an elastic two space scatterer formalism. Because of mode conversion at the boundary of the single obstacle, it is necessary for us to use exciting fields that are longitudinal (p) and transverse (s). We assume the two space scatterer formalism in elasticity to consist of two distinct and linear scattering processes. They are longitudinal

and transverse due to exciting coherent fields ${}_p\phi_{\hat{\kappa}}(\mathbf{r})$, ${}_s\phi_{\hat{\kappa}}(\mathbf{r})$.

Let $\hat{\kappa}$ be the bulk propagation vector and $\hat{\epsilon}$ be the bulk polarization vector such that

$$\hat{\kappa}\cdot\hat{\epsilon}=0. \tag{20}$$

Let κ_p and κ_s be the bulk longitudinal and transverse wave propagation numbers, respectively. We represent the coherent exciting fields as

$${}_p\phi_{\hat{\kappa}}(\mathbf{r})=\hat{\kappa}e^{i\kappa_p\hat{\kappa}\cdot\mathbf{r}},\quad {}_s\phi_{\hat{\kappa}}(\mathbf{r})=\hat{\epsilon}e^{i\kappa_s\hat{\kappa}\cdot\mathbf{r}}. \tag{21}$$

The total outside fields due to the incident coherent waves are given by

$${}_x\psi_{\hat{\kappa}}(\mathbf{r})={}_x\phi_{\hat{\kappa}}(\mathbf{r})+{}_x\mathbf{u}_{\hat{\kappa}}(\mathbf{r}). \tag{22}$$

Similar to (12) and (13), as $r\to\infty$ we write the two space scatterer formalism vector scattering amplitudes as

$${}_x\mathbf{g}_p(\hat{\mathbf{r}},\hat{\kappa})=(ik_p/4\pi\rho c_p^2)\{\hat{\mathbf{r}}\hat{\mathbf{r}}e^{-ik_p\hat{\mathbf{r}}\cdot\mathbf{r}'},[{}_x\mathbf{u}_{\hat{\kappa}}(\mathbf{r}')]\}, \tag{23}$$

and

$${}_x\mathbf{g}_s(\hat{\mathbf{r}},\hat{\kappa})=(ik_s/4\pi\rho c_s^2)\{\tilde{\mathbf{I}}_\cdot e^{-ik_s\hat{\mathbf{r}}\cdot\mathbf{r}'},[{}_x\mathbf{u}_{\hat{\kappa}}(\mathbf{r}')]\}. \tag{24}$$

IV. MULTIPLE SCATTERING

We consider a fixed configuration of N-identical spheres with centers located by $\mathbf{r}_m$ $(m=1,2,\ldots,N)$ (11:12, 12:11). We represent the total multiple scattering field ${}_x\Psi$ corresponding to an incoming x wave (where x is p or s) as

$${}_x\Psi={}_x\phi+\sum_{m=1}^{N}{}_x\mathbf{U}_m(\mathbf{r}-\mathbf{r}_m). \tag{25}$$

Here, we use the obvious decomposition into p and s fields

$${}_x\mathbf{U}_m(\mathbf{r}-\mathbf{r}_m)={}_x\mathbf{U}_{m:p}+{}_x\mathbf{U}_{m:s} \tag{26}$$

for the multiple scattered wave which radiates from the scatterer fixed at $\mathbf{r}_m$.

Equivalently, with reference to the scatterer located by $\mathbf{r}_t$,

$${}_x\Psi_t(\mathbf{r})={}_x\phi(\mathbf{r})+\sum_m{}'{}_x\mathbf{U}_m+{}_x\mathbf{U}_t(\mathbf{r}-\mathbf{r}_t), \tag{27}$$

where

$${}_x\mathbf{U}_t(\mathbf{r}-\mathbf{r}_t)={}_x\mathbf{U}_{t:p}+{}_x\mathbf{U}_{t:s}, \tag{28}$$

and

$$\sum_m{}'=\sum_{m\neq t}. \tag{29}$$

Using the surface integral representation of Sec. II, the multiple scattered wave associated with the scatterer located at t is

$${}_x\mathbf{U}_t=\{\tilde{\Gamma}_p,[{}_x\mathbf{U}_t(\mathbf{r}_t+\mathbf{r}')]\}+\{\tilde{\Gamma}_s,[{}_x\mathbf{U}_t(\mathbf{r}_t+\mathbf{r}')]\}. \tag{30}$$

Asymptotically, for $r\to\infty$ (30) becomes

$${}_x\mathbf{U}_t(\mathbf{R}_t)=h(k_p|\mathbf{R}_t|)[{}_x\mathbf{G}_{t:p}]+h(k_s|\mathbf{R}_t|)[{}_x\mathbf{G}_{t:s}], \tag{31}$$

where

$$\mathbf{R}_t=\mathbf{r}-\mathbf{r}_t,\quad \mathbf{G}=\mathbf{G}(\mathbf{r}_t;\hat{\mathbf{r}}). \tag{32}$$

Here $h(k_x|\mathbf{r}-\mathbf{r}_t|)$ is the Hankel's function of the first kind; ${}_x\mathbf{G}_{t:p}$ and ${}_x\mathbf{G}_{t:s}$ are the longitudinal and transverse multiple vector scattering amplitudes of the scatterer at t due an incoming x wave.

Following the development leading to (12) and (13), the multiple vector scattering amplitudes are

$${}_x\mathbf{G}_{t:p}(\mathbf{r}_t,\hat{\mathbf{r}})=(ik_p/4\pi\rho c_p^2)\{\hat{\mathbf{r}}\hat{\mathbf{r}}e^{-ik_p\hat{\mathbf{r}}\cdot\mathbf{r}'},({}_x\mathbf{U}_t)\}, \tag{33}$$

and

$${}_x\mathbf{G}_{t:s}(\mathbf{r}_t;\hat{\mathbf{r}})=(ik_s/4\pi\rho c_s^2)\{\tilde{\mathbf{I}}_\cdot e^{-ik_s\hat{\mathbf{r}}\cdot\mathbf{r}'},({}_x\mathbf{U}_t)\}, \tag{34}$$

$${}_x\mathbf{U}_t={}_x\mathbf{U}_t(\mathbf{r}_t+r').$$

The spectral representation of the multiple scattered wave analog to (19) is

$${}_x\mathbf{U}_t(\mathbf{R}_t)=\int_c\{e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_t}({}_x\mathbf{G}_{t:pc})+e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_t}({}_x\mathbf{G}_{t:sc})\}, \tag{35}$$

where ${}_x\mathbf{G}_{t:yc}={}_x\mathbf{G}_{t:y}(\hat{\mathbf{r}}_c)$, $x,y\in\{p,s\}$, and

$$\int_c=\frac{1}{2\pi}\int_c\{\ \}d\Omega_c. \tag{36}$$

V. SELF-CONSISTENT INTEGRAL EQUATIONS OF THE CONFIGURATIONAL MULTIPLE SCATTERING AMPLITUDES

We derive two self-consistent integral equations for the multiple scattering amplitudes in terms of the known single scattering amplitudes of Sec. II for the case of an initial incidence longitudinal p wave. The development for an incidence s wave is similar. For more details see Refs. 20 and 21.

We consider for the t-scatterer two arbitrary single solutions ${}_p\psi_t(\mathbf{r}_a)$, and ${}_s\psi_t(\mathbf{r}_a)$ [where $(\mathbf{r}_a,\mathbf{r}_b)$ are arbitrary directions of propagation] and the configurational multiple solution ${}_p\Psi_{t:b}$. The requirement that they satisfy the same conditions on the surface S_t of the scatterer and its volume v_t corresponds to

$$\{{}_p\psi_t(\mathbf{r}_a),{}_p\Psi_{t:b}\}_t=0,\quad \{{}_s\psi_t(\mathbf{r}_a),{}_p\Psi_{t:b}\}_t=0. \tag{37}$$

In the first part of (37), replace ${}_p\psi_t(\mathbf{r}_a)$ and ${}_p\Psi_{t:p}$ to obtain

$$\left\{{}_p\phi_a+{}_p\mathbf{u}_{t_a},\left({}_p\phi_b+\sum_m{}'{}_p\mathbf{U}_m+{}_p\mathbf{U}_t\right)\right\}_t=0. \tag{38}$$

Using brace algebra of Refs. 20 and 21 in (38) with the vanishing of the nonsingular and the S_∞ terms gives

$$\begin{aligned}\{{}_p\phi_a,[{}_p\mathbf{U}_t(\mathbf{r}_t+\mathbf{r}')]\}_t&=\{{}_p\phi_b,({}_p\mathbf{u}_{t_a})\}_t\\&\quad+\left\{\sum_m{}'{}_p\mathbf{U}_m(\mathbf{r}-\mathbf{r}_m),({}_p\mathbf{u}_{t_a})\right\}_t.\end{aligned} \tag{39}$$

Substituting

$${}_p\phi_w=\hat{\mathbf{r}}_w\cdot(\hat{\mathbf{r}}_w\hat{\mathbf{r}}_w e^{ik_p\hat{\mathbf{r}}_w\cdot\mathbf{r}'}),\qquad w\in\{a,b\}, \tag{40}$$

$\mathbf{r}=\mathbf{r}_t+\mathbf{r}'$, $\mathbf{R}_{tm}=\mathbf{r}_t-\mathbf{r}_m$, and (35) into (39) we have

$$\begin{aligned}&\hat{\mathbf{r}}_a\cdot\{\hat{\mathbf{r}}_a\hat{\mathbf{r}}_a e^{ik_p\hat{\mathbf{r}}_a\cdot\mathbf{r}'},({}_p\mathbf{U}_t)\}_t\\&\quad=\hat{\mathbf{r}}_b\cdot\{\hat{\mathbf{r}}_b\hat{\mathbf{r}}_b e^{ik_p\hat{\mathbf{r}}_b\cdot\mathbf{r}'},({}_p\mathbf{u}_{t_a})\}_t\\&\qquad+\sum_m{}'\int_c\left[e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_{tm}}\left\{e^{i\mathbf{k}_{pc}\cdot\mathbf{r}'}\mathbf{G}_{m:p},({}_p\mathbf{u}_{t_a})\right\}_t\right.\\&\qquad\left.+e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_{tm}}\left\{e^{i\mathbf{k}_{sc}\cdot\mathbf{r}'}{}_p\mathbf{G}_{m:s},({}_p\mathbf{u}_{t_a})\right\}_t\right].\end{aligned} \tag{41}$$

Define the vector multiple scattering amplitudes as follows:

$$
\begin{aligned}
{}_p\mathbf{G}_{m:p}(\hat{\mathbf{r}}_c) &= {}_p\hat{\boldsymbol{\epsilon}}_{m:p}[{}_pG_{m:p}(\hat{\mathbf{r}}_c)], \\
{}_p\mathbf{G}_{m:s}(\hat{\mathbf{r}}_c) &= {}_p\hat{\boldsymbol{\epsilon}}_{m:s}[{}_pG_{m:s}(\hat{\mathbf{r}}_c)],
\end{aligned} \tag{42}
$$

where ${}_p\hat{\boldsymbol{\epsilon}}_{m:p}, {}_p\hat{\boldsymbol{\epsilon}}_{m:s}$ are the configurational polarization vectors.

With (12), (33), and (42), we reduce (41) to

$$
\begin{aligned}
&(c_p^2/k_p)\hat{\mathbf{r}}_a \cdot {}_p\mathbf{G}_{t:p}(-\hat{\mathbf{r}}_a) \\
&\quad = \frac{c_p^2}{k_p}\hat{\mathbf{r}}_b \cdot ({}_p\mathbf{g}_{t:p_b}) e^{i\mathbf{k}_p\cdot\mathbf{r}_t} + \frac{c_p^2}{k_p}\sum_m{}' \int_c e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_{tm}}\, {}_p\hat{\boldsymbol{\epsilon}}_{m:p} \\
&\qquad \cdot ({}_p\mathbf{g}_{t:p_c})\, {}_pG_{m:p}(\hat{\mathbf{r}}_c) + \frac{c_s^2}{k_s}\sum_m{}' \int_c e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_{tm}}\, {}_p\hat{\boldsymbol{\epsilon}}_{m:s} \\
&\qquad \cdot ({}_p\mathbf{g}_{t:s_c})\, {}_pG_{m:s}(\hat{\mathbf{r}}_c),
\end{aligned} \tag{43}
$$

$$
{}_p\mathbf{g}_{t:x_t} = {}_p\mathbf{g}_{t:x}(-\hat{\mathbf{r}}_t, \hat{\mathbf{r}}_a),
$$

$$
x \in \{p,s\}, \quad \iota \in \{b,c\}.
$$

Using the reciprocity relations

$$
\hat{\mathbf{k}}_1 \cdot {}_p\mathbf{g}_p(-\hat{\mathbf{k}}_1, \hat{\mathbf{k}}_2) = \hat{\mathbf{k}}_2 \cdot {}_p\mathbf{g}_p(-\hat{\mathbf{k}}_2, \hat{\mathbf{k}}_1), \tag{44}
$$

and

$$
\hat{\mathbf{k}}_1 \cdot {}_s\mathbf{g}_p(-\hat{\mathbf{k}}_1, \hat{\mathbf{k}}_2) = (k_s/k_p)^{-3}\hat{\boldsymbol{\epsilon}}_{s_2} \cdot {}_p\mathbf{g}_s(-\hat{\mathbf{k}}_2, \hat{\mathbf{k}}_1) \tag{45}
$$

of (21:2.47, 2.51); $-\hat{\mathbf{r}}_a = \hat{\mathbf{r}}$, and $\hat{\mathbf{r}}_b = \hat{\mathbf{k}}$ in (43) lead to the desired result

$$
\begin{aligned}
{}_pG_{t:p}(\hat{\mathbf{r}}) &= {}_pg_{t:p}(\hat{\mathbf{r}}, \hat{\mathbf{k}}) e^{i\mathbf{k}_p\cdot\mathbf{r}_t} \\
&\quad + \sum_m{}' \int_c \left[e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_{tm}} ({}_pg_{t:p}(\hat{\mathbf{r}}, \hat{\mathbf{r}}_c))\, {}_pG_{m:p}(\hat{\mathbf{r}}_c) \right. \\
&\quad \left. + e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_{tm}} ({}_sg_{t:p}(\hat{\mathbf{r}}, \hat{\mathbf{r}}_c))\, {}_pG_{m:s}(\hat{\mathbf{r}}_c) \right].
\end{aligned} \tag{46}
$$

Similarly, the second part of (37) gives the second self-consistent integral equation

$$
\begin{aligned}
{}_pG_{t:s}(\hat{\mathbf{r}}) &= {}_pg_{t:s}(\hat{\mathbf{r}}, \hat{\mathbf{k}}) e^{i\mathbf{k}_p\cdot\mathbf{r}_t} \\
&\quad + \sum_m{}' \int_c \left[e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_{tm}} ({}_pg_{t:s}(\hat{\mathbf{r}}, \hat{\mathbf{r}}_c))\, {}_pG_{m:p}(\hat{\mathbf{r}}_c) \right. \\
&\quad \left. + e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_{tm}} ({}_sg_{t:s}(\hat{\mathbf{r}}, \hat{\mathbf{r}}_c))\, {}_pG_{m:s}(\hat{\mathbf{r}}_c) \right].
\end{aligned} \tag{47}
$$

Equations (46) and (47) determine the multiple scattering amplitudes corresponding to an incoming wave ${}_p\phi$ in terms of the responses of the individual scatterers.

A similar system of two equations for the multiple scattering amplitudes ${}_sG_{t:p}$ and ${}_sG_{t:s}$ results from the consideration of an incoming ${}_s\phi$ wave.

Equation (46) reduces to the acoustic results of Twersky (12:16) whenever the medium cannot sustain transverse waves.

Barratt gave one combined equation for the multiple scattering amplitudes. The differences between the representation of Barratt and ours lie in the way the single scattering amplitudes are obtained. For Barratt, the single scattering amplitudes came directly from the longitudinal or transverse part of the scattered field. In our case, the single scattering amplitudes are computed from the net resultant scattered field.

VI. ENSEMBLE AVERAGE

The average of ${}_p\Psi(\mathbf{r})$ over a statistically homogeneous ensemble of configurations of N-identical spheres whose centers $\mathbf{r}_m$ are uniformly distributed in a volume V is given by Twersky (11:17).

$$
\langle {}_p\Psi(\mathbf{r}) \rangle = \frac{1}{V}\int_V \langle {}_p\Psi \rangle_m \, dV(\mathbf{r}_m), \tag{48}
$$

where $\langle {}_p\Psi(\mathbf{r}) \rangle$ is independent of the configurational variables $(\mathbf{r}_1, \mathbf{r}_2, \ldots, \mathbf{r}_N)$ and $\langle\ \rangle_m$, the average with $\mathbf{r}_m$ held fixed, depends only on $\mathbf{r}_m$.

Replacing ${}_p\Psi(\mathbf{r})$ by its value inside the volume integral of (48) leads to

$$
\langle {}_p\Psi(\mathbf{r}) \rangle = {}_p\phi(\mathbf{r}) + \rho_n \int_V \langle {}_p\mathbf{U}_m(\mathbf{r} - \mathbf{r}_m) \rangle_m \, dV(\mathbf{r}_m), \tag{49}
$$

where $\rho_n = N/V$ is the average number of centers per unit volume.

Decomposing the volume V into

$$
V = (V - \nu_m + \nu_m), \tag{50}
$$

where ν_m is the volume of the scatterer located by $\mathbf{r}_m$, and representing ${}_p\mathbf{U}_m(\mathbf{r} - \mathbf{r}_m)$ by

$$
{}_p\mathbf{U}_m(\mathbf{k}_p, \mathbf{k}_s) \equiv {}_p\mathbf{U}_{m:p}(\mathbf{k}_p, \mathbf{k}_s) + {}_p\mathbf{U}_{m:s}(\mathbf{k}_p, \mathbf{k}_s), \tag{51}
$$

we rewrite (49) as

$$
\begin{aligned}
\langle {}_p\Psi(\mathbf{r}) \rangle &= {}_p\phi(\mathbf{r}) + \rho_n \int_{V-\nu_m} \langle {}_p\mathbf{U}_m(\mathbf{k}_p, \mathbf{k}_s) \rangle_m \, dV(\mathbf{r}_m) \\
&\quad + \rho_n \int_{\nu_m} \langle {}_p\mathbf{U}_m(\mathbf{k}'_p, \mathbf{k}'_s) \rangle_m \, dV(\mathbf{r}_m),
\end{aligned} \tag{52}
$$

with $\langle {}_p\mathbf{U}_m(\mathbf{k}'_p, \mathbf{k}'_s) \rangle_m$ being the average scattered wave inside the volume ν_m.

We can directly average (33) and (34) to obtain

$$
\begin{aligned}
\langle {}_p\mathbf{G}_{t:p}(\mathbf{r}_t; \hat{\mathbf{r}}) \rangle_t &= \frac{ik_p}{4\pi\rho c_p^2}\{\hat{\mathbf{r}}\hat{\mathbf{r}} e^{-i\mathbf{k}_p\hat{\mathbf{r}}\cdot\mathbf{r}'}, \langle {}_p\mathbf{U}_t \rangle_t\} \\
&\equiv {}_p\mathbf{G}_p(\mathbf{r}_t; \hat{\mathbf{r}}), \\
\langle {}_p\mathbf{G}_{t:s}(\mathbf{r}_t; \hat{\mathbf{r}}) \rangle_t &= \frac{ik_s}{4\pi\rho c_s^2}\{\tilde{\mathbf{I}} \cdot e^{-i\mathbf{k}_s\hat{\mathbf{r}}\cdot\mathbf{r}'}, \langle {}_p\mathbf{U}_t \rangle_t\} \\
&\equiv {}_p\mathbf{G}_s(\mathbf{r}_t; \hat{\mathbf{r}}).
\end{aligned} \tag{53}
$$

The ensemble average vector scattering amplitudes due to an incoming s incident wave can be obtained from (53) as

$$
\begin{aligned}
\langle {}_s\mathbf{G}_{t:p}(\mathbf{r}_t; \hat{\mathbf{r}}) \rangle_t &= \frac{ik_p}{4\pi\rho c_p^2}\{\hat{\mathbf{r}}\hat{\mathbf{r}} e^{-i\mathbf{k}_p\hat{\mathbf{r}}\cdot\mathbf{r}'}, \langle {}_s\mathbf{U}_t \rangle_t\} \\
&\equiv {}_s\mathbf{G}_p(\mathbf{r}_t; \hat{\mathbf{r}}), \\
\langle {}_s\mathbf{G}_{t:s}(\mathbf{r}_t; \hat{\mathbf{r}}) \rangle_t &= \frac{ik_s}{4\pi\rho c_s^2}\{\tilde{\mathbf{I}} \cdot e^{-i\mathbf{k}_s\hat{\mathbf{r}}\cdot\mathbf{r}'}, \langle {}_s\mathbf{U}_t \rangle_t\} \\
&\equiv {}_s\mathbf{G}_s(\mathbf{r}_t; \hat{\mathbf{r}}).
\end{aligned} \tag{54}
$$

VII. DISPERSION RELATIONS FOR THE EFFECTIVE MEDIUM

We consider only the case of an initial p wave. The cases of an initial s wave as well as a general wave can be handled

similarly. To derive the dispersion relations needed to compute the bulk wave propagation numbers, we combine the ensemble average of (45), (46), and use with our identical scatterers the quasicrystalline approximation of Lax[23] to obtain (21:4.5, 24:59)

$$
\begin{aligned}
{}_{\alpha}G(\mathbf{r}_t,\hat{\mathbf{r}}) &= {}_{\alpha}g(\hat{\mathbf{r}},\hat{\mathbf{k}})e^{i\mathbf{k}_p\cdot\mathbf{r}_t} + \rho_n \int_{V-\nu} d\mathbf{r}_m f(\mathbf{R}_{tm}) \\
&\quad \cdot \left\{ \int_c e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_{tm}} ({}_{a}g_c)\,{}_{p}G_p \right. \qquad (55) \\
&\quad \left. + \int_c e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_{tm}} ({}_{\beta}g_c)\,{}_{p}G_s \right\}, \\
{}_{p}G_x &\equiv {}_{p}G_x(\mathbf{r}_m,\hat{\mathbf{r}}_c), \\
{}_{q}g_c &\equiv {}_{q}g(\hat{\mathbf{r}},\hat{\mathbf{r}}_c), \quad x\in\{p,s\}, \quad q\in\{\alpha,\beta\},
\end{aligned}
$$

and

$$
\begin{aligned}
{}_{\alpha}G(\mathbf{r}_t;\hat{\mathbf{r}}) &\equiv {}_{p}G_p(\mathbf{r}_t;\hat{\mathbf{r}}) + {}_{p}G_s(\mathbf{r}_t;\hat{\mathbf{r}}), \\
{}_{\alpha}g(\hat{\mathbf{r}},\hat{\mathbf{k}}) &\equiv {}_{p}g_p(\hat{\mathbf{r}},\hat{\mathbf{k}}) + {}_{p}g_s(\hat{\mathbf{r}},\hat{\mathbf{k}}), \qquad (56) \\
{}_{\beta}g(\hat{\mathbf{r}},\hat{\mathbf{k}}) &\equiv {}_{s}g_p(\hat{\mathbf{r}},\hat{\mathbf{k}}) + {}_{s}g_s(\hat{\mathbf{r}},\hat{\mathbf{k}}).
\end{aligned}
$$

Here $f(\mathbf{R}_{tm})$ is the radial distribution function associated with the distribution of scatterers in the region V. It is zero for $|\mathbf{R}_{tm}| < 2a$, and goes to 1 when $\mathbf{R}_{tm}$ goes to infinity.

To take advantage of the two space scatterer formalism of Sec. III, we introduce the equivalent medium approximation expansion as

$$
\begin{aligned}
{}_{p}G_x(\mathbf{r}_t;\hat{\mathbf{r}}) &= \sum_{j=1}^{2} {}_{p}g_x(\mathbf{k}_{xr}|\boldsymbol{\kappa}_{jx})e^{i\boldsymbol{\kappa}_{jx}\cdot\mathbf{r}_t}, \\
&\quad x\in\{p,s\}. \qquad (57)
\end{aligned}
$$

The internal field for the distribution corresponds in general to four waves. They are a p and an s wave with their reflected counterparts.

For the two space scatterer formalism, the vector scattering amplitudes corresponding to the equivalent medium are

$$
\begin{aligned}
{}_{p}\mathbf{g}_p(\mathbf{k}_{pr}|\boldsymbol{\kappa}_{jp}) &= \frac{ik_p}{4\pi\rho c_p^2}\{\hat{\mathbf{r}}\hat{\mathbf{r}}e^{-i\mathbf{k}_{pr}\cdot\mathbf{r}'}, [{}_{p}\mathbf{u}^j_{eq}(\mathbf{r}')]\}, \\
{}_{p}\mathbf{g}_s(\mathbf{k}_{sr}|\boldsymbol{\kappa}_{js}) &= \frac{ik_s}{4\pi\rho c_s^2}\{\bar{\mathbf{I}}_t e^{-i\mathbf{k}_{sr}\cdot\mathbf{r}'}, [{}_{p}\mathbf{u}^j_{eq}(\mathbf{r}')]\}, \qquad (58)
\end{aligned}
$$

where ${}_{p}\mathbf{u}^j_{eq}(\mathbf{r}')$ is the net scattered field of the single obstacle equivalent to the distribution of identical scatterers in a volume V bounded by a slab of thickness d.

The volume integral of (55) can be decomposed into two integrals

$$
\int_{V-\nu} f = \int_{V-\nu} 1 + \int_{V-\nu} (f-1). \qquad (59)
$$

In addition, the integral

$$
\int_{V-\nu} (1) = \int_{V_B+V_L+V_{L'}} (1), \qquad (60)
$$

where V_B is the bulk volume, V_L, $V_{L'}$ are boundary layers, and the integral

$$
\int_{V-\nu} (f-1) = \int_{V_\infty-\nu} (f-1), \qquad (61)
$$

since $f(\mathbf{R}_{tm}) \sim 1$ as $\mathbf{R}_{tm} \to \infty$.

In V_B, and $V_B - \nu$ we use G of (57) and reduce $\int_{V_B}$ to integrals over the layer surfaces ($z = l, d - l'$) and exclusion surface $S(b)$ by Green's theorem (21:4.19, 12:55–56, 13:61–62). The layer integrals combine with ${}_{\alpha}g(\hat{\mathbf{r}},\hat{\mathbf{k}})e^{i\mathbf{k}_p\cdot\mathbf{r}_t}$ to produce vanishing sums of coefficients. The dispersion relations are given by the coefficients of $e^{i\boldsymbol{\kappa}_{jp}\cdot\mathbf{r}_t}$ and $e^{i\boldsymbol{\kappa}_{js}\cdot\mathbf{r}_t}$. Hence, the dispersion relations are (12:60, 13:65)

$$
\begin{aligned}
&{}_{p}g_x(\mathbf{k}_{xr}|\boldsymbol{\kappa}_{jx}) \\
&\quad = [\rho_n/(\kappa_{jx}^2 - k_x^2)]\{[e^{-i\boldsymbol{\kappa}_{jx}\cdot\mathbf{R}}, ({}_{q}u_{jx})]\}_b \\
&\qquad + \rho_n \int_{V_\infty-\nu} d(\mathbf{R})(f(\mathbf{R})-1)e^{-i\boldsymbol{\kappa}_{jx}\cdot\mathbf{R}}({}_{q}u_{jx}), \qquad (62)
\end{aligned}
$$

$$
x = p, \quad q = \alpha; \quad \text{or } x = s, \quad q = \beta, \qquad (63)
$$

and $\mathbf{R} = \mathbf{R}_{tm}$.

Here, the radiative functions ${}_{\alpha}u_{jp}$, ${}_{\beta}u_{js}$ of the form of (19) satisfy

$$
(\nabla^2_{|\mathbf{R}|} + k_{xc}^2)\,{}_{q}u_{jx} = 0, \qquad (64)
$$

and for ($x = p, q = \alpha$) or ($x = s, q = \beta$) are given by

$$
\begin{aligned}
{}_{q}u_{jx} &= {}_{q}u_{jx}(\mathbf{k}_{pr},\mathbf{k}_{sr},\mathbf{k}_{xc}\cdot\mathbf{R};\boldsymbol{\kappa}_{jx}) \\
&\equiv \int_c e^{i\mathbf{k}_{xc}\cdot\mathbf{R}}[{}_{q}g(\hat{\mathbf{r}},\hat{\mathbf{r}}_c)]\,{}_{p}\mathbf{p}_x(\mathbf{k}_{xc}|\boldsymbol{\kappa}_{jx}). \qquad (65)
\end{aligned}
$$

The brace operator in (62) is an operator introduced by Twersky (13:63) given by

$$
\{[f,g]\}_b \equiv \int_{S(b)} [f\partial_n(g) - g\partial_n(f)]\,dS(b), \qquad (66)
$$

where

$$
S(b) = S\hat{\mathbf{n}} \qquad (67)
$$

is the exclusion surface with $\hat{\mathbf{n}}$ being the outward unit normal from $\nu(b)$.

To introduce the associated radiative amplitudes ${}_{q}F_x$, we rewrite (65) as

$$
{}_{q}u_{jx} \equiv \int_c e^{i\mathbf{k}_{xc}\cdot\mathbf{R}}\,{}_{q}F_x, \qquad (68)
$$

where

$$
\begin{aligned}
{}_{q}F_x(\mathbf{k}_{pr},\mathbf{k}_{sr},\mathbf{k}_{xc}|\boldsymbol{\kappa}_{jx}) &\equiv [{}_{q}g(\hat{\mathbf{r}},\hat{\mathbf{r}}_c)]\,{}_{p}g_x(\mathbf{k}_{xc}|\boldsymbol{\kappa}_{jx}) \\
&= \{[e^{-i\mathbf{k}_{xc}\cdot\mathbf{R}}, ({}_{q}u_{jx})]\}_b; \qquad (69)
\end{aligned}
$$

or

$$
{}_{q}F_x\{\mathbf{k}_{pr},\mathbf{k}_{xr},\boldsymbol{\kappa}_{xc}|\boldsymbol{\kappa}_{jx}\} \equiv \{[e^{-i\boldsymbol{\kappa}_{jx}\cdot\mathbf{R}}, ({}_{q}u_{jx})]\}_b. \qquad (70)
$$

We substitute (68)–(70) into (62) to obtain the associated dispersion relations (12:64, 13:66)

$$
\begin{aligned}
&{}_{p}g_p(\mathbf{k}_{pr}|\boldsymbol{\kappa}_{jp}) \\
&\quad = -[\rho_n/(\kappa_{jp}^2 - k_p^2)]\,{}_{\alpha}F_p\{\mathbf{k}_{pr},\mathbf{k}_{sr},\boldsymbol{\kappa}_{jp}|\boldsymbol{\kappa}_{jp}\} \\
&\qquad + \int_c \mathcal{M}(\mathbf{k}_{pc},\boldsymbol{\kappa}_{jp})[{}_{\alpha}F_p(\mathbf{k}_{pr},\mathbf{k}_{rs},\mathbf{k}_{pc}|\boldsymbol{\kappa}_{jp})], \qquad (71)
\end{aligned}
$$

and

$$
\begin{aligned}
&{}_pg_s(\mathbf{k}_{sr}|\boldsymbol{\kappa}_{js}) \\
&\quad = -[\rho_n/(\kappa_{js}^2 - k_s^2)]_\beta F_s\{\mathbf{k}_{pr},\mathbf{k}_{sr},\boldsymbol{\kappa}_{js}|\boldsymbol{\kappa}_{js}\} \\
&\qquad + \int_c \mathscr{M}(\mathbf{k}_{sc},\boldsymbol{\kappa}_{js})[{}_\beta F_s(\mathbf{k}_{pr},\mathbf{k}_{rs},\mathbf{k}_{sc}|\boldsymbol{\kappa}_{js})],
\end{aligned} \tag{72}
$$

with

$$
\mathscr{M}(\mathbf{k},\boldsymbol{\kappa}) = \rho_n \int_{V_\infty - v} d(\mathbf{R})(f(\mathbf{R}) - 1)e^{-(\mathbf{k}-\boldsymbol{\kappa})\cdot\mathbf{R}}. \tag{73}
$$

VIII. RAYLEIGH APPROXIMATION

We approximate the associated radiative amplitudes of (69) and (70) for $(x = p,q = \alpha)$ or $(x = s,q = \beta)$ as

$$
\begin{aligned}
{}_qF_x\{\ \} &= [{}_qg(\hat{\mathbf{r}},\hat{\mathbf{r}})]_p g_x(\mathbf{k}_{xc}|\boldsymbol{\kappa}_{jx}) \\
&\quad + \{[(e^{-i\boldsymbol{\kappa}_{jx}\cdot\mathbf{R}} - e^{-i\mathbf{k}_{xc}\cdot\mathbf{R}}),({}_qu_{jx})]\}_b.
\end{aligned} \tag{74}
$$

For $\hat{\mathbf{r}} = \hat{\mathbf{r}}_c = \hat{\mathbf{k}}$, we have

$$
{}_qF_x\{\mathbf{k}_{pr},\mathbf{k}_{sr},\boldsymbol{\kappa}_{jx}|\boldsymbol{\kappa}_{jx}\} \simeq {}_qg(\hat{\mathbf{k}},\hat{\mathbf{k}})_p g_x(\mathbf{k}_{xr}|\boldsymbol{\kappa}_{jx}). \tag{75}
$$

We substitute (75) into (71) and (72), and solve for the bulk wave propagation numbers with $\rho_n \simeq 0$, $\mathscr{M}(\mathbf{k},\boldsymbol{\kappa}) = 0$ (Ref. 24) to obtain (12:65, 13:69)

$$
\kappa_p^2 - k_p^2 \simeq -\rho_n[{}_pg_p(\hat{\mathbf{k}},\hat{\mathbf{k}})], \tag{76}
$$

and

$$
\kappa_s^2 - k_s^2 \simeq -\rho_n[{}_sg_s(\hat{\mathbf{k}},\hat{\mathbf{k}})]. \tag{77}
$$

The forward scattering amplitudes of (76) and (77) for a rigid sphere of radius a are given by Barratt and Collins .[25] We use their results in (76) and (77) and

$$
\begin{aligned}
\kappa_p^2 - k_p^2 &\simeq -\rho_n \frac{3}{2} a\left(\frac{k_p}{k_s}\right)^2\left(1 + \frac{k_p^2}{2k_s^2}\right)^{-1} \\
&\quad \times\left[1 + ik_s a\left(1 + \frac{k_p^3}{2k_s^3}\right)\left(1 + \frac{k_p^2}{2k_s^2}\right)^{-1}\right],
\end{aligned} \tag{78}
$$

and

$$
\begin{aligned}
\kappa_s^2 - k_s^2 &\simeq -\rho_n \tfrac{3}{2} a(1 + k_p^2/2k_s^2)^{-1} \\
&\quad \times[1 - ik_s a(1 + k_p^3/2k_s^3)(1 + k_p^2/2k_s^2)^{-1}].
\end{aligned} \tag{79}
$$

IX. TWO SPACE SCATTERER FORMALISM APPROXIMATION

We rewrite the associated radiative amplitudes of (70) as

$$
\begin{aligned}
{}_qF_x\{\ \} &\equiv [{}_qg(\hat{\mathbf{r}},\hat{\boldsymbol{\kappa}})]_p g_x(\mathbf{k}_{x\hat{\kappa}}|\boldsymbol{\kappa}_{jx}) \\
&\quad + \{[(e^{-i\boldsymbol{\kappa}_{jx}\cdot\mathbf{R}} - e^{-i\mathbf{k}_{x\hat{\kappa}}\cdot\mathbf{R}}),({}_qu_{jx})]\}_b, \\
&\quad \mathbf{k}_{p\hat{\kappa}} = k_p\hat{\boldsymbol{\kappa}}, \quad \mathbf{k}_{s\hat{\kappa}} = k_s\hat{\boldsymbol{\kappa}}.
\end{aligned} \tag{80}
$$

Taking the leading term in product of scattering amplitudes, we reduce (80) with $\hat{\mathbf{r}} = \hat{\boldsymbol{\kappa}}$ and $(x = p,q = \alpha)$ or $(x = s,q = \beta)$ to

$$
{}_qF_x\{\mathbf{k}_{pr},\mathbf{k}_{sr},\boldsymbol{\kappa}_{jx}|\boldsymbol{\kappa}_{jx}\} \simeq [{}_qg(\hat{\boldsymbol{\kappa}},\hat{\boldsymbol{\kappa}})]_p g_x(\mathbf{k}_{x\hat{\kappa}}|\boldsymbol{\kappa}_{jx}). \tag{81}
$$

Substituting (81) into (71) and (72), retaining only the leading term in product of scattering amplitudes, and solving for the bulk wave propagation numbers, we obtain (12:68)

$$
\kappa_p^2 - k_p^2 \simeq -(\rho_n)_\alpha \mathscr{G}(\mathbf{k}_{pc},\mathbf{k}_{p\hat{\kappa}}|\boldsymbol{\kappa}_{jp})[{}_\alpha g(\hat{\boldsymbol{\kappa}},\hat{\boldsymbol{\kappa}})] \tag{82}
$$

and

$$
\kappa_s^2 - k_s^2 \simeq -(\rho_n)_\beta \mathscr{G}(\mathbf{k}_{sc},\mathbf{k}_{s\hat{\kappa}}|\boldsymbol{\kappa}_{js})[{}_\beta g(\hat{\boldsymbol{\kappa}},\hat{\boldsymbol{\kappa}})]. \tag{83}
$$

Here, for $(x = p,q = \alpha)$ or $(x = s,q = \beta)$

$$
{}_q\mathscr{G}(\mathbf{k}_{xc},\mathbf{k}_{x\hat{\kappa}}|\boldsymbol{\kappa}_{jx}) \equiv [1 - \langle {}_q\mathscr{M}\rangle/{}_pg_x(\mathbf{k}_{x\hat{\kappa}}|\boldsymbol{\kappa}_{jx})]^{-1} \tag{84}
$$

and

$$
\langle {}_q\mathscr{M}\rangle \equiv \int_c \mathscr{M}(\mathbf{k}_{xc},\boldsymbol{\kappa}_{jx})[{}_qg(\hat{\boldsymbol{\kappa}},\hat{\mathbf{r}}_c)]_p g_x(\mathbf{k}_{xc}|\boldsymbol{\kappa}_{jx}). \tag{85}
$$

Using the approximations

$$
{}_pg_p(\mathbf{k}_{p\hat{\kappa}}|\boldsymbol{\kappa}_{jp}) \simeq {}_pg_p(\hat{\mathbf{r}},\hat{\boldsymbol{\kappa}}) + {}_pg_s(\hat{\mathbf{r}},\hat{\boldsymbol{\kappa}}) \tag{86}
$$

and

$$
{}_pg_s(\mathbf{k}_{s\hat{\kappa}}|\boldsymbol{\kappa}_{js}) \simeq {}_sg_p(\hat{\mathbf{r}},\hat{\boldsymbol{\kappa}}) + {}_sg_s(\hat{\mathbf{r}},\hat{\boldsymbol{\kappa}}) \tag{87}
$$

of Sec. III, and keeping only the leading term in (84), with $\hat{\mathbf{r}} = \hat{\boldsymbol{\kappa}}$, we obtain the two space scatterer formalism approximation for the ${}_q\mathscr{G}$ operator of (82) and (83). Hence, (84) for $(x = p,q = \alpha)$ or $(x = s,q = \beta)$ using the notation of (36) becomes

$$
{}_q\mathscr{G}(\mathbf{k}_{xc},\mathbf{k}_{xr}|\boldsymbol{\kappa}_{jx}) \simeq 1 + \langle {}_q\mathscr{M}_x\rangle/{}_qg(\hat{\boldsymbol{\kappa}},\hat{\boldsymbol{\kappa}}) \tag{88}
$$

and

$$
\langle {}_q\mathscr{M}_x\rangle \equiv \int_c \mathscr{M}(\mathbf{k}_{xc},\boldsymbol{\kappa}_{jx})_q g(\hat{\boldsymbol{\kappa}},\hat{\mathbf{r}}_c)_q g_x(\hat{\mathbf{r}}_c,\hat{\boldsymbol{\kappa}}). \tag{89}
$$

Substituting (88) and (89) into (82) and (83), we have the desired results

$$
\kappa_p^2 - k_p^2 \simeq -(\rho_n)_\alpha g(\hat{\boldsymbol{\kappa}},\hat{\boldsymbol{\kappa}}) + \langle {}_\alpha\mathscr{M}_p\rangle \tag{90}
$$

and

$$
\kappa_s^2 - k_s^2 \simeq -(\rho_n)_\beta g(\hat{\boldsymbol{\kappa}},\hat{\boldsymbol{\kappa}}) + \langle {}_\beta\mathscr{M}_s\rangle. \tag{91}
$$

When $\langle {}_q\mathscr{M}_x\rangle = 0$, Eqs. (90) and (91) correspond to the forward scattering results of Devaney (16:4.13a, 4.13b). The two space scatterer formalism scattering amplitudes are given in Appendix B of Ref. 21.

Equations (62), (71), and (72) can be specialized to specific geometry or extended to cover the case of an inhomogeneous and anisotropic elastic host medium. From the self-consistent integral equations (46) and (47), we can obtain asymptotic expansions for the configurational scattering amplitudes similar to Barratt and Twersky.

[1] V. D. Kupradze, *Progress in Solid Mechanics. III. Dynamical Problems in Elasticity* (North-Holland, Amsterdam, 1963).

[2] V. Twersky, "Scattering of waves by two objects," in *Electromagnetic Waves*, edited by R. E. Langer (Univ. of Wisconsin, Madison, 1962), pp. 361–69.

[3] C. F. Ying and R. Truell, "Scattering of a plane longitudinal wave by a spherical obstacle in an isotropically elastic solid," J. Appl. Phys. **27**(9), 1086 (1956).

[4] N. G. Einspruch, E. J. Witterholt, and R. Truell, "Scattering of a plane transverse wave by a spherical obstacle in an elastic medium," J. Appl. Phys. **31**(5), 806 (1960).

[5] N. G. Einspruch and R. Truell, "Scattering of a plane longitudinal wave by a spherical fluid obstacle in an elastic medium," J. Acoust. Soc. Am. **32**, 214 (1960).

[6] Y.-H. Pao and C. C. Mow, "Scattering of plane compressional waves by a spherical obstacle," J. Appl. Phys. **34**(3), 493 (1963).

[7] G. Dassios and K. Kiriaki, "The low-frequency theory of elastic wave

scattering," Quart. Appl. Math. **42**, 225 (1984).
[8]P. J. Barratt, "Multiple scattering of plane harmonic elastic waves in an infinite solid by an arbitrary configuration of obstacles," Proc. Cambridge Philos. Soc. **66**, 469 (1969).
[9]V. Twersky, "On a general class of scattering problems," J. Math. Phys. **3**, 716 (1962).
[10]V. Twersky, "On scattering of waves by random distributions. II. Two-space scatterer formalism," J. Math. Phys. **3**, 724 (1962).
[11]V. Twersky, "Propagation parameters in random distributions of scatterers," J. D'Analyse Math. **30**, 498 (1976).
[12]V. Twersky, "Coherent scalar field in pair-correlated random distributions of aligned scatterers," J. Math. Phys. **18**, 2468 (1977).
[13]V. Twersky, "Coherent electromagnetic waves in pair-correlated random distributions of aligned scatterers," J. Math. Phys. **19**, 215 (1978).
[14]V. Twersky, *Multiple Scattering of Waves by Correlated Distributions. Mathamatical Methods and Applications of Scattering Theory* (Spring-Verlag, New York, 1980).
[15]L. L. Foldy, "The multiple scattering of waves," Phys. Rev. **67**, 107 (1945).
[16]A. J. Devaney, "Multiple scattering theory for discrete, elastic, random media," J. Math. Phys. **21**, 2603 (1980).
[17]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Chap. 13, p. 1987.
[18]F. Noether, in *Theory of Functions*, edited by R. Rother, F. Ollendorf, and K. Polhausen (Technology, Cambridge, MA, 1948), pp. 167–70.
[19]A. Sommerfeld, *Partial Differential Equations in Physics* (Academic, New York, 1949).
[20]V. Twersky, "Multiple scattering of electromagnetic waves by arbitrary configurations," J. Math. Phys. **8**, 589 (1967).
[21]D. D. Phanord, "Multiple scattering of elastic waves by a distribution of identical spheres," Ph.D. thesis, University of Illinois at Chicago, Chicago (1988).
[22]V. Twersky, "On scattering of waves by random distributions. I. Free-space scatterer formalism," J. Math. Phys. **3**, 700 (1962).
[23]M. Lax, "The effective field in dense system," Phys. Rev. **88**, 621 (1952).
[24]L. Rayleigh, "On the transmission of light through an atmosphere containing small particles in suspension, and on the origin of the blue sky," Philos. Mag. **47**, 375 (1889).
[25]P. J. Barratt and W. D. Collins, "The scattering cross sections of an obstacle in an elastic solid for plane harmonic waves," Proc. Cambridge Philos. Soc. **61**, 969 (1965).

Reciprocity relations and energy theorems for multiple scattering of elastic waves

D. D. Phanord and N. E. Berger
Department of Mathematical Sciences, University of Alabama in Huntsville, Huntsville, Alabama 35899, and Department of Mathematics, Statistics, and Computer Science, University of Illinois at Chicago, Chicago, Illinois 60680

(Received 12 December 1989; accepted for publication 11 July 1990)

Reciprocity relations, energy theorems, and scattering cross sections are derived for a fixed configuration of N-identical and similarly aligned lossless scatterers whose centers are uniformly distributed in a given volume. The general results are specialized to scattering in the forward direction.

I. INTRODUCTION

Reciprocity relations, energy theorems, and scattering cross sections represent an important segment of scattering theory for acoustics, elasticity, electromagnetics, quantum mechanics, and other connected disciplines. The total scattering cross section, which is the sum of the absorption and the scattering cross sections, measures the disturbance caused by the scatterer or scatterers to the incident wave. It is defined as the ratio of the average at which energy is absorbed and scattered by the single or many scatterers to the average rate at which the energy of the incident wave crosses a unit area normal to the direction of propagation.

Barratt and Collins[1] used a method due to Jones[2] for the asymptotic evaluation of double integral to express the scattering cross section due to a single object by means of the spherical scattering amplitudes in the forward directions. Tan[3] extended the general Betti–Rayleigh reciprocity theorems to problems of elastodynamics. In Ref. 4, he used method of De Hoop[5] to establish theorems on scattering cross sections for the scatterer in isolation.

Varatharajulu,[6] employed the Helmholtz decomposition of the total displacement vector, and the techniques of Jones[7] to derive reciprocity relations and formulas for the scattering cross sections corresponding to different combinations of longitudinal or transverse incident waves. Dassios *et al.*[8] extended into elasticity the techniques developed by Twersky[9] for the individual acoustic scatterer. They provided a direct way of evaluating the scattering cross sections from the scattering amplitudes in both general and forward directions.

In this paper, we extend the results of Twersky[10] to multiple scattering by elastic bodies. We derive four reciprocity relations, energy theorems, and formulas for the scattering cross sections of a fixed configuration of N-identical and similarly aligned lossless scatterers whose centers are uniformly distributed in a give volume. We specialize these new results in the forward scattering direction.

To introduce notations and representations necessary for further development, we consider briefly the single elastic scatterer in isolation,[11–15] and a fixed configuration of N-identical and similarly aligned scatterers[10,16–20] to obtain with the self-consistent approach of Twersky[16] and total outside field.

In general, we work with arbitrary scatterers and specialize to spherical geometry. We use boldface to indicate vectors. A hat on the top of a vector indicates that the vector has unit magnitude. A tilde on the top of a capital letter denotes a dyadic (second rank tensor). For brevity, we use (3:5) for Eq. (5) of Ref. 3, and retain where it is possible the equational forms of Refs. 10, 15–19.

II. DEFINITIONS AND REPRESENTATION

We consider the scattering of an incident plane harmonic elastic wave ϕ propagating in a direction $\hat{\mathbf{k}}$ by a spherical object embedded in an isotropic and homogeneous elastic host medium specified by Lamé's constants λ, μ and density ρ different from λ', μ', and ρ' of the obstacle.

The problem is to determine the total scattered field due to the incident wave. The incident elastic wave can be either a longitudinal wave denoted as a p-incident wave and given by ${}_p\phi$ or a transverse wave denoted as an s incident wave and given by ${}_s\phi$. The spherical object can be either a perfect elastic sphere, a rigid sphere, or spherical cavity.

Let v be the volume occupied by the scatterer and V be the volume outside of the scatterer. The surface of the spherical object is S, its radius is a, and it is centered at $\mathbf{r} = \mathbf{0}$. The p incident wave is parallel to the propagation vector $\hat{\mathbf{k}}$ while the s incident wave lies along the polarization vector $\hat{\boldsymbol{\epsilon}}_s$ normal to the direction of propagation

$$\hat{\boldsymbol{\epsilon}}_s \cdot \hat{\mathbf{k}} = 0. \tag{1}$$

The propagation parameters k_p, k_s satisfy

$$\omega = c_p k_p = c_s k_s, \tag{2}$$

where $c_p^2 = (\lambda + 2\mu)/\rho$, $c_s^2 = \mu/\rho$ are the longitudinal and transverse phase velocity squared.

We suppress the time-harmonic dependence $e^{-i\omega t}$ and define the p and s incident waves as

$$\begin{aligned} {}_p\phi(\mathbf{r}) &= \hat{\mathbf{k}} e^{ik_p \hat{\mathbf{k}}\cdot\mathbf{r}}, \\ {}_s\phi(\mathbf{r}) &= \hat{\boldsymbol{\epsilon}}_s e^{ik_s \hat{\mathbf{k}}\cdot\mathbf{r}}, \end{aligned} \tag{3}$$

Hence, all quantities in the entire paper have at most a spatial dependence. A factor of $e^{-i\omega t}$ is assumed in the final answer.

In the region exterior to the scatterer, the total field due to an incoming x wave (where x is a p or an s wave)

$${}_x\psi = {}_x\phi + {}_x\mathbf{u}, \tag{4}$$

satisfies the time-independent linearized equation of dynam-

 0022-2488/90/112626-07$03.00

ic elasticity in the absence of body forces

$$[c_p^2\nabla(\nabla\cdot) - c_s^2\nabla\times(\nabla\times) + \omega^2]\psi = \mathbf{0}. \tag{5}$$

The radiative functions $\mathbf{u}$ in (4) which in general consist of p or s wave fields are given by (15:23) or (18:2.8)

$$\mathbf{u}(\mathbf{r}) = \{\tilde{\Gamma},\mathbf{u}\}$$

$$\equiv \frac{1}{4\pi}\int_S [\tilde{\Gamma}\cdot\mathbf{T}_{\mathbf{r}'}\mathbf{u} - \mathbf{u}\cdot\mathbf{T}_{\mathbf{r}'}\tilde{\Gamma}]\,dS(\mathbf{r}'). \tag{6}$$

Here,

$$\mathbf{T}_{\mathbf{r}} \equiv 2\mu\hat{\mathbf{n}}\cdot\nabla_{\mathbf{r}} + \lambda\hat{\mathbf{n}}\nabla_{\mathbf{r}}\cdot + \mu\hat{\mathbf{n}}\times(\nabla\times) \tag{7}$$

is the surface stress operator and $\hat{\mathbf{n}}$ is the exterior unit normal to S. The operator in (6) is a surface integral which arises in Betti's[24] identities for elasticity. Here $\tilde{\Gamma}(\mathbf{r},\mathbf{r}')$ is the fundamental dyadic solution of

$$\mathsf{L}\{\tilde{\Gamma}(\mathbf{r},\mathbf{r}')\} = -4\pi\delta(\mathbf{r}-\mathbf{r}')\tilde{\mathbf{I}}, \tag{8}$$

with

$$\mathsf{L} = c_s^2\nabla_{\mathbf{r}'}^2 + (c_p^2 - c_s^2)\nabla_{\mathbf{r}'}(\nabla_{\mathbf{r}'}\cdot) + \omega^2, \tag{9}$$

where $\tilde{\mathbf{I}} = \hat{\mathbf{r}}\hat{\mathbf{r}} + \hat{\boldsymbol{\theta}}\hat{\boldsymbol{\theta}} + \hat{\boldsymbol{\varphi}}\hat{\boldsymbol{\varphi}}$ is the identity dyadic and $\delta(\mathbf{r}-\mathbf{r}')$ represents the Dirac measure concentrated at $\mathbf{r}$.

Decomposing the net scattered fields of the single obstacle due to an incoming x wave into the form of a p plus s field we have

$${}_x\mathbf{u}(\mathbf{r}) = {}_x\mathbf{u}_p(\mathbf{r}) + {}_x\mathbf{u}_s(\mathbf{r}). \tag{10}$$

Using (15:35), (15:36), and (18:2.20–2.22) for $r\to\infty$, we obtain from (10) the asymptotic form

$${}_x\mathbf{u}(\mathbf{r}) \sim \left[\frac{e^{ik_p r}}{ik_p r}\right]{}_x\mathbf{g}_p(\hat{\mathbf{r}},\hat{\mathbf{k}}) + \left[\frac{e^{ik_s r}}{ik_s r}\right]_x \mathbf{g}_s(\hat{\mathbf{r}},\hat{\mathbf{k}}),$$

$$x\in\{p,s\}, \tag{11}$$

where

$${}_x\mathbf{g}_p(\hat{\mathbf{r}},\hat{\mathbf{k}}) = \frac{ik_p}{4\pi\rho c_p^2}\{\hat{\mathbf{r}}\hat{\mathbf{r}}e^{-ik_p\hat{\mathbf{r}}\cdot\mathbf{r}'},[{}_x\mathbf{u}(\mathbf{r}')]\}, \tag{12}$$

$${}_x\mathbf{g}_s(\hat{\mathbf{r}},\hat{\mathbf{k}}) = \frac{ik_s}{4\pi\rho c_s^2}\{\tilde{\mathbf{I}}_\cdot e^{-ik_s\hat{\mathbf{r}}\cdot\mathbf{r}'},[{}_x\mathbf{u}(\mathbf{r}')]\}, \tag{13}$$

and

$$\tilde{\mathbf{I}}_\cdot = (\tilde{\mathbf{I}} - \hat{\mathbf{r}}\hat{\mathbf{r}}). \tag{14}$$

In a manner similar to Refs. 10, 17–24, we write the spectral representation of the scattered waves due to a single object as

$${}_x\mathbf{u}(\mathbf{r}) = \frac{1}{2\pi}\int_c \{e^{i\mathbf{k}_{pc}\cdot\mathbf{r}}\,{}_x\mathbf{g}_p + e^{i\mathbf{k}_{sc}\cdot\mathbf{r}}\,{}_x\mathbf{g}_s\}\,d\Omega_c,$$

$$x\in\{p,s\},\quad \mathbf{g} = \mathbf{g}(\hat{\mathbf{r}}_c,\hat{\mathbf{k}}). \tag{15}$$

In (15), $\mathbf{k}_{kc} = k_p\hat{\mathbf{r}}_c$ and $\mathbf{k}_{sc} = k_s\hat{\mathbf{r}}_c$. The unit vector $\hat{\mathbf{r}}_c = \hat{\mathbf{r}}_c(\theta_c,\varphi_c) = (\cos\theta_c\sin\varphi_c)\hat{\mathbf{i}} + (\sin\theta_c\sin\varphi_c)\hat{\mathbf{j}} + (\cos\varphi_c)\hat{\mathbf{k}}$, and c is the Sommerfeld's path.[23] The contour path c is chosen to insure that $\mathrm{Im}[\hat{r}_c]\cdot(\mathbf{r}-\mathbf{r}')\geqslant 0$ (see Refs. 10, 17–23).

III. MULTIPLE SCATTERING

We consider a fixed configuration of N identical spheres with centers located by $\mathbf{r}_m$ $(m = 1,2,\ldots,N)$ (16:12, 17:11). The total multiple ${}_x\Psi$ correspond to an incoming x wave is

$${}_x\Psi = {}_x\phi + \sum_{m=1}^{N} {}_x\mathbf{U}_m(\mathbf{r}-\mathbf{r}_m), \tag{16}$$

where ${}_x\mathbf{U}_m$ represents the total multiple scattered field.

Decomposing $\mathbf{U}$ into p and s fields

$${}_x\mathbf{U}_m(\mathbf{r}-\mathbf{r}_m) = {}_x\mathbf{U}_{m:p} + {}_x\mathbf{U}_{m:s}, \tag{17}$$

for the multiple-scattered wave that radiates from the scatterer fixed at $\mathbf{r}_m$, we have, with reference to the scatterer located by $\mathbf{r}_t$,

$${}_x\Psi_t(\mathbf{r}) = {}_x\phi(\mathbf{r}) + \sum_m{}' {}_x\mathbf{U}_m + {}_x\mathbf{U}_t(\mathbf{r}-\mathbf{r}_t), \tag{18}$$

where

$${}_x\mathbf{U}_t(\mathbf{r}-\mathbf{r}_t) = {}_x\mathbf{U}_{t:p} + {}_x\mathbf{U}_{t:s}, \tag{19}$$

and

$$\sum_m{}' = \sum_{m\neq t}. \tag{20}$$

Using the surface integral representation of Sec. II, the multiple-scattered wave is

$${}_x\mathbf{U}_t = \{\tilde{\Gamma}_p,[{}_x\mathbf{U}_t(\mathbf{r}_t+\mathbf{r}')]\} + \{\tilde{\Gamma}_s,[{}_x\mathbf{U}_t(\mathbf{r}_t+\mathbf{r}')]\}. \tag{21}$$

Asymptotically, for $r\to\infty$, (21) becomes

$${}_x\mathbf{U}_t(\mathbf{R}_t) = h(k_p|\mathbf{R}_t|)\,{}_x\mathbf{G}_{t:p} + h(k_s|\mathbf{R}_t|)\,{}_x\mathbf{G}_{t:s}, \tag{22}$$

where

$$\mathbf{R}_t = \mathbf{r} - \mathbf{r}_t,\quad \mathbf{G} = \mathbf{G}(\mathbf{r}_t;\hat{\mathbf{r}}). \tag{23}$$

Here, $h(k_x|\mathbf{r}-\mathbf{r}_t|)$ is the Hankel's function of the first kind, ${}_x\mathbf{G}_{t:p}$ and ${}_x\mathbf{G}_{t:s}$ are the longitudinal and transverse multiple vector scattering amplitudes due an incoming x wave.

Using the same techniques, as in (12) and (13), the multiple vector scattering amplitudes can be represented by

$${}_x\mathbf{G}_{t:p}(\mathbf{r}_t,\hat{\mathbf{r}}) = (ik_p/4\pi\rho c_p^2)\{\hat{\mathbf{r}}\hat{\mathbf{r}}e^{-ik_p\hat{\mathbf{r}}\cdot\mathbf{r}'},({}_x\mathbf{U}_t)\}, \tag{24}$$

and

$${}_x\mathbf{G}_{t:s}(\mathbf{r}_t;\hat{\mathbf{r}}) = (ik_s/4\pi\rho c_s^2)\{\tilde{\mathbf{I}}_\cdot e^{-ik_s\hat{\mathbf{r}}\cdot\mathbf{r}'},({}_x\mathbf{U}_t)\}, \tag{25}$$

where ${}_x\mathbf{U}_t = {}_x\mathbf{U}_t(\mathbf{r}_t+\mathbf{r}')$.

The spectral representation of the multiple-scattered wave analog to (15) is

$${}_x\mathbf{U}_t(\mathbf{R}_t) = \int_c \{e^{i\mathbf{k}_{pc}\cdot\mathbf{R}}\,{}_x\mathbf{G}_{t:pc} + e^{i\mathbf{k}_{sc}\cdot\mathbf{R}}\,{}_x\mathbf{G}_{t:sc}\}, \tag{26}$$

where ${}_x\mathbf{G}_{t:yc} = {}_x\mathbf{G}_{t:y}(\hat{\mathbf{r}}_c)$, $y\in\{p,s\}$, and

$$\int_c\{\ \} = \frac{1}{2\pi}\int_c\{\ \}\,d\Omega_c. \tag{27}$$

IV. RECIPROCITY RELATIONS FOR MULTIPLE ELASTIC SCATTERING

In this section, we establish reciprocity relations for the multiple vector scattering amplitudes corresponding to different combinations of the incoming p or s waves. In general, we consider two arbitrary outside total multiple solutions ${}_x\Psi_1$, ${}_y\Psi_2$ due to two arbitrary incident waves ${}_x\phi_1$, ${}_y\phi_2$, where x and y can be either p or s. We will derive the $x = p$, $y = p$ and $x = p$, $y = s$ cases. The remaining two cases are similar.

For the $x=p, y=p$ case, we consider for the t scatterer two arbitrary total outside configurational solutions ${}_p\Psi_{t:1}$, ${}_p\Psi_{t:2}$. The requirement that ${}_p\Psi_{t:1}, {}_p\Psi_{t:2}$ satisfy the same conditions on the surface S_t of the scatterer and within the volume v_t gives

$$\{{}_p\Psi_{t:1},({}_p\Psi_{t:2})\}_t = 0. \tag{28}$$

Applying (18) to (28), we have

$$\{{}_p\Phi_1 + {}_pU_{t_1}(\mathbf{r}_t+\mathbf{r}'),[{}_p\Phi_2 + {}_pU_{t_2}(\mathbf{r}_t+\mathbf{r}')]\}_t = 0, \tag{29}$$

where

$${}_p\Phi_i = {}_p\phi_i + \sum_m{}' {}_pU_{m_i}(\mathbf{r}-\mathbf{r}_m); \quad i=1,2. \tag{30}$$

Using the brace algebra of Refs. 10, 16–18, we recast (29) and (30) into

$$\{{}_p\phi_1,[{}_pU_{t_2}]\}_t + \left\{\sum_m{}' {}_pU_{m_1}(\mathbf{r}-\mathbf{r}_m),[{}_pU_{t_2}]\right\}_t = \{{}_p\phi_2,[{}_pU_{t_1}]\}_t + \left\{\sum_m{}' {}_pU_{m_2}(\mathbf{r}-\mathbf{r}_m),[{}_pU_{t_1}]\right\}_t. \tag{31}$$

At m_1, m_2 we use spectral representation or (26) to obtain

$$\{{}_p\phi_1,({}_pU_{t_2})\}_t + \sum_m{}' \int_c \{(e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_m}{}_pG_{m_1:pc} + e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_m}{}_pG_{m_1:sc}),({}_pU_{t_2})\}_t = \{{}_p\phi_2,({}_pU_{t_1})\}_t + \sum_m{}' \int_c \{(e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_m}{}_pG_{m_2:pc} + e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_m}{}_pG_{m_2:sc}),({}_pU_{t_1})\}_t, \tag{32}$$

$\mathbf{R}_m = \mathbf{r}-\mathbf{r}_m$.

Using

$${}_p\phi_i = \hat{\mathbf{k}}_i(\hat{\mathbf{k}}_i\hat{\mathbf{k}}_i e^{ik_p\hat{\mathbf{k}}_i\cdot\mathbf{r}'}),$$
$${}_pG_{m:xc} = {}_pG_{m:x}(\hat{\mathbf{r}}_c) = {}_p\hat{\boldsymbol{\epsilon}}_{m:x}[{}_pG_{m:x}(\hat{\mathbf{r}}_c)], \tag{33}$$

with $i\in[1,2]$, and $x\in\{p,s\}$, working similarly to (10:20, 121–126), and (18:3.18–3.20) with $\mathbf{R}_{tm} = \mathbf{r}_t - \mathbf{r}_m$, we transform (32) into

$$\hat{\mathbf{k}}_1\cdot\{\hat{\mathbf{k}}_1\hat{\mathbf{k}}_1 e^{ik_p\hat{\mathbf{k}}_1\cdot\mathbf{r}'},({}_pU_{t_2})\}_t + \sum_m{}' \int_c e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_{tm}}{}_p\hat{\boldsymbol{\epsilon}}_{m_1:p}\cdot\{\hat{\mathbf{r}}_c\hat{\mathbf{r}}_c e^{i\mathbf{k}_{pc}\cdot\mathbf{r}'},({}_pU_{t_2})\}_t\,{}_pG_{m_1:pc}$$
$$+ \sum_m{}' \int_c e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_{tm}}{}_p\hat{\boldsymbol{\epsilon}}_{m_1:s}\cdot\{\tilde{\mathbf{I}}_{*_c} e^{i\mathbf{k}_{sc}\cdot\mathbf{r}'},({}_pU_{t_2})\}_t\,{}_pG_{m_1:sc}$$
$$= \hat{\mathbf{k}}_2\cdot\{\hat{\mathbf{k}}_2\hat{\mathbf{k}}_2 e^{ik_p\hat{\mathbf{k}}_2\cdot\mathbf{r}'},({}_pU_{t_1})\}_t + \sum_m{}' \int_c e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_{tm}}{}_p\hat{\boldsymbol{\epsilon}}_{m_2:p}\cdot\{\hat{\mathbf{r}}_c\hat{\mathbf{r}}_c e^{i\mathbf{k}_{pc}\cdot\mathbf{r}'},({}_pU_{t_1})\}_t\,{}_pG_{m_2:pc}$$
$$+ \sum_m{}' \int_c e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_{tm}}{}_p\epsilon_{m_2:s}\cdot\{\tilde{\mathbf{I}}_{*_c} e^{i\mathbf{k}_{sc}\cdot\mathbf{r}'},({}_pU_{t_1})\}_t\,{}_pG_{m_2:sc}. \tag{34}$$

Here we define

$$\tilde{\mathbf{I}}_c = (\hat{\mathbf{r}}_c\hat{\mathbf{r}}_c) + (\hat{\boldsymbol{\theta}}_c\hat{\boldsymbol{\theta}}_c + \hat{\boldsymbol{\varphi}}_c\hat{\boldsymbol{\varphi}}_c), \quad \tilde{\mathbf{I}}_{*_c} = \tilde{\mathbf{I}}_c - (\hat{\mathbf{r}}_c\hat{\mathbf{r}}_c), \tag{35}$$

with ${}_p\hat{\boldsymbol{\epsilon}}_{m:p}\|\hat{\mathbf{r}}_c$ with ${}_p\hat{\boldsymbol{\epsilon}}_{m:s}\perp\hat{\mathbf{r}}_c$. In (34), the six quantities inside the braces are proportional to the configurational vector scattering amplitudes due to the different longitudinal incident waves. Combining (24) and (25) with (34) leads to the p–p multiple elastic reciprocity relation

$$\frac{c_p^2}{k_p}\hat{k}_1\cdot{}_pG_{t:p}(-\hat{k}_1,\hat{k}_2) + \frac{c_p^2}{k_p}\sum_m{}' \int_c e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_{tm}}{}_p\hat{\boldsymbol{\epsilon}}_{m_1:p}\cdot{}_pG_{t:p}(-\hat{r}_c,\hat{k}_2)[{}_pG_{m_1:p}(\hat{r}_c)]$$
$$+ \frac{c_s^2}{k_s}\sum_m{}' \int_c e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_{tm}}{}_p\hat{\boldsymbol{\epsilon}}_{m_1:s}\cdot{}_pG_{t:s}(-\hat{r}_c,\hat{k}_2)[{}_pG_{m_1:s}(\hat{r}_c)]$$
$$= \frac{c_p^2}{k_p}\hat{\mathbf{k}}_2\cdot{}_pG_{t:p}(-\hat{\mathbf{k}}_2,\hat{\mathbf{k}}_1) + \frac{c_p^2}{k_p}\sum_m{}' \int_c e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_{tm}}{}_p\hat{\boldsymbol{\epsilon}}_{m_2:p}\cdot{}_pG_{t:p}(-\hat{\mathbf{r}}_c,\hat{\mathbf{k}}_1)[{}_pG_{m_2:p}(\hat{\mathbf{r}}_c)]$$
$$+ \frac{c_s^2}{k_s}\sum_m{}' \int_c e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_{tm}}{}_p\hat{\boldsymbol{\epsilon}}_{m_2:s}\cdot{}_pG_{t:s}(-\hat{\mathbf{r}}_c,\hat{\mathbf{k}}_2)[{}_pG_{m_2:s}(\hat{\mathbf{r}}_c)]. \tag{36}$$

The result for two transverse incident waves is obtained from (36) with p replaced by s, and $\hat{\mathbf{k}}_i$ replaced by $\hat{\boldsymbol{\epsilon}}_i$ in the appropriate locations. Hence, the s–s multiple elastic reciprocity relation is

$$\frac{c_s^2}{k_s}\hat{\epsilon}_{s1}\cdot{}_sG_{t:s}(-\hat{k}_1,\hat{k}_2)+\frac{c_p^2}{k_p}\sum_m{}'\int_c e^{ik_{pc}\cdot R_{tm}}{}_s\hat{\epsilon}_{m_1:p}\cdot{}_sG_{t:p}(-\hat{r}_c,\hat{k}_2)[{}_sG_{m_1:p}(\hat{r}_c)]$$
$$+\frac{c_s^2}{k_s}\sum_m{}'\int_c e^{ik_{sc}\cdot R_{tm}}{}_s\hat{\epsilon}_{m_1:s}\cdot{}_sG_{t:s}(-\hat{r}_c,\hat{k}_2)[{}_sG_{m_1:s}(\hat{r}_c)]$$
$$=\frac{c_s^2}{k_s}\hat{\epsilon}_{s2}\cdot{}_sG_{t:s}(-\hat{k}_2,\hat{k}_1)+\frac{c_p^2}{k_p}\sum_m{}'\int_c e^{ik_{pc}\cdot R_{tm}}{}_s\hat{\epsilon}_{m_2:p}\cdot{}_sG_{t:p}(-\hat{r}_c,\hat{k}_1)[{}_sG_{m_2:p}(\hat{r}_c)]$$
$$+\frac{c_s^2}{k_s}\sum_m{}'\int_c e^{ik_{sc}\cdot R_{tm}}{}_s\hat{\epsilon}_{m_2:s}\cdot{}_sG_{t:s}(-\hat{r}_c,\hat{k}_2)[{}_sG_{m_2:s}(\hat{r}_c)]. \tag{37}$$

Here, $\hat{\epsilon}_{s1},\hat{\epsilon}_{s2}$ are the polarization vectors of the two transverse incident waves.

For the case of $x=p$, $y=s$, we proceed as we did in the p–p incident waves. With p and s incident waves, Eqs. (29) and (30) read

$$\{{}_p\Phi_1+{}_pU_{t_1}(r_t+r'),[{}_s\Phi_2+{}_sU_{t_2}(r_t+r')]\}_t=0, \tag{38}$$

with

$${}_p\Phi_1={}_p\phi_1+\sum_m{}'{}_pU_{m_1}(r-r_m),\quad {}_s\Phi_2={}_s\phi_2+\sum_m{}'{}_sU_{m_2}(r-r_m). \tag{39}$$

Following (31), we transform (38) into

$$\{{}_p\phi_1,({}_sU_{t_2})\}_t+\sum_m{}'\int_c\{(e^{ik_{pc}\cdot R_m}{}_pG_{m_1:pc}+e^{ik_{sc}\cdot R_m}{}_pG_{m_1:sc}),({}_sU_{t_2})\}_t$$
$$=\{{}_s\phi_2,({}_pU_{t_1})\}_t+\sum_m{}'\int_c\{(e^{ik_{pc}\cdot R_m}{}_sG_{m_2:pc}+e^{ik_{sc}\cdot R_m}{}_sG_{m_2:sc}),({}_pU_{t_1})\}_t. \tag{40}$$

As in Refs. 10 and 18, and with

$${}_sG_{m:xc}={}_sG_{m:x}(\hat{r}_c)={}_s\hat{\epsilon}_{m:x}[{}_sG_{m:x}(\hat{r}_c)],\quad {}_s\phi_2=\hat{\epsilon}_{s_2}\cdot\{\tilde{I}_*e^{ik_s\hat{k}_2\cdot r'}\}, \tag{41}$$

we reduce (40) to

$$\hat{k}_1\cdot\{\hat{k}_1\hat{k}_1e^{ik_p\hat{k}_1\cdot r'},({}_sU_{t_2})\}_t+\sum_m{}'\int_c e^{ik_{pc}\cdot R_{tm}}{}_p\hat{\epsilon}_{m_1:p}\cdot\{\hat{r}_c\hat{r}_ce^{ik_{pc}\cdot r'},({}_sU_{t_2})\}_t\,{}_pG_{m_1:pc}$$
$$+\sum_m{}'\int_c e^{ik_{sc}\cdot R_{tm}}{}_p\hat{\epsilon}_{m_1:s}\cdot\{\tilde{I}_*e^{ik_{sc}\cdot r'},({}_sU_{t_2})\}_t\,{}_pG_{m_1:sc}$$
$$=\hat{\epsilon}_2\cdot\{\tilde{I}_*e^{ik_s\hat{k}_2\cdot r'},({}_pU_{t_1})\}_t+\sum_m{}'\int_c e^{ik_{pc}\cdot R_{tm}}{}_s\hat{\epsilon}_{m_2:p}\cdot\{\hat{r}_c\hat{r}_ce^{ik_{pc}\cdot r'},({}_pU_{t_1})\}_t\,{}_sG_{m_2:pc}$$
$$+\sum_m{}'\int_c e^{ik_{sc}\cdot R_{tm}}{}_s\hat{\epsilon}_{m_2:s}\cdot\{\tilde{I}_*e^{ik_{sc}\cdot r'},({}_pU_{t_1})\}_t\,{}_sG_{m_2:sc}. \tag{42}$$

The six quantities inside the braces in (42) are proportional to the configurational vector scattering amplitudes due to either p or s incident waves. Combining (24) and (25) with (42) gives the p–s multiple elastic reciprocity relation

$$\frac{c_p^2}{k_p}\hat{k}_1\cdot{}_sG_{t:p}(-\hat{k}_1,\hat{k}_2)+\frac{c_p^2}{k_p}\sum_m{}'\int_c e^{ik_{pc}\cdot R_{tm}}{}_p\hat{\epsilon}_{m_1:p}\cdot{}_sG_{t:p}(-\hat{r}_c,\hat{k}_2)[{}_pG_{m_1:p}(\hat{r}_c)]$$
$$+\frac{c_s^2}{k_s}\sum_m{}'\int_c e^{ik_{sc}\cdot R_{tm}}{}_p\hat{\epsilon}_{m_1:s}\cdot{}_sG_{t:s}(-\hat{r}_c,\hat{k}_2)[{}_pG_{m_1:s}(\hat{r}_c)]$$
$$=\frac{c_s^2}{k_s}\hat{\epsilon}_{s_2}\cdot{}_pG_{t:s}(-\hat{k}_2,\hat{k}_1)+\frac{c_p^2}{k_p}\sum_m{}'\int_c e^{ik_{pc}\cdot R_{tm}}{}_s\hat{\epsilon}_{m_2:p}\cdot{}_pG_{t:p}(-\hat{r}_c,\hat{k}_1)[{}_sG_{m_2:p}(\hat{r}_c)]$$
$$\frac{c_s^2}{k_s}\sum_m{}'\int_c e^{ik_{sc}\cdot R_{tm}}{}_s\hat{\epsilon}_{m_2:s}\cdot{}_pG_{t:s}(-\hat{r}_c,\hat{k}_1)[{}_sG_{m_2:s}(\hat{r}_c)]. \tag{43}$$

With s and p incident waves, we obtain from (40) the s–p multiple elastic reciprocity relation

$$\frac{c_s^2}{k_s}\hat{\mathbf{e}}_{s_1}\cdot{}_p\mathbf{G}_{t:s}(-\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2)+\frac{c_p^2}{k_p}\sum_m{}'\int_c e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_{tm}}{}_s\hat{\mathbf{e}}_{m_{1:p}}\cdot{}_p\mathbf{G}_{t:p}(-\hat{\mathbf{r}}_c,\hat{\mathbf{k}}_2)[{}_sG_{m_1:p}(\hat{\mathbf{r}}_c)]$$

$$+\frac{c_s^2}{k_s}\sum_m{}'\int_c e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_{tm}}{}_s\hat{\mathbf{e}}_{m_{1:s}}\cdot{}_p\mathbf{G}_{t:s}(-\hat{\mathbf{r}}_c,\hat{\mathbf{k}}_2)[{}_sG_{m_1:s}(\hat{\mathbf{r}}_c)]$$

$$=\frac{c_p^2}{k_p}\hat{\mathbf{k}}_2\cdot{}_s\mathbf{G}_{t:p}(-\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2)+\frac{c_p^2}{k_p}\sum_m{}'\int_c e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_{tm}}{}_p\hat{\mathbf{e}}_{m_{2:p}}\cdot{}_s\mathbf{G}_{t:p}(-\hat{\mathbf{r}}_c,\hat{\mathbf{k}}_1)[{}_pG_{m_2:p}(\hat{\mathbf{r}}_c)]$$

$$+\frac{c_s^2}{k_s}\sum_m{}'\int_c e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_{tm}}{}_p\hat{\mathbf{e}}_{m_{2:s}}\cdot{}_s\mathbf{G}_{t:s}(-\hat{\mathbf{r}}_c,\hat{\mathbf{k}}_1)[{}_pG_{m_2:s}(\hat{\mathbf{r}}_c)]. \tag{44}$$

Equations (36), (37), (43), and (44) are the multiple scattering reciprocity relations.

V. ENERGY THEOREMS AND SCATTERING CROSS SECTIONS

From Ref. 19, we have

$$\{{}_p\Psi_1^*,{}_p\Psi_2\}_t=c_0\mathbf{S}_a, \tag{45}$$

where $\mathbf{S}_a$ is the configurational absorption cross section which is null for elastic solid[6] made up of lossless scatterers. The $*$ means complex conjugation.

The four different multiple elastic reciprocity relations lead to four different energy theorems. We derive the results for p–p and p–s incident waves, and use them to obtain energy theorems for the s–p and s–s incident cases.

For the p–p case, we have starting with (45), and (30)

$$\{{}_p\phi_1^*,({}_p\mathbf{U}_{t_2})\}_t+\sum_m{}'\{{}_p\mathbf{U}_{m_1}^*,({}_p\mathbf{U}_{t_2})\}_t$$

$$+\{{}_p\mathbf{U}_{t_1}^*,({}_p\mathbf{U}_{t_2})\}_t \tag{46}$$

$$+\{{}_p\mathbf{U}_{t_1}^*,({}_p\phi_2)\}_t+\left\{{}_p\mathbf{U}_{t_1}^*,\sum_m{}'{}_p\mathbf{U}_{m_2}\right\}_t=0.$$

In (46), the first term is transformed by (24) to give

$$\{{}_p\phi_1^*,({}_p\mathbf{U}_{t_2})\}_t=-\beta_p\hat{\mathbf{k}}_1\cdot{}_p\mathbf{G}_{t:p}(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2), \tag{47}$$

and the second term, after applying spectral representation at m_1, complex conjugation, and (24) and (25), becomes

$$\sum_m{}'\{{}_p\mathbf{U}_{m_1}^*,({}_p\mathbf{U}_{t_2})\}_t$$

$$=-\sum_{x=p,s}\beta_x\sum_m{}'\int_c{}^*e^{-i\mathbf{k}_{xc}^*\cdot\mathbf{R}_{tm}}{}_p\hat{\mathbf{e}}_{m1:x}^*$$

$$\cdot{}_pG_{t:x}(\hat{\mathbf{r}}_c^*,\hat{\mathbf{k}}_2)[{}_pG_{m_1:x}^*(\hat{\mathbf{r}}_c)], \tag{48}$$

$\beta_x=(4\pi ic_x^2/k_x)$.

In the third term of (46), by Betti's theorem for the exterior, we replace the surface S_t of the t scatterer by the surface at infinity. From the asymptotic form (22) for ${}_p\mathbf{U}_p(\mathbf{R}_t)$, we obtain[8,10,15,19]

$$\{{}_p\mathbf{U}_{t_1}^*,{}_p\mathbf{U}_{t_2}\}_t$$

$$=\{{}_p\mathbf{U}_{t_1}^*,({}_p\mathbf{U}_{t_2})\}_\infty$$

$$=-4\pi\omega^2i\rho\left(\int_S\{k_p^{-3}{}_p\mathbf{G}_{t:p}^*(\hat{\mathbf{r}},\hat{\mathbf{k}}_1)\cdot{}_p\mathbf{G}_{t:p}(\hat{\mathbf{r}},\hat{\mathbf{k}}_2)\right. \tag{49}$$

$$\left.+k_s^{-3}{}_p\mathbf{G}_{t:s}^*(\hat{\mathbf{r}},\hat{\mathbf{k}}_1)\cdot{}_p\mathbf{G}_{t:s}(\hat{\mathbf{r}},\hat{\mathbf{k}}_2)\}d\Omega_{\hat{\mathbf{r}}}\right)$$

$$=-\frac{2\omega^2i\rho}{k_p}{}_{(p,p)}\mathbf{S}_{\text{scat}}(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2),$$

where ${}_{(p,p)}\mathbf{S}_{\text{scat}}(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2)$ is the configurational scattering cross section due to p–p incident waves, and $d\Omega_{\hat{\mathbf{r}}}$ is the differential solid angle around $\hat{\mathbf{r}}$. The integration in (49) is over all angles of observation.

Using (6), we rewrite the fourth and fifth terms of (46) to obtain

$$\{{}_p\mathbf{U}_{t_1}^*,({}_p\phi_2)\}_t=-\{{}_p\phi_2^*,({}_p\mathbf{U}_{t_1})\}_t^* \tag{50}$$

and

$$\{{}_p\mathbf{U}_{t_1}^*,\sum_m{}'{}_p\mathbf{U}_{m_2}\}_t=-\sum_m{}'\{{}_p\mathbf{U}_{m_2}^*,({}_p\mathbf{U}_{t_1})\}_t^*. \tag{51}$$

Proceeding similarly to (48) and (49), (50) gives

$$\{{}_p\mathbf{U}_{t_1}^*,({}_p\phi_2)\}_t=-\{{}_p\phi_2^*,({}_p\mathbf{U}_{t_1})\}_t^*$$

$$=-\beta_p\hat{\mathbf{k}}_2\cdot{}_p\mathbf{G}_{t:p}^*(\hat{\mathbf{k}}_2,\hat{\mathbf{k}}_1), \tag{52}$$

and (51) leads to

$$\left\{{}_p\mathbf{U}_{t_1}^*,\sum_m{}'{}_p\mathbf{U}_{m_2}\right\}_t=-\sum_m{}'\{{}_p\mathbf{U}_{m_2}^*,({}_p\mathbf{U}_{t_1})\}_t^*$$

$$=-\sum_{x=p,s}\beta_x\sum_m{}'\int_c e^{i\mathbf{k}_{xc}\cdot\mathbf{R}_{tm}}{}_p\hat{\mathbf{e}}_{m_{2:x}}$$

$$\cdot{}_p\mathbf{G}_{t:x}^*(\hat{\mathbf{r}}_c^*,\hat{\mathbf{k}}_1)[{}_pG_{m_2:x}(\hat{\mathbf{r}}_c)]. \tag{53}$$

Combining (46) with (47)–(49), (52), and (53) leads to the general p–p "energy theorem"

$$\frac{4\pi}{k_p^3}[\hat{\mathbf{k}}_1\cdot{}_p\mathbf{G}_{t:p}(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2)+\hat{\mathbf{k}}_2\cdot{}_p\mathbf{G}_{t:p}^*(\hat{\mathbf{k}}_2,\hat{\mathbf{k}}_1)]+\frac{4\pi}{k_p^3}\left[\sum_m{}'\int_c{}^*e^{-i\mathbf{k}_{pc}^*\cdot\mathbf{R}_{tm}}{}_p\hat{\mathbf{e}}_{m_{1:p}}^*\cdot{}_p\mathbf{G}_{t:p}(\hat{\mathbf{r}}_c^*,\hat{\mathbf{k}}_2)[{}_pG_{m_1:p}^*(\hat{\mathbf{r}}_c)]\right.$$

$$\left.+\sum_m{}'\int_c e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_{tm}}{}_p\hat{\mathbf{e}}_{m_{2:p}}\cdot{}_p\mathbf{G}_{t:p}^*(\hat{\mathbf{r}}_c^*,\hat{\mathbf{k}}_1)[{}_pG_{m_2:p}(\hat{\mathbf{r}}_c)]\right]+\frac{4\pi}{k_s^3}\left[\sum_m{}'\int_c{}^*e^{-i\mathbf{k}_{sc}^*\cdot\mathbf{R}_{tm}}{}_p\hat{\mathbf{e}}_{m_{1:s}}^*\cdot{}_p\mathbf{G}_{t:s}(\hat{\mathbf{r}}_c^*,\hat{\mathbf{k}}_2)[{}_pG_{m_1:s}^*(\hat{\mathbf{r}}_c)]\right] \tag{54}$$

$$\left.+\sum_m{}'\int_c e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_{tm}}{}_p\hat{\mathbf{e}}_{m_{2:s}}\cdot{}_p\mathbf{G}_{t:s}^*[{}_pG_{m_2:s}(\hat{\mathbf{r}}_c)]\right]=-\frac{2}{k_p}{}_{(p,p)}\mathbf{S}_{\text{scat}}(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2).$$

For the forward direction, assuming that the directions of longitudinal and transverse waves coincide, we obtain the p–p "forward energy theorem"

$$-\frac{4\pi}{k_p^2}\Re\left[\hat{\mathbf{k}}\cdot{}_pG_{t:p}(\hat{\mathbf{k}},\hat{\mathbf{k}})+\sum_m{}'\int_c e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_{tm}}{}_p\hat{\boldsymbol{\epsilon}}_{m:p}\cdot{}_pG^*_{t:p}(\hat{\mathbf{r}}_c^*,\hat{\mathbf{k}})[\,{}_pG_{m:p}(\hat{\mathbf{r}}_c)]\right]={}_{(p,p)}S_{\mathrm{scat}}(\hat{\mathbf{k}},\hat{\mathbf{k}}),\tag{55}$$

where $\Re$ denotes the real part.

We can use (54) to write the geneal s–s "energy theorem"

$$\begin{aligned}&\frac{4\pi}{k_s^3}[\hat{\boldsymbol{\epsilon}}_{s_1}\cdot{}_sG_{t:s}(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2)+\hat{\boldsymbol{\epsilon}}_{s_2}\cdot{}_sG^*_{t:s}(\hat{\mathbf{k}}_2,\hat{\mathbf{k}}_1)]+\frac{4\pi}{k_p^3}\left[\sum_m{}'\int_c{}^* e^{-i\mathbf{k}^*_{pc}\cdot\mathbf{R}_{tm}}{}_s\hat{\boldsymbol{\epsilon}}^*_{m_1:p}\cdot{}_sG_{t:p}(\hat{\mathbf{r}}_c^*,\hat{\mathbf{k}}_2)[\,{}_sG^*_{m_1:p}(\hat{\mathbf{r}}_c)]\right.\\&\quad+\sum_m{}'\int_c e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_{tm}}{}_s\hat{\boldsymbol{\epsilon}}_{m_2:p}\cdot{}_sG^*_{t:p}(\hat{\mathbf{r}}_c^*,\hat{\mathbf{k}}_1)[\,{}_sG_{m_2:p}(\hat{\mathbf{r}}_c)]\Big]+\frac{4\pi}{k_s^3}\left[\sum_m{}'\int_c{}^* e^{-i\mathbf{k}^*_{sc}\cdot\mathbf{R}_{tm}}{}_s\hat{\boldsymbol{\epsilon}}^*_{m_1:s}\cdot{}_sG_{t:s}(\hat{\mathbf{r}}_c^*,\hat{\mathbf{k}}_2)[\,{}_sG^*_{m_1:s}(\hat{\mathbf{r}}_c)]\right.\\&\quad+\sum_m{}'\int_c e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_{tm}}{}_s\hat{\boldsymbol{\epsilon}}_{m_2:s}\cdot{}_sG^*_{t:s}(\hat{\mathbf{r}}_c^*,\hat{\mathbf{k}}_1)[\,{}_sG_{m_2:s}(\hat{\mathbf{r}}_c)]\Big]=-\frac{2}{k_s}{}_{(s,s)}S_{\mathrm{scat}}(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2),\end{aligned}\tag{56}$$

and

$$\begin{aligned}&{}_{(s,s)}S_{\mathrm{scat}}(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2)\\&\quad=2\pi k_s\sum_{x=p,s}k_x^{-3}\Big(\int_S{}_sG^*_{t:x}(\hat{\mathbf{r}},\hat{\mathbf{k}}_1)\\&\qquad\cdot{}_sG_{t:x}(\hat{\mathbf{r}},\hat{\mathbf{k}}_2)d\Omega_{\hat{\mathbf{r}}}\Big).\end{aligned}\tag{57}$$

In particular, in the forward direction, we obtain the s–s "forward energy theorem:"

$$\begin{aligned}&-\frac{4\pi}{k_s^2}\Re\left[\hat{\boldsymbol{\epsilon}}\cdot{}_sG_{t:s}(\hat{\mathbf{k}},\hat{\mathbf{k}})+\sum_m{}'\int_c e^{i\mathbf{k}_{sc}\cdot\mathbf{R}_{tm}}{}_s\hat{\boldsymbol{\epsilon}}_{m:s}\right.\\&\quad\left.\cdot{}_sG^*_{t:s}(\hat{\mathbf{r}}_c^*,\hat{\mathbf{k}})[\,{}_sG_{m:s}(\hat{\mathbf{r}}_c)]\right]={}_{(s,s)}S_{\mathrm{scat}}(\hat{\mathbf{k}},\hat{\mathbf{k}}).\end{aligned}\tag{58}$$

The results of (55) and (58) differ from the results of Barratt and Collins (1:3.5, 3.9, 3.10) or Barratt (20:4.5,4.6), only in the way the vector multiple scattering amplitudes are defined. When the fixed configuration is reduced to a single scatterer, we recover the results of Dassios *et al.* (8:91, 92, 94), and Varatharajulu (6:3.9, 3.10).

For p–s incident waves, we reduce (46) to

$$\begin{aligned}&\{{}_p\phi_1^*,({}_sU_{t_2})\}_t+\sum_m{}'\{{}_pU^*_{m_1},({}_sU_{t_2})\}_t\\&\quad+\{{}_pU^*_{t_1},({}_sU^*_{t_2})\}_t+\{{}_pU^*_{t_1},({}_s\phi_2)\}_t\\&\quad+\{{}_pU^*_{t_1},\sum_m{}'{}_sU_{m_2}\}_t=0.\end{aligned}\tag{59}$$

Therefore, (47) and (48) become

$$\{{}_p\phi_1^*,({}_sU_{t_2})\}_t=-\beta_p\hat{\mathbf{k}}_1\cdot{}_sG_{t:p}(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2),\tag{60}$$

and

$$\begin{aligned}&\sum_m{}'\{{}_pU^*_{m_1},({}_sU_{t_2})\}_t\\&\quad=-\sum_{x=p,s}\beta_x\sum_m{}'\int_c{}^* e^{-i\mathbf{k}^*_{xc}\cdot\mathbf{R}_{tm}}({}_p\hat{\boldsymbol{\epsilon}}^*_{m_1:x})\\&\quad\cdot{}_sG_{t:x}(\hat{\mathbf{r}}_c^*,\hat{\mathbf{k}}_2)[\,{}_pG^*_{m_1:x}(\hat{\mathbf{r}}_c)].\end{aligned}\tag{61}$$

The p–s forms of (49) are

$$\begin{aligned}\{{}_pU^*_{t_1},({}_sU_{t_2})\}_t&=\{{}_pU^*_{t_1},({}_sU_{t_2})\}_\infty\\&=-4\pi\omega^2i\rho\Big(\sum_{x=p,s}\int_S k_x^{-3}{}_pG^*_{t:x}(\hat{\mathbf{r}},\hat{\mathbf{k}}_1)\\&\quad\cdot{}_sG_{t:x}(\hat{\mathbf{r}},\hat{\mathbf{k}}_2)d\Omega_{\hat{\mathbf{r}}}\Big)\\&=-(2\omega^2i\rho/k_p)_{(p,s)}S_{\mathrm{scat}}(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2),\end{aligned}\tag{62}$$

and

$${}_{(p,s)}S_{\mathrm{scat}}(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2)=2\pi k_p\Big(\sum_{x=p,s}\int_S k_x^{-3}{}_pG^*_{t:x}(\hat{\mathbf{r}},\hat{\mathbf{k}}_1)\cdot{}_sG_{t:x}(\hat{\mathbf{r}},\hat{\mathbf{k}}_2)d\Omega_{\hat{\mathbf{r}}}\Big).\tag{63}$$

In a manner similar to (50) and (51), we transform (52) and (53) into

$$\begin{aligned}\{{}_pU^*_{t_1},({}_s\phi_2)\}_t&=-\{{}_s\phi_2^*,({}_pU_{t_1})\}_t^*\\&=-\beta_s\hat{\boldsymbol{\epsilon}}_{s_2}\cdot{}_pG^*_{t:s}(\hat{\mathbf{k}}_2,\hat{\mathbf{k}}_1),\end{aligned}\tag{64}$$

and

$$\begin{aligned}\{{}_pU^*_{t_1},\sum_m{}'{}_sU_{m_2}\}_t&=-\sum_m{}'\{{}_sU^*_{m_2},({}_pU_{t_1})\}_t^*\\&=-\sum_{x=p,s}\beta_x\sum_m{}'\int_c e^{i\mathbf{k}_{xc}\cdot\mathbf{R}_{tm}}({}_s\hat{\boldsymbol{\epsilon}}_{m_2:x})\\&\quad\cdot{}_pG^*_{t:x}(\hat{\mathbf{r}}_c^*,\hat{\mathbf{k}}_1)[\,{}_sG_{m_2:x}(\hat{\mathbf{r}}_c)].\end{aligned}\tag{65}$$

Hence, the general p–s "energy theorem" is obtained by combining (60), (61), (62), (64), and (65) with (59). We have

$$\begin{aligned}&\left[\frac{4\pi}{k_p^3}\hat{\mathbf{k}}_1\cdot{}_sG_{t:p}(\hat{\mathbf{k}}_1,\hat{\mathbf{k}}_2)+\frac{4\pi}{k_s^3}\hat{\boldsymbol{\epsilon}}_{s_2}\cdot{}_pG_{t:s}(\hat{\mathbf{k}}_2,\hat{\mathbf{k}}_1)\right]+\frac{4\pi}{k_p^3}\left[\sum_m{}'\int_c{}^* e^{-i\mathbf{k}^*_{pc}\cdot\mathbf{R}_{tm}}({}_p\hat{\boldsymbol{\epsilon}}^*_{m_1:p})\cdot{}_sG_{t:p}(\hat{\mathbf{r}}_c^*,\hat{\mathbf{k}}_2)[\,{}_pG^*_{m_1:p}(\hat{\mathbf{r}}_c)]\right.\\&\quad+\sum_m{}'\int_c e^{i\mathbf{k}_{pc}\cdot\mathbf{R}_{tm}}({}_s\hat{\boldsymbol{\epsilon}}_{m_2:p})\cdot{}_pG^*_{t:p}(\hat{\mathbf{r}}_c^*,\hat{\mathbf{k}}_1)[\,{}_sG_{m_2:p}(\hat{\mathbf{r}}_c)]\Big]+\frac{4\pi}{k_s^3}\left[\sum_m{}'\int_c e^{-i\mathbf{k}^*_{sc}\cdot\mathbf{R}_{tm}}({}_p\hat{\boldsymbol{\epsilon}}^*_{m_1:s})\cdot{}_sG_{t:s}(\hat{\mathbf{r}}_c^*,\hat{\mathbf{k}}_2)[\,{}_pG^*_{m_1:s}(\hat{\mathbf{r}}_c)]\right.\end{aligned}$$

$$+\sum_m{}' \int_c e^{ik_{sc}\cdot R_{tm}}({}_s\hat{\epsilon}_{m_2;s})\cdot{}_pG^*_{t;s}(\hat{r}^*_c,\hat{k}_1)\,[{}_sG_{m_2;s}(\hat{r}_c)]\Big] = -\frac{2}{k_p}\,{}_{(p,s)}S_{scat}(\hat{k}_1,\hat{k}_2). \tag{66}$$

The general s–p "energy theorem" is deduced from (66), giving

$$\left[\frac{4\pi}{k_s^3}\,\hat{\epsilon}_{s_1}\cdot{}_pG_{t;s}(\hat{k}_1,\hat{k}_2)+\frac{4\pi}{k_p^3}\,\hat{k}_2\cdot{}_sG^*_{t;p}(\hat{k}_2,\hat{k}_1)\right]+\frac{4\pi}{k_p^3}\Big[\sum_m{}'\int_c{}^* e^{-ik^*_{pc}\cdot R_{tm}}({}_s\hat{\epsilon}^*_{m_1;p})\cdot{}_pG_{t;p}(\hat{r}^*_c,\hat{k}_2)\,[{}_sG^*_{m_1;p}(\hat{r}_c)]$$
$$+\sum_m{}'\int_c e^{ik_{pc}\cdot R_{tm}}({}_p\hat{\epsilon}_{m_2;p})\cdot{}_sG^*_{t;p}(\hat{r}^*_c,\hat{k}_1)\,[{}_pG_{m_2;p}(\hat{r}_c)]\Big]+\frac{4\pi}{k_s^3}\Big[\sum_m{}'\int_c{}^* e^{-ik^*_{sc}\cdot R_{tm}}({}_s\hat{\epsilon}^*_{m_1;s})\cdot{}_pG_{t;s}(\hat{r}^*_c,\hat{k}_2)\,[{}_sG^*_{m_1;s}(\hat{r}_c)]$$
$$+\sum_m{}'\int_c e^{ik_{sc}\cdot R_{tm}}({}_p\hat{\epsilon}_{m_2;s})\cdot{}_sG^*_{t;s}(\hat{r}^*_c,\hat{k}_1)\,[{}_pG_{m_2;s}(\hat{r}_c)]\Big]=-\frac{2}{k_s}\,{}_{(s,p)}S_{scat}(\hat{k}_1,\hat{k}_2). \tag{67}$$

Equations (66) and (67) show that

$${}_{(p,s)}S_{scat}(\hat{k},\hat{k}) = {}_{(s,p)}S_{scat}(\hat{k},\hat{k}) = 0 \tag{68}$$

in the forward direction, since each component of (65) and (66) contains a mixed scattering amplitude which is null for $\hat{k}_1 = \hat{k}_2$. When the configuration is reduced to a single scatterer, Eqs. (65) and (66) correspond to the p–s, s–p reciprocity relations of Dassios *et al.* (8:80, 81) or Varatharajulu (6:3.13).

[1]P. J. Barratt and W. D. Collins, Proc. Cambridge Philos. Soc. **61**, 969 (1965).
[2]D. S. Jones, Proc. Cambridge Philos. Soc. **48**, 736 (1952).
[3]T. H. Tan, J. Acoust. Soc. Am. **61**, 928 (1977).
[4]T. H. Tan, J. Acoust. Soc. Am. **60**, 1265 (1976).
[5]A. T. De Hoope, Appl. Soc. Res. Sec. B **7**, 463 (1954).
[6]V. Varatharajulu, J. Math. Phys. **18**, 537 (1977).
[7]D. S. Jones, *The Theory of Electromagnetism* (Pergamon, Oxford, 1964).
[8]G. Dassios, K. Kiriaki, and D. Polysos, Quart. Appl. Math. (to be published).
[9]V. Twersky, J. Math. Phys. **3**, 83 (1962).
[10]V. Twersky, J. Math. Phys. **8**, 589 (1967).
[11]C. F. Ying and R. Truell, J. Appl. Phys. **27**, 1086 (1956).
[12]N. G. Einspruch and E. J. Witterholt, J. Appl. Phys. **31**, 806 (1960).
[13]N. G. Einspruch and R. Truell, J. Acoust. Soc. Am. **32**, 214 (1960).
[14]Y-H. Pao and C. C. Mow, J. Appl. Phys. **34**, 493 (1963).
[15]G. Dassios and K. Kiriaki, Quart. Appl. Math. **42**, 225 (1984).
[16]V. Twersky, J. D'Analyse Math. **30**, 498 (1976).
[17]V. Twersky, J. Math. Phys. **18**, 2468 (1977).
[18]D. D. Phanord, "Multiple scattering of elastic waves by a distribution of identical spheres," Ph.D. thesis, University of Illinois at Chicago (1988).
[19]V. Twersky, J. Acoust. Soc. Am. **53**, 96 (1973).
[20]P. J. Barratt, Proc. Cambridge Philos. Soc. **66**, 469 (1969).
[21]P. M. Morse, and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Chap. 13, p. 1987.
[22]F. Noether, *In Theory of Functions*, edited by R. Rother, F. Ollendorf, and K. Polhausen (Technology, Cambridge, MA, 1948), pp. 167–170.
[23]A. Sommerfeld, *Partial Differential Equations in Physics* (Academic, New York, 1949).
[24]V. D. Kupradze, "Dynamical Problems in Elasticity," in *Progress in Solid Mechanics III* (North-Holland, Amsterdam, 1963).

Nonlinear corrections to the Schrödinger equation from geometric quantum mechanics

Carlos Castro
Physics Department, Center for Particle Theory, University of Texas at Austin, Austin, Texas 78712

(Received 22 February 1989; accepted for publication 5 July 1990)

The nonlinear corrections to the Schrödinger equation based on the results of geometric quantum mechanics (GQM) are derived. Such theory derives quantum mechanics from the underlying Weyl space-time geometry associated with the classical ensemble of particle paths given by the solutions of the Hamilton–Jacobi equation.

I. INTRODUCTION

Geometric quantum mechanics[1,2] (GQM) is a theory that attempts to describe all of the quantum mechanical phenomena based on the underlying Weyl space-time geometry associated with the classical statistical ensemble of particle paths (linked to a particle) given by the solutions of the Hamilton–Jacobi equation. (Of course that one requires the Weyl symmetry to be broken at a certain length scale to yield a Riemannian space-time at scales $l > l_w$; where $l_w =$ scale at which the Weyl symmetry is broken.) The ensemble density acts as a true curvature source that affects the particle trajectories and, in turn, the particle trajectories affect the curvature which, in turn, depends on the ensemble density, and so forth. It is as a result of this nonlinear mechanism that the mass of the particle acquires an explicit and physical dependence on the Weyl scalar curvature. The relationship between the Weyl gauge potential and the ensemble density is obtained from a variational principle that yields the origins of the wave particle duality.[2] When one imposes the condition that the ensemble average of the line integral (given by the worldline of the particle's path) must be related to the average of the Einstein–Hilbert action one obtains the desired relationship between R and ρ (see Refs. 1 and 2). Such a relationship is nothing but the celebrated Bohm quantum potential:

$$\frac{1}{4}\left(\frac{D-2}{D-1}\right)R = -\frac{1}{\sqrt{\rho}}\Box(\sqrt{\rho}) \qquad \begin{pmatrix}\text{disregarding the Riemannian}\\ \text{component in } R_{\text{Weyl}}\end{pmatrix}, \tag{1}$$

which is obtained after the substitution $\phi_\mu = [2/(D-2)]\partial_\mu(\ln \rho/m)$ is made in Weyl's expression for the scalar curvature in any number of dimensions:

$$R = R_{\text{Riemann}} - [(D-1)(D-2)/4]\phi^\mu\phi_\mu - (D-1)\partial_\mu\phi^\mu, \tag{2}$$

$\rho \equiv$ ensemble density,

$\Box \equiv -\partial^2_t + \nabla^2$.

For convenience we give the expression to be used later:

$$R_s/6 = -(1/\sqrt{\rho})\nabla^2(\sqrt{\rho})$$

$\equiv$ Spatial component of the Weyl scalar curvature in $D=4$,

where ϕ_μ is Weyl's gauge potential that gauges dilations.[1,2] The derivation of the Schrödinger equation can be obtained from the nonrelativistic limit of the Hamilton–Jacobi equation

$$(\partial_\mu S)^2 + m^2 = 0. \tag{3}$$

It is the purpose of this paper to explicitly derive the first-order nonlinear corrections to the Schrödinger equation. There has recently been work on the precision tests of quantum mechanics and on the limitations that nonlinear corrections might impose on it.[3] Today we have at our disposal experimental techniques that are perfectly suited to test the theory of geometric quantum mechanics. This is the motivation of this work.

II. NONLINEAR CORRECTIONS TO THE SCHRÖDINGER EQUATION

We are going to give a very simple derivation of the nonlinear terms that appear in the most general version of the Schrödinger equation. Since we perform the nonrelativistic limit of the Hamilton–Jacobi equation we must first introduce to the reader our conventions. We work in a four-dimensional Weyl space-time, although our derivation works in any dimension as well. The signature is a Lorentzian one: $(-,+,+,+)$. We work in a Riemannian-flat background but *not* on a Weyl flat one. After all, our aim is to reproduce all of our equations in a flat Minkowski space. This is hinting to us that the metric of space-time is a passive physical entity in GQM and it seems no longer farfetched that it is plausible that for all these years we have been asking a totally meaningless question: Can geometry be quantized? (There has been work in the past where gravity appears as a result of a spontaneous symmetry breakdown of a underlying metric-affine theory.) GQM shows that it is the *Weyl affine connection*, combined with classical statistical mechanics, which is responsible for quantum effects and that the metric has nothing to do with it at all. Greek indices run over 0,1,2,3. $\hbar = c = 1$.

The relativistic Hamilton–Jacobi equation for a particle is[1,2]

 0022-2488/90/112633-04$03.00

$$(p_\mu)^2 + m^2(R) = 0, \tag{4}$$

where the mass function is[1,2]

$$m^2(R) = m_0^2 + \tfrac{1}{6}R. \tag{5}$$

The nonrelativistic limit is performed as usual:[2]

$$(p^0)^2 = (\mathbf{p})^2 + m_0^2 + \frac{1}{\sqrt{\rho}}\frac{\partial^2\sqrt{\rho}}{\partial t^2} - \frac{1}{\sqrt{\rho}}\nabla^2\sqrt{\rho}.$$

Neglecting the time derivative:

$$(p^0)^2 \approx (\mathbf{p})^2 + m_0^2 + \frac{R_s}{6} = m_0^2\left(1 + \frac{(\mathbf{p})^2}{m_0^2} + \frac{R_s}{6m_0^2}\right). \tag{6}$$

The binomial expansion yields:

$$p^0 - m_0 \approx \frac{1}{2m_0}(\mathbf{p})^2 + \frac{R_s}{12m_0} - \frac{m_0}{8}\left(\frac{\mathbf{p}^2}{m_0^2} + \frac{R_s}{6m_0^2}\right)^2 + \cdots$$

$$[\text{valid for } -1 < (\hbar^2/m_0^2c^2)R < 1].$$

If we wish to reproduce the standard form of the Schrödinger equation we just keep the terms linear in R only:

$$p^0 - m_0 \approx (1/2m_0)(\mathbf{p})^2 + (1/12m_0)R_s$$

$$= \text{effective kinetic energy}.$$

The Hamilton–Jacobi equation is

$$\Rightarrow -\frac{\partial S}{\partial t} = \frac{1}{2m_0}(\nabla S)^2 - \frac{1}{2m_0}\frac{1}{\sqrt{\rho}}\nabla^2(\sqrt{\rho}). \tag{7}$$

The Schrödinger equation is easily obtained after we use the explicitly Weyl-noncovariant form of the continuity equation for the probability current of the particle:[2]

$$\nabla_a\left(\frac{\rho}{m_0}\nabla^a S\right) = -\frac{\partial\rho}{\partial t} \quad (a = 1,2,3) \tag{8}$$

and, after the substitution $\psi = \sqrt{\rho}\exp(iS)$ is made:

$$i\frac{\partial\psi}{\partial t} = -\frac{1}{2m_0}\nabla^2\psi. \tag{9}$$

Notice that it is upon the use of Eq. (8) that the underlying Weyl symmetry is explicitly broken. Had we used an explicit Weyl covariant continuity equation we would have obtained Weyl covariant expressions. In particular, the norm$^2 = |\psi^*\psi|$ could have been gauged to a constant value as a result of the Weyl gauge degrees of freedom after imposing the gauge:

$$\phi_\mu = \partial_\mu \ln(\rho/m_0) = 0,$$

$$\Rightarrow \rho/m_0 = \text{constant in units of } (\hbar/m_0c)^{-2}. \tag{10}$$

(We know that at the atomic scales there is no Weyl symmetry any longer since the spectrum of the atoms is independent on the *locations* of these. Clearly $l_w < l_{\text{atomic}}$.) For more details see Ref. 2. [The connection between the breakdown of the Weyl symmetry and the use of Eq. (8) will be explained in a future paper.]

It is a straightforward exercise to show that upon the inclusion of the curvature square terms one has

$$i\frac{1}{\psi}\frac{\partial\psi}{\partial t} = -\frac{1}{2m_0}\frac{1}{\psi}\nabla^2\psi - \frac{1}{8m_0}\left[i\frac{\partial}{\partial t}\ln|\psi^*\psi| + \frac{1}{m_0\psi}\nabla^2\psi\right]^2. \tag{11}$$

Also straightforward is the inclusion of an external potential V in the rhs of the above equation. In presence of electromagnetic interactions, the prescription $\mathbf{p} - ie\mathbf{A}$ applies also. This inclusion in the canonical momentum will clearly reflect the effects of the nonlinear corrections through the presence of $\mathbf{A}^4$ terms. For consistency purposes we require that the solutions to Eq. (11) do in fact maintain $(\hbar/m_0c)^2|R| \leqslant 1$.

Above, the continuity equation has been used and the following identity:

$$i\frac{1}{\psi}\frac{\partial\psi}{\partial t} = \frac{i}{2}\frac{\partial\ln\rho}{\partial t} - \frac{\partial S}{\partial t},$$

which is obtained from

$$\psi = \sqrt{\rho}\exp(iS) \tag{12}$$

in the explicit evaluation of

$$\frac{\partial}{\partial t}[\sqrt{\rho}(\exp iS)].$$

The form of Eq. (11) is homogeneous as mentioned by Weinberg:[3] Under $\psi \to \lambda\psi$; Eq. (11) is invariant and it has the form

$$i\frac{\partial\psi}{\partial t} = \frac{\partial h(\psi^*,\psi,t)}{\partial\psi^*},$$

which is more general than the case studied by Weinberg. (Weinberg studied the special case that $\partial|\psi^*\psi|/\partial t = 0$ and also the special case where the Hamiltonian had no explicit time dependence. Our result is thus more general. In this work $|\psi^*\psi|$ does *not* mean the integrated norm but simply the probability.)

It is up to the experiments to verify the validity of Eq. (11). At first sight, the simplest thing to look for is corrections of the type $\mathbf{A}^4$ in the presence of electromagnetic fields.

III. CONCLUSION

We have explicitly derived the first-order nonlinear corrections to the standard Schrödinger equation directly from the theory of geometric quantum mechanics. Such corrections agree with the prescriptions of homogeneity given by Weinberg. Experimental confirmation of these results is warranted to test the validity of this theory. The homogeneity condition can be very simply understood as remnants of the Weyl invariance. This indicates that quantum effects could be derived from an underlying Weyl invariance. For more details concerning this we refer the reader to Refs. 1 and 2. The consequences of this theory could be very important in all realms of physics, esspecially in defining quantum chaos, etc. The subject of current work under investigation deals with the vigorous study of

how the Weyl symmetry is broken down to yield a Riemannian space-time. Notice that the expansion of Eq. (6) is performed in terms of a dimensionless parameter

$$-1<\tfrac{1}{6}(\hbar/m_0c)^2R_{\rm Weyl}<1.$$

We know that Planck's length is basically obtained when the Compton wavelength of a particle of mass $M_{\rm Planck}$ is of the order of the Schwarzchild radius $r_s\equiv 2GM/c^2$. Clearly, at r_s:$R_{\rm Riemann}\approx 1/r_s^2$.

Therefore, it is natural to expect that the signals of Weyl invariance begin to show up at M_p for a particle of mass equal to M_p. Numerical solutions to Eq. (11) are currently under investigation for the simplest models. It is *there* that we should look for departures from quantum mechanics if, indeed, they exist. Therefore, it is *not* necessary to probe the Planck's length regime to begin to see the departures from linear quantum mechanics. However, if we wish to observe these departures we require to perform experiments with an extreme precision as pointed out by Weinberg.[3] Also, we may depart from the nonrelativistic regime to cross the relativistic and even ultrarelativistic regime. Since it is difficult to set *bounds* on $R_{\rm Weyl}$ it is far more convenient to solve Eq. (11) and *then* to look for an experimental confirmation in our current laboratories and accelerators without having to probe the Planck scale regime.

ACKNOWLEDGMENTS

I would like to thank Longhorn Copies for typing the manuscript and M. Marder for his help in the completion of some parts of the Appendix.

APPENDIX: THE HAMILTONIAN FUNCTION AND FURTHER TOPICS

We are going to present the explicit form of the Hamiltonian function that renders Eq. (11) in the form

$$i\frac{\partial\psi}{\partial t}=\frac{\partial}{\partial\psi^*}h(\psi^*,\psi,t).$$

In our case the Hamiltonian does not have an explicit time dependence although, in general, it might have. The energy in our case is thus a constant of the motion. The Hamiltonian h consists of two parts (set $\hbar=c=1$):

$$h_0\equiv-(1/2m_0)\psi^*\nabla^2\psi,$$

$$h_1\equiv-\frac{1}{8m_0}\left\{(\psi^*\psi)\left(\frac{\partial}{\partial t}\ln(\psi^*\psi)\right)^2+\frac{i}{m_0}(\psi^*\nabla^2\psi)\frac{\partial}{\partial t}\left(\ln\frac{\psi^*\psi}{\|\psi^*\psi\|}\right)^2+\frac{1}{m_0^2}\left[\frac{1}{\psi^*\psi}(\psi^*\nabla^2\psi)^2\right]\right\}.$$

(One should add the Hermitean conjugate to h in order to have a real-valued expectation value for h. On the other hand, our results predict a nonlinear-induced decay for a "stable" particle through the presence of a small imaginary component in the energy. Of course, stationary state solutions would *not* suffer from this. Since the scope of a relativistic Hamiltonian framework is very limited we are assuming that Hamilton's equations hold and that the Hamiltonian coincides with the energy in our case which is *not* strictly a purely nonrelativistic one.) Since h is homogeneous of degree 1 (both in ψ and ψ^*):$\psi\to\lambda\psi$;$h\to\lambda h$. One must properly introduce a suitable normalization in the term

$$\frac{\partial}{\partial t}\left(\ln\frac{\psi^*\psi}{\|\psi^*\psi\|}\right)^2,$$

where $\|\psi^*\psi\|$ is the *integrated* norm which we can choose to be equal to 1. We remind the reader that our expression $|\psi^*\psi|$ is simply equal to ρ and *not* equal to the total integrated norm as in the case discussed by Weinberg.

We turn attention to the presence of nodes. Notice that a node for ψ implies a singularity in the expression for R:

$$R\approx(1/\sqrt{\rho})\nabla^2(\sqrt{\rho})$$

and the expansion in Eq. (6) *breaks* down. However, it is important to realize that nodes can in fact be present. The

$$i\frac{1}{\psi}\frac{\partial\psi}{\partial t}$$

term in the lhs of Eq. (11) is balanced against the term

$$-\frac{1}{2m_0}\frac{1}{\psi}\nabla^2\psi.$$

Similarly, the term (purely imaginary)

$$i\frac{\partial}{\partial t}\ln(\psi^*\psi)$$

is balanced against the term

$$\frac{1}{m_0}\frac{1}{\psi}\nabla^2\psi=2\left(\frac{1}{2m_0}\frac{1}{\psi}\nabla^2\psi\right)$$

in the rhs of Eq. (11). If nodes are to be present for the wave function ψ the term

$$\frac{1}{\psi}\frac{\partial\psi}{\partial t}$$

must approach infinity along the *real* axis whereas the term

$$\frac{1}{2m_0}\frac{1}{\psi}\nabla^2\psi$$

must approach infinity along the *imaginary* axis. (If we wish to have nodes for stationary states then we must require that ψ has an inflection point at the node, i.e., $\nabla^2\psi$ is zero at such node.) Notice the crucial importance of the *numerical* coefficients in the nonlinear terms of Eq. (11). It is fairly obvious that a plane-wave solution (a stationary solution) satisfies Eq. (11). We could try to plug-in a Gaussian-wave packet

$$\psi(x,t) \approx \left[1 + \frac{i\hbar t}{2(\Delta x)_0^2 m}\right]^{-1/2} \exp\left(-i\frac{Et}{\hbar} + \frac{ipx}{\hbar}\right)$$
$$\times \exp\left[\frac{-x^2}{4(\Delta x)_0^2(1 + i\hbar t/2(\Delta x)_0^2 m)}\right]$$

and see to what degree it satisfies Eq. (11) or gets close to satisfying it. We know that it satisfies the ordinary Schrödinger equation and it would be interesting to see how approximate a solution to Eq. (11) this Gaussian can be.

What about Galilean invariance? We should expect it in the absence of *external* potentials which might break it. However, Eq. (6) contains an explicit factor of $1/c^2$ in the R^2 term once we insert the suitable c and $\hbar$ factors. (There is an explicit cancellation of c in the term linear in R.) Since the Galilean algebra is the Wigner–Ionu contraction of the Poincaré algebra [ISO (3,1)] as a result of taking the $c \to \infty$ limit, it is fairly clear that the $1/c^2$ terms in Eq. (11) are not going to be consistent with this Galilean limit. The ordinary Schrödinger equation is Galilean invariant although, in general, the introduction of external potentials can break the Galilean invariance. Galilean invariance is surely not a true symmetry of nature. Lorentz invariance is. Weinberg[3] has failed to construct an explicit homogeneous, nonbilinear realization of the generators of the Galilean algebra, in particular, the Hamiltonian function (the generator of time translations). This, of course, does not imply that it is impossible to do so. However, if we could prove that there are *no* such homogeneous, nonbilinear realizations of the Galilean algebra this would surely be a signal that Galilean invariance should not be a cornerstone in the formulation of nonlinear quantum mechanics. It is true that our results are *not* strictly nonrelativistic! Geometric quantum mechanics predicts nonlinear corrections to quantum mechanics—not only in the relativistic regime (see Ref. 2)—but also in the expansion of Eq. (6). True, a $1/c^2$ appears because we rely on a geometric formulation and, hence, the speed of light and the curvature are natural ingredients in our theory. Notice that we were able to neglect the term

$$\frac{1}{c^2}\frac{1}{\sqrt{\rho}}\frac{\partial^2(\sqrt{\rho})}{\partial t^2}$$

in Eq. (6) with respect to

$$(1/\sqrt{\rho})\nabla^2(\sqrt{\rho}),$$

since the latter should be of the same order of magnitude as

$$\frac{1}{c}\frac{1}{\sqrt{\rho}}\frac{\partial(\sqrt{\rho})}{\partial t}.$$

For logarithmic nonlinear corrections to the relativistic Klein–Gordon equation and the Dirac equation, see Ref. 2. There is a whole other list of requirements that we must satisfy:[4] to see whether we have invariance under space and time reflections; conservation of total momentum and angular momentum for isolated systems; the solutions should satisfy the Ehrenfest theorem:

$$m\frac{d}{dt}\langle \mathbf{r}\rangle = \langle \mathbf{p}\rangle,$$

$$\frac{d}{dt}\langle \mathbf{p}\rangle = \langle -\nabla U\rangle$$

(where U is a potential). Also the principle of *separability* of noninteracting subsystems should hold. In general, nonlinear equations introduce correlations even for noninteracting subsystems. This is not highly desirable because no physical predictions could be made if the rest of the world, in the absence of forces, influences the detailed behavior of an isolated particle.

These and more issues are the subject of further investigation.

[1] E. Santamato, Phys. Rev. D **29**, 216 (1984); Phys. Rev. D **32**, 2615 (1985); Phys. Lett. A **130**, 199 (1988).

[2] C. Castro, "On Weyl Geometry, Random Processes and Geometric Quantum Mechanics," to appear in Foundations Phys.; Texas preprint: DOE#111-87.

[3] S. Weinberg, "Precision Tests of Quantum Mechanics," UTTG#29-88 and "Particle States as Realizations (Linear or Nonlinear) of Space Time Symmetries," UTTG#15-88.

[4] I. Bialynicki-Biruli and J. Mycielski, Ann. Phys. **100**, 62 (1976).

Quantum mechanics with higher time derivatives: Some mathematical facts and the inclusion of gravitational self-interaction

Alla D. Popova
All-Union Center for Surface and Vacuum Research, Andreyevskaya Naberezhnaya 2, Moscow 117334, USSR

(Received 1 June 1989; accepted for publication 6 June 1990)

The inclusion of self-interaction, in particular, gravitational into the quantum mechanics of a particle (first quantization) implies two necessary features: dealing with quantum mechanics involving the second time derivative and the construction of a variational principle maintaining the rescaling invariance of the wave function. The relevant variational principle is constructed which is not the principle of extremality of any action. Some general mathematical facts lain in the basis of this construction are discussed.

I. INTRODUCTION

In Ref. 1, it was shown that, in a stationary situation, it is possible to include gravitational self-interaction into the quantum mechanics of a particle without unsolvable problems. The inclusion of full gravity, without any approximations, suggests one to consider the relativistic Klein–Gordon equation. This explains exactly the author's interest to quantum mechanics whose evolution equation contains higher-order time derivatives. Certainly, the Klein–Gordon equation is the most primitive example of such an equation. However, although we do not know realistic situations in which an equation with third- or higher-order time derivatives describes the evolution of a quantum system, we consider theories[2] with arbitrary-order time derivatives because all of them possess common features better understood in the general form. Note that the usual Schrödinger theory with the first time derivative is a somewhat degenerate case.

There exists an objection against the probability interpretation of the Klein–Gordon equation founded on the observation that this equation provides no positive definite probability density. Generally, with a norm square defined as a functional whose conservation follows from the evolution equation, the main "physical" trouble of quantum mechanics with higher time derivatives is that the norm square may acquire not only positive values. The definition that can be chosen for the norm square is, however, not unique: There are diverse conserved functionals. For instance, for the Klein–Gordon theory with electromagnetic interaction, different definitions were made in Refs. 3 and 4 on the one hand and Ref. 5 on the other hand.

Following the definition of Refs. 3 and 4, the norm square acquires all real values: positive, negative, and zero. (In the latter case, the so-called associated vectors in the Hilbert space of states emerge.[6,7]) Restriction to solutions with the positive norm square only, their superposition is unable to form a general solution; thus some modification of the probabilistic description is required. The interpretation of the Klein–Gordon equation for a charged relativistic particle in an external electromagnetic potential was discussed in Ref. 3 where a two-component formalism was constructed in a Hamiltonian form symmetric in the wave function and its extended time derivative (involving the electric potential) of the first order. The difficulty of the negative norm square was avoided[3] by considering, instead of eigenvalues of energy, some energy expectation values always having a positive sign to be observables. In Ref. 4, a quantum system of two spinless charged particles in the cases of scalar and relativistic Coulomb interactions was investigated, the problems of normalization were discussed in the context of general stationary wave equations, and an example of an associated vector for a charged particle in an electric potential well was given. In the theory of the Bethe–Salpeter equation describing the relativistic two-body problem in quantum field theory, a similar normalization arises (the physical literature devoted to this subject is very volumenous, see, for review, Ref. 8). This is also the case for the quantum-relativistic many-particle system in bound states.[9]

Another definition for the norm square as an "energy integral" in Ref. 5 is always positive for free particles and remains positive for a wide class of physically meaningful potentials. Another version of a two-component formalism was constructed[5] with a Hamiltonian that is self-adjoint in a Hilbert space with positive metric.

The inclusion of self-interaction, in particular, gravitational, into quantum mechanics encounters a lot of problems, and the most principal of them is exposed here. The theory of any complex field Ψ with linear in Ψ equations may be attempted to interprete as quantum mechanical, Ψ being a wave function. The cause is that linear equations are invariant under the rescaling transformation

$$\Psi \to a\Psi, \quad + \text{complex conjugate}, \tag{1.1}$$

where a is an arbitrary complex number with the modulus generally unequal to unity.[10] This invariance allows one to supply the norm square (full probability) with a fixed value, for example unity, if it is possible. But the ordinary action bilinear in Ψ^* and Ψ used in the field theories and leading to linear equations which looks like[11]

$$\bar{J}_{(2)} = -\int d^4x(-\det g_{\mu\nu})^{1/2}(\Psi^*_{,\lambda}\Psi^{,\lambda} + \mu^2\Psi^*\Psi), \tag{1.2}$$

for the Klein–Gordon theory, is not invariant under (1.1), so, if one added the Einstein–Hilbert action to (1.2), the

 0022-2488/90/112637-16$03.00

resulting equations would not be invariant under (1.1). In order to maintain probability interpretation when including self-interaction, the candidate action must be invariant under (1.1), guaranteeing that the norm alteration in the process of interaction is not catastrophic. Such a variational principle exists, but, unfortunately, for discrete stationary states only.

For wave functions with exponential time dependence (stationary states) the evolution equation with higher time derivatives is reduced to an eigenvalue problem with nonlinear dependence on the spectral parameter. A variational principle leading to such a problem was discovered by Rayleigh in 1873 (Ref. 12) for a particular case of quadratic dependence in the context of classical dynamics. Later, it was extended to the cases of arbitrary quadratic dependence[13] and general polynomial dependence.[14] A great amount of purely mathematical studies were carried out on this subject; the relevant references and a subsequent account on the subject can be found in Ref. 15. In the Schrödinger quantum mechanics, this principle is known in its trivial form.[16]

The Rayleigh functionals in the role of some (we add eigen-) actions, as well as stationary equations and external currents derived from them, are invariant under (1.1); the external currents are equal to stationary bilinear ones divided by the norm square. However, the time dependence is inevitably lost.

We now ask the question of whether a variational principle exists combining the positive features of the two above principles, namely, time dependence and invariance under (1.1)? In other words, what variational principle could give us equations and external currents that are both time dependent and rescaling invariant? The answer is negative in the sense that there exists no suitable variational principle of extremality for any action. However, nobody forbids us to abandon the extremality principle and to try to construct a variational principle of averaged extremals for some action that possesses some desired qualities and resolves partly to the problem. This is the main program of this paper; to accomplish it we need certain mathematical constructions. Although we often refer to the previous paper[1] where many physical questions were discussed, all the necessary information is presented here.

The paper proceeds as follows. First of all, we devote Secs. II and III to the properties of general solutions to the evolution equation with higher time derivatives. A part of the results known earlier is adduced in a general form. In Sec. II, the basic concepts, namely, equations, operators, and bilinear quantities are introduced. Some conserved functionals are derived, and the simplest of them allows us to define a norm which is appropriate for our purposes. The revision of the norm concept entails the revision of expectation values that we define in our own way having the advantage of the definition in Ref. 3. We also give necessary and sufficient conditions for the modified expectation values to be conserved in time.

Unlike Ref. 3, in Sec. III, the two-component formalism of the Klein–Gordon theory is constructed more generally in a nonsymmetric form and further extended to the n-component formalism for a theory with the evolution equation containing time derivatives of orders up to n. Our method allows us to see what quantity plays the role of a (matrix) Hamiltonian and to demonstrate correspondence to the results of Sec. II.

Stationary states are considered in Secs. IV–VI. In Sec. IV, the Rayleigh variational principle and functionals are described and applied to the Schrödinger and Klein–Gordon theories. In Sec. V, we study the criterion for each of the Rayleigh functionals associated with the same evolution equation to have coinciding sets of eigenfunctions and discuss modified orthogonality properties for eigenfunctions with different eigenvalues. In Sec. VI, the results of Secs. IV and V are reproduced in and compared with the n-component formalism.

In Sec. VII, we describe our receipt for the formulation of variational principles of averaged extremals without and with gravity.

II. EQUATIONS, BILINEAR ACTION, CONSERVED FUNCTIONALS, AND EXPECTATION VALUES

Let a particle be described by a wave function defined in the Minkowski or Riemann space-time on the time semiaxis $0 \leqslant t < \infty$ and denoted by $\Psi(t,\mathbf{x})$ [or sometimes $\Phi(t,\mathbf{x})$]. Assume that at each fixed moment of time t_f wave functions, e.g., $\Psi(t,\mathbf{x})$, and all their time derivatives up to some order m ($m \leqslant n$), e.g.,

$$\left(i\frac{\partial}{\partial t}\right)^m \Psi(t,\mathbf{x})|_{t=t_f},$$

which we need are given in a closed three-space domain Ω and belong to the complex Hilbert space $\mathscr{L}_2(\Omega)$ of quadratic integrable over Ω functions. The standard definition of the Hilbert space suggests the following inner product and norm:

$$\begin{aligned} &\langle \Phi^*(t_f,\mathbf{x}),\Psi(t_f,\mathbf{x})\rangle = \int d\Omega\, \Phi^*\Psi, \\ &\mathscr{N}_{\mathscr{L}_2} = (\langle \Psi^*,\Psi\rangle)^{1/2}. \end{aligned} \tag{2.1}$$

[At the same time (2.1) serves as the definition of the Dirac brackets to be used below, where $d\Omega$ is the invariant three-volume element, but unlike standard notations we prefer to reserve the asterisk for complex conjugate functions.] Generally, the quantities (2.1) are not conserved in time.

Let $\Psi(t,\mathbf{x})$ satisfy the evolution equation

$$\hat{A}_{(n)}\left(i\frac{\partial}{\partial t}\right)\Psi = 0, \tag{2.2}$$

$\hat{A}_{(n)}$ being a formal differential operator of the order n linear in Ψ:

$$\hat{A}_{(n)}\left(i\frac{\partial}{\partial t}\right) = \sum_{k=0}^{n} \hat{A}_k \left(i\frac{\partial}{\partial t}\right)^k, \tag{2.3}$$

where formal differential operators $\hat{A}_k$ acting in $\mathscr{L}_2(\Omega)$ involve partial derivatives with respect to three-space coordinates $\mathbf{x}$ whose orders may, in principle, be unrestricted; the coefficients by these derivatives carry no dependence on time but may contain some smooth functions of $\mathbf{x}$, which we de-

note collectively by $\alpha(\mathbf{x})$ and their three-space derivatives of any order,

$$\hat{A}_k = \hat{A}_k(\{\alpha(\mathbf{x})\}).$$

Here and below, figure brackets mean that not only the function but also its derivatives of some order with respect to its argument enter the corresponding quantity. We consider only the operators $\hat{A}_k$ of this kind. If Eq. (2.2) is general relativistic, then each operator $\hat{A}_k$ should involve three-space derivatives of the order n-k. In principle, instead of (2.3) it is possible to consider nonpolynomial operators expanded into Taylor series if they are analytic functions of the argument $i\partial/\partial t$.

We will deal with a class of functions $\mathscr{L}_2'(\Omega)$, $\mathscr{L}_2'(\Omega) \subset \mathscr{L}_2(\Omega)$, which obey only such boundary conditions on $\partial\Omega$, the boundary of Ω, that each formal differential operator $\hat{A}_k$ can be associated with a uniquely defined bounded Hermitian operator denoted by the same symbol,

$$\langle \Phi^*, \hat{A}_k \Psi \rangle \equiv \langle \hat{A}_k^T \Phi^*, \Psi \rangle = \langle \hat{A}_k^* \Phi^*, \Psi \rangle,$$

or, formally,

$$\hat{A}_k = \hat{A}_k^{T*} = \hat{A}_k^\dagger.$$

Below we use variational (functional) derivatives of two kinds applied to the two kinds of functional performed by integration either over Ω or over Ω and t:

$$\underset{(3)}{\frac{\delta}{\delta a(\mathbf{x})}} \int d\Omega\, L(\{a(\mathbf{x})\}) = \frac{\partial L}{\partial a} - \left(\frac{\partial L}{\partial a_{,i}}\right)_{,i} + \left(\frac{\partial L}{\partial a_{,ik}}\right)_{,ik} - \cdots,$$

$$\underset{(4)}{\frac{\delta}{\delta a(t,\mathbf{x})}} \int dt\, d\Omega\, L(\{a(t,\mathbf{x})\}) = \frac{\partial L}{\partial a} - \left(\frac{\partial L}{\partial a_{,\mu}}\right)_{,\mu} + \left(\frac{\partial L}{\partial a_{,\mu\nu}}\right)_{,\mu\nu} - \cdots,$$

$$(i,k = 1,2,3; \mu,\nu = 0,1,2,3).$$

Consider an action functional bilinear in Ψ^* and Ψ at a finite time interval $[0,T]$:

$$\bar{J}_{(n)} = \frac{1}{2}\int_0^T dt \sum_{k=0}^{n} \left[\left\langle \hat{A}_k^* \left(-i\frac{\partial}{\partial t}\right)^k \Psi^*, \Psi \right\rangle + \left\langle \Psi^*, \hat{A}_k \left(i\frac{\partial}{\partial t}\right)^k \Psi \right\rangle\right]. \tag{2.4}$$

Equation (2.2) can be obtained from the variational principle stating the extremality of (2.4), the extremum condition being

$$\underset{(4)}{\frac{\delta \bar{J}_{(n)}}{\delta \Psi^*}} = 0. \tag{2.5}$$

The conjugate to (2.2) equation

$$\hat{A}_{(n)}^* \left(i\frac{\partial}{\partial t}\right)\Psi^* = 0, \tag{2.2'}$$

where the complex conjugate operator has the form:

$$\hat{A}_{(n)}^* \left(i\frac{\partial}{\partial t}\right) = \sum_{k=0}^{n} \hat{A}_k^* \left(-i\frac{\partial}{\partial t}\right)^k, \tag{2.3'}$$

can be obtained from the complex conjugate variational principle:

$$\underset{(4)}{\frac{\delta \bar{J}_{(n)}}{\delta \Psi}} = 0. \tag{2.5'}$$

[$\bar{J}_{(n)} = \bar{J}_{(n)}^*$ due to the integrand's form symmetric in time derivatives of Ψ^* and Ψ; this form is, however, not unique: Integrands that differ from that in (2.4) by a total time derivative are also possible.] Unless necessary, henceforth we omit complex conjugate equalities and statements since they can be obtained trivially.

Suppose formally $\alpha(\mathbf{x}) = \alpha(t,\mathbf{x})$ in the operators $\hat{A}_k$ in order to define an external bilinear α-current, $Q^{\text{bil}}_{(\alpha)}$, associated with the field α:

$$Q^{\text{bil}}_{(\alpha)}(t,\mathbf{x}) \equiv \underset{(4)}{\frac{\delta}{\delta \alpha(t,\mathbf{x})}} \bar{J}_{(n)}[\{\Psi^*\},\{\Psi\},\{\alpha\}]. \tag{2.6}$$

For instance, if the functions α are deciphered as the components of the metric tensor $g^{\mu\nu}(t,\mathbf{x})$, then (2.6) is proportional to the stress-energy tensor $T_{\mu\nu}$ (see below).

Now we exhibit the known examples of the operator (2.3) to be studied further. The Schrödinger theory is described by the operator

$$\hat{A}_{(1)} = \hat{A}_1\left(i\frac{\partial}{\partial t}\right) + \hat{A}_0, \tag{2.7a}$$

$$\hat{A}_0 = -\hat{H}, \quad \hat{A}_1 = 1, \tag{2.7b}$$

where $\hat{H}$ is a nonrelativistic Hamiltonian.

In a general Klein–Gordon theory, we must write

$$\hat{A}_{(2)} = \hat{A}_2\left(i\frac{\partial}{\partial t}\right)^2 + \hat{A}_1\left(i\frac{\partial}{\partial t}\right) + \hat{A}_0. \tag{2.8}$$

To specify $\hat{A}_k$ $(k = 0,1,2)$ consider the Klein–Gordon equation that ensues from the variational principle (2.5) applied to the action (1.2):

$$(-\det g_{\mu\nu})^{1/2}(\Box - \mu^2)\Psi = 0, \tag{2.9}$$

where $\Box$ is the D'Alambertian; hence, for a relativistic particle in the flat Minkowski space-time,

$$\hat{A}_0^f = \Delta - \mu^2, \quad \hat{A}_1^f = 0, \quad \hat{A}_2^f = 1, \tag{2.8a}$$

where Δ is the three-Laplacian. As for a free particle in a stationary $(3+1)$-splitted Riemann space-time, we introduce, in accordance with Ref. 11, the lapse function N, the shift functions N_a, and the metric tensor H_{ab} of the Riemann three-space V_3 $(\Omega \subset V_3)$ which is employed for raising and lowering Latin indices and for performing the three-covariant differentiation ∇_a. Then, taking into account that $(-\det g_{\mu\nu})^{1/2} = N(\det H_{ab})^{1/2}$ and omitting the factor $(\det H_{ab})^{1/2}$, we get from (2.9):

$$\hat{A}_0^g = \nabla_a\left(NH^{ab} - \frac{N^a N^b}{N}\right)\nabla_b - \mu^2 N,$$

$$\hat{A}_1^g = -i\left(\frac{N^a}{N}\nabla_a + \nabla_a\frac{N^a}{N}\right), \quad \hat{A}_2^g = \frac{1}{N}. \tag{2.8b}$$

Here, we derive functionals conserved in time. Let Ψ and Φ^* satisfy Eq. (2.2) and its conjugate (2.2'), respectively. Contracting (multiplying and integrating over Ω) the former equation for Ψ with $(-i\partial/\partial t)^m \Phi^*$ and the latter one for Φ^* with $(i\partial/\partial t)^m \Psi$ gives

$$\left\langle \left(-i\frac{\partial}{\partial t}\right)^m \Phi^*, \hat{A}_{(n)}\left(i\frac{\partial}{\partial t}\right)\Psi\right\rangle = 0, \tag{2.10a}$$

$$\left\langle \hat{A}^*_{(n)}\left(i\frac{\partial}{\partial t}\right)\Phi^*, \left(i\frac{\partial}{\partial t}\right)^m \Psi\right\rangle = 0. \tag{2.10b}$$

Subtracting (2.10b) from (2.10a), using the time independence and Hermiticity of $\hat{A}_k$ and applying the formula

$$\left\langle \Phi^*, \left(i\frac{\partial}{\partial t}\right)^p \Psi\right\rangle - \left\langle \left(-i\frac{\partial}{\partial t}\right)^p \Phi^*, \Psi\right\rangle = i\frac{\partial}{\partial t}\left[\sum_{q=0}^{p-1}\left\langle \left(-i\frac{\partial}{\partial t}\right)^q \Phi^*, \left(i\frac{\partial}{\partial t}\right)^{p-q-1}\Psi\right\rangle\right],$$

which can be proved by the induction method, one may check the validity of the integral conservation laws

$$i\frac{\partial}{\partial t} C^m_{(n)}\,[\{\Phi^*\},\{\Psi\}] = 0, \tag{2.11}$$

for the functionals

$$C^m_{(n)} = \sum_{k=0}^{n} D^m_k[\{\Phi^*\},\{\Psi\}], \tag{2.12}$$

where

$$D^m_k = \sum_{l=0}^{k-m-1}\left\langle \left(-i\frac{\partial}{\partial t}\right)^{l+m}\Phi^*, \hat{A}_k\left(i\frac{\partial}{\partial t}\right)^{k-l-1}\Psi\right\rangle, \quad m \leqslant k-1;$$

$$D^m_k = 0, \quad m = k;$$

$$D^m_k = -\sum_{l=0}^{m-k-1}\left\langle \left(-i\frac{\partial}{\partial t}\right)^{l+k}\Phi^*, \hat{A}_k\left(i\frac{\partial}{\partial t}\right)^{m-l-1}\Psi\right\rangle, \quad m \geqslant k+1.$$

There are $n+1$ functionals (2.12) for $m = 0,1,\ldots, n$ involving time derivatives of Φ^* and Ψ of orders not higher than $n-1$. [For operators (2.3) with an infinite Taylor expansion there exists an infinite number of such functionals.] The functionals (2.12) are also suitable for $m \geqslant n+1$, but they contain time derivatives of orders equal to or greater than n. Generally speaking, the functionals (2.12) are independent in the sense that any functional $C^p_{(n)}$ cannot be obtained by any algebraic and differential operations from the functionals with $m < p$ without reference to Eqs. (2.2) and (2.2'), although it is not the case for such special solutions as stationary states (see Secs. IV and V).

The functionals (2.12) have the property

$$C^m_{(n)}\,[\{\Phi^*\},\{\Psi\}] = (C^m_{(n)}\,[\{\Psi^*\},\{\Phi\}])^*, \tag{2.13}$$

consequently, they are real at equal arguments, $\operatorname{Im} C^m_{(n)}\,[\{\Psi^*\},\{\Psi\}] = 0$.

In the Schrödinger theory for $m = 0$ and 1,

$$C^0_{(1)} = \langle \Phi^*, \Psi\rangle$$

is the conserved standard inner product coinciding with that in (2.1) and

$$C^1_{(1)} = \langle \Phi^*, \hat{H}\Psi\rangle$$

is the conserved transition probability for the Hamiltonian.

In the Klein–Gordon theory for $m = 0,1,2$, one has

$$C^0_{(2)} = \left\langle \left(-i\frac{\partial}{\partial t}\right)\Phi^*, \hat{A}_2\Psi\right\rangle + \left\langle \Phi^*, \hat{A}_2\left(i\frac{\partial}{\partial t}\right)\Psi\right\rangle + \langle \Phi^*, \hat{A}_1\Psi\rangle,$$

$$C^1_{(2)} = \left\langle \left(-i\frac{\partial}{\partial t}\right)\Phi^*, \hat{A}_2\left(i\frac{\partial}{\partial t}\right)\Psi\right\rangle - \langle \Phi^*, \hat{A}_0\Psi\rangle,$$

$$C^2_{(2)} = -\left\langle \left(-i\frac{\partial}{\partial t}\right)\Phi^*, \hat{A}_1\left(i\frac{\partial}{\partial t}\right)\Psi\right\rangle - \left\langle \left(-i\frac{\partial}{\partial t}\right)\Phi^*, \hat{A}_0\Psi\right\rangle - \left\langle \Phi^*, \hat{A}_0\left(i\frac{\partial}{\partial t}\right)\Psi\right\rangle.$$

By analogy with the Schrödinger theory, define the conserved inner product as follows:[3,4]

$$C^0_{(n)} = \sum_{k=1}^{n}\sum_{l=0}^{k-1}\left\langle \left(-i\frac{\partial}{\partial t}\right)^l \Phi^*, \hat{A}_k\left(i\frac{\partial}{\partial t}\right)^{k-l-1}\Psi\right\rangle, \tag{2.14}$$

the conserved norm square being

$$\mathcal{N}^2 = C^0_{(n)}\,[\{\Psi^*\},\{\Psi\}]. \tag{2.15}$$

Taking (2.14), one could define the modified Hilbert space with an indefinite metric and there would exist a continuum of identical modified Hilbert spaces referred to every moment of time.

Note that the "energy" inner product in Ref. 5 corresponds to the functional $C^1_{(2)}$, in the case considered there, $\hat{A}_2 = 1$ and the question of the positiveness of $C^1_{(2)}$ reduces to that of nonpositive definiteness of $\hat{A}_0$. In the Schrödinger theory, $C^1_{(1)}$ has the meaning of an energy functional as well, suggesting that $C^1_{(n)}$ is always the energy functional for every n.

Now consider a Hermitian operator $\hat{f}$ acting in $\mathcal{L}_2(\Omega)$ which may depend on time, $\partial\hat{f}/\partial t \neq 0$, i.e., the coefficients by three-spatial derivatives in $\hat{f}$ may depend on time. If one tries to modify the definition of the expectation value of $\hat{f}$ in the manner of Ref. 3,

$$(\mathcal{N}[\{\Psi^*\},\{\Psi\}])^{-2}\sum_{k=1}^{n}\sum_{l=0}^{k-1}\left\langle \left(-i\frac{\partial}{\partial t}\right)^l \Psi^*, \hat{A}_k\hat{f}\left(i\frac{\partial}{\partial t}\right)^{k-l-1}\Psi\right\rangle, \tag{2.16}$$

then this requires the modification of Hermitian conjugation for noncommuting $\hat{f}$ and $\hat{A}_k$. A more relevant definition seems to be the following (denoted by a double bar over the symbol):

$$\overline{\overline{\hat{f}}} = (\mathcal{N}[\{\Psi^*\}\{\Psi\}])^{-2}\frac{1}{2}\sum_{k=1}^{n}\sum_{l=0}^{k-1}\left\langle \left(-i\frac{\partial}{\partial t}\right)^l \Psi^*, \{\hat{A}_k\hat{f}\}\left(i\frac{\partial}{\partial t}\right)^{k-l-1}\Psi\right\rangle, \tag{2.17}$$

where, as usual,

$$\{\hat{A}_k, \hat{f}\} = \hat{A}_k\hat{f} + \hat{f}\hat{A}_k.$$

In employing (2.17), no modification of Hermitian conjugation is required since the anticommutator of any Hermitian operators (including noncommuting ones) is always Hermitian:

$\{\hat{A}_k, \hat{f}\} = (\{\hat{A}_k, \hat{f}\})^\dagger$;

so $\left(\overline{\overline{\hat{f}}}\right) = \left(\overline{\overline{\hat{f}}}\right)^*$. The definitions (2.16) and (2.17) are equivalent for commuting $\hat{f}$ and $\hat{A}_k$; in the Schrödinger theory, both of them are reduced to the standard expressions.

Give the necessary and sufficient conditions for the expectation value (2.17) to be conserved in time. Differentiating (2.17) with respect to time and using Eqs. (2.2) and (2.2') yields

$$[\hat{A}_0, \hat{f}] + \frac{1}{2}\left\{\hat{A}_1, i\frac{\partial}{\partial t}\hat{f}\right\} = 0, \tag{2.18.1}$$

$$[\hat{A}_1, \hat{f}] + \left\{\hat{A}_2, i\frac{\partial}{\partial t}\hat{f}\right\} = 0, \tag{2.18.2}$$

$$[\hat{A}_{p-1}, \hat{f}] = 0, \quad \left\{\hat{A}_p, i\frac{\partial}{\partial t}\hat{f}\right\} = 0, \quad p = 2,\ldots,n. \tag{2.18.p}$$

In the Schrödinger theory, the total time derivative of an operator is determined from the condition that its expectation value is equal to the total time derivative of the operator expectation value. Thus the condition (2.18.1) for the Schrödinger operators (2.7a) says that the total time derivative of $\hat{f}$ vanishes:

$$\frac{d}{dt}\hat{f} = -i(2.18.1) = i[\hat{H}, \hat{f}] + \frac{\partial}{\partial t}\hat{f} = 0, \tag{2.19}$$

the conditions (2.18.2) and the subsequent ones becoming trivial. In fact, (2.19) is just the definition of the Schrödinger total time derivative. To extend this definition to the theories with higher time derivatives using the modified expectation value (2.17) one should solve Eq. (2.2) about $(i\partial/\partial t)\Psi$ that means to deal with operators inverse to $\hat{A}_k$.

For the operators $\hat{f}$ independent on time, the set of Eqs. (2.18) degenerates into the conditions for $\hat{f}$ and each $\hat{A}_k$ to commute.

Although the functional (2.15) may take both positive and negative values (zero values should be excluded) and the norm may become imaginary, it may occur that the expectation value (2.17) which is always real has the positive sign irrespective of the sign of $\mathcal{N}^2$. By the way, the unit operator describing the probability density always has the expectation value equal to unity.

III. THE n-COMPONENT FORMALISM

It is possible to describe a quantum system whose evolution is governed by Eq. (2.2) in the framework of the formalism that we call n-component, in contrast to that presented in Sec. II called further monocomponent. The monocomponent formalism is more universal in the sense that it can be applied to nonpolynomial operators as well. However, the n-component formalism has an attractive feature, namely, the possibility to introduce the concept of a Hamiltonian for a quantum system.

For the sake of clarity, start with the general Klein–Gordon theory with Eq. (2.8) and construct the general two-component formalism. Redenoting

$$\Psi = \Psi_0, \tag{3.1}$$

and defining

$$i\frac{\partial}{\partial t}\Psi_0 = \Psi_1, \tag{3.2}$$

we rewrite Eq. (2.8) in the form

$$\hat{A}_2 i\frac{\partial}{\partial t}\Psi_1 = -\hat{A}_1\Psi_1 - \hat{A}_0\Psi_0. \tag{3.3}$$

Thus (3.2) and (3.3) form a set of simultaneous equations to find Ψ_0 and Ψ_1. However, instead of this set it is more convenient to take some operational combinations of Eqs. (3.2) and (3.3). First, applying the operator $\hat{A}_1$ to (3.2) and adding the result to (3.3), we obtain

$$i\frac{\partial}{\partial t}(\hat{A}_1\Psi_0 + \hat{A}_2\Psi_1) = -\hat{A}_0\Psi_0. \tag{3.4}$$

After that apply $\hat{A}_2$ to (3.2):

$$i\frac{\partial}{\partial t}(\hat{A}_2\Psi_0) = \hat{A}_2\Psi_1. \tag{3.5}$$

The set of Eqs. (3.4) and (3.5) may be reformulated in the matrix form

$$i\frac{\partial}{\partial t}(\hat{B}_{IK}\Psi_K) = \hat{H}_{IK}\Psi_K, \tag{3.6}$$

where $I,K = 0,1$, the summation convention over repeated indices is used;

$$\Psi_K = \begin{pmatrix}\Psi_0\\ \Psi_1\end{pmatrix}, \quad \hat{B}_{IK} = \begin{pmatrix}\hat{A}_1 & \hat{A}_2\\ \hat{A}_2 & 0\end{pmatrix}, \quad \hat{H}_{IK} = \begin{pmatrix}-\hat{A}_0 & 0\\ 0 & \hat{A}_2\end{pmatrix},$$

Ψ_K is a two-component column wave function belonging to the direct product of the Hilbert spaces $\mathscr{L}_2'(\Omega) \otimes \mathscr{L}_2'(\Omega) = (\mathscr{L}_2'(\Omega))^2$ at each moment of time, whose conjugate element is the two-component row function

$$\Psi_I^* = (\Psi_0^*, \Psi_1^*).$$

The elements of 2×2 matrices $\hat{B}_{IK}$ and $\hat{H}_{IK}$ are Hermitian operators:

$$\hat{B}^\dagger_{IK} = \hat{B}_{IK}, \quad \hat{H}^\dagger_{IK} = \hat{H}_{IK}, \tag{3.7a}$$

and the matrices are symmetric themselves:

$$\hat{B}_{IK} = \hat{B}_{KI}, \quad \hat{H}_{IK} = \hat{H}_{KI}, \tag{3.7b}$$

hence, they are self-conjugate in $(\mathscr{L}_2(\Omega))^2$, i.e.,

$$\hat{B}^\dagger_{IK} = \hat{B}_{KI}, \quad \hat{H}^\dagger_{IK} = \hat{H}_{KI}. \tag{3.7c}$$

The matrix $\hat{H}_{IK}$ has the meaning of a generalized matrix Hamiltonian.

Conserved functionals can be derived by using the property (3.7c) only. Indeed, consider some two-component row function Φ_I^*, satisfying, by analogy with Eq. (2.2'), the equation complex conjugate to (3.6), with I and K replaced,

$$-i\frac{\partial}{\partial t}(\hat{B}^*_{KI}\Phi_I^*) = \hat{H}^*_{KI}\Phi_I^*. \tag{3.6'}$$

As before, contracting, first, Eq. (3.6) with Φ_I^* and Eq. (3.6') with Ψ_K [and, second, with $(-i\partial/\partial t)\Phi_I^*$ and $(i\partial/\partial t)\Psi_K$, respectively], subtracting one result from the other and recalling (3.7c), we find the two conservation laws

$$i\frac{\partial}{\partial t}C^{0,1} = 0, \tag{3.8a}$$

with

$$C^0 = \langle \Phi_I^*, \hat{B}_{IK} \Psi_K \rangle, \tag{3.8b}$$

$$C^1 = \langle \Phi_I^*, \hat{H}_{IK} \Psi_K \rangle, \tag{3.8c}$$

where, in components,

$$C^0 = \langle \Phi_0^*, \hat{A}_1 \Psi_0 \rangle + \langle \Phi_0^*, \hat{A}_2 \Psi_1 \rangle + \langle \Phi_1^*, \hat{A}_2 \Psi_0 \rangle,$$

$$C^1 = -\langle \Phi_0^*, \hat{A}_0 \Psi_0 \rangle + \langle \Phi_1^*, \hat{A}_2 \Psi_1 \rangle.$$

Returning to the identifications (3.1) and (3.2), on the one hand these expressions coincide with the functionals $C^0_{(2)}$ (the inner product) and $C^1_{(2)}$ in the monocomponent formalism for the Klein–Gordon theory; on the other hand, they are some generalized analogs of the functionals $C^0_{(1)}$ and $C^1_{(1)}$ in the Schrödinger theory. Evidently, the meaning of the matrix $\hat{B}_{IK}$ is to define the inner product (3.8b) and the norm square in the two-component formalism.

We could not derive the third conserved functional, a counterpart of $C^2_{(2)}$, by the above way, we could actually do it by using the column

$$\begin{pmatrix} i\dfrac{\partial}{\partial t} & \Psi_1 \\ i\dfrac{\partial}{\partial t} & \Psi_0 \end{pmatrix} = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix} \begin{pmatrix} i\dfrac{\partial}{\partial t} & \Psi_0 \\ i\dfrac{\partial}{\partial t} & \Psi_1 \end{pmatrix},$$

for contacting with (3.6'), etc. However, it is impossible to proceed without referring to (3.1) and (3.2) and, strictly speaking, their differential consequences.

Extend the above construction to the nth order operator (2.3). Recall (3.1) and define

$$i\frac{\partial}{\partial t}\Psi_0 = \Psi_1, \tag{3.9.1}$$

$$i\frac{\partial}{\partial t}\Psi_1 = \Psi_2, \tag{3.9.2}$$

$$\cdots\cdots\cdots\cdots, \qquad \cdots\cdot$$

$$i\frac{\partial}{\partial t}\Psi_{n-2} = \Psi_{n-1}. \tag{3.9.$n-1$}$$

Using Eqs. (3.9), rewrite (2.2):

$$\hat{A}_n i\frac{\partial}{\partial t}\Psi_{n-1} = -\hat{A}_0\Psi_0 - \hat{A}_1\Psi_1 - \cdots - \hat{A}_{n-2}\Psi_{n-2} - \hat{A}_{n-1}\Psi_{n-1}, \tag{3.10}$$

and reorganize the set of Eqs. (3.9) and (3.10) in the following way.

At the first step, apply the operators $\hat{A}_1, \hat{A}_2, \ldots, \hat{A}_{n-1}$ to each of Eqs. (3.9), respectively, and add all the results and Eq. (3.10) together:

$$i\frac{\partial}{\partial t}(\hat{A}_1\Psi_0 + \hat{A}_2\Psi_1 + \cdots + \hat{A}_{n-1}\Psi_{n-2} + \hat{A}_n\Psi_{n-1}) = -\hat{A}_0\Psi_0. \tag{3.11.1}$$

At the second step, apply the operators $\hat{A}_2, \hat{A}_3, \ldots, \hat{A}_n$ to each of Eqs. (3.9), respectively, and add all the results together:

$$i\frac{\partial}{\partial t}(\hat{A}_2\Psi_0 + \hat{A}_3\Psi_1 + \cdots + \hat{A}_n\Psi_{n-2}) = \hat{A}_2\Psi_1 + \hat{A}_3\Psi_2 + \cdots + \hat{A}_n\Psi_{n-1}. \tag{3.11.2}$$

At the third step, apply the operators $\hat{A}_3, \hat{A}_4, \ldots, \hat{A}_n$ to the first $(n-2)$ equations of (3.9) [except the last, (3.9.$n-1$)] and so on. This process is continued until, at the nth step, it remains to apply the operator $\hat{A}_n$ to the first equation, (3.9.1):

$$i\frac{\partial}{\partial t}(\hat{A}_n\Psi_0) = \hat{A}_n\Psi_1. \tag{3.11.n}$$

The reorganized set of Eqs. (3.11) rewritten in the matrix form looks exactly like (3.6) but with $I,K = 0,1,\ldots,n-1$, and

$$\Psi_K = \begin{pmatrix} \Psi_0 \\ \Psi_1 \\ \Psi_2 \\ \vdots \\ \Psi_{n-2} \\ \Psi_{n-1} \end{pmatrix},$$

$$\hat{B}_{IK} = \begin{pmatrix} \hat{A}_1 & \hat{A}_2 & \hat{A}_3 & \cdots & \hat{A}_{n-1} & \hat{A}_n \\ \hat{A}_2 & \hat{A}_3 & \cdots & \hat{A}_{n-1} & \hat{A}_n & 0 \\ \hat{A}_3 & \cdots & \cdots & \hat{A}_n & 0 & 0 \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ \hat{A}_{n-1} & \hat{A}_n & 0 & \cdots & 0 & 0 \\ \hat{A}_n & 0 & \cdots & \cdots & 0 & 0 \end{pmatrix},$$

$$\hat{H}_{IK} = \begin{pmatrix} -\hat{A}_0 & 0 & 0 & \ldots & 0 & 0 \\ 0 & \hat{A}_2 & \hat{A}_3 & \ldots & \hat{A}_{n-1} & \hat{A}_n \\ 0 & \hat{A}_3 & \ldots & \hat{A}_{n-1} & \hat{A}_n & 0 \\ \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots & \cdots\cdots \\ 0 & \hat{A}_{n-1} & \hat{A}_n & 0 & \ldots & 0 \\ 0 & \hat{A}_n & 0 & \ldots & \ldots & 0 \end{pmatrix},$$

where Ψ_K is an n-component column, that is, an element of $(\mathscr{L}_2(\Omega))^n$ at each moment of time, a conjugate n-component row element being

$$\Psi_I^* = (\Psi_0^*, \Psi_1^*, \ldots, \Psi_{n-2}^*, \Psi_{n-1}^*),$$

$\hat{B}_{IK}$ and $\hat{H}_{IK}$ are $n \times n$ operator matrices acting in $(\mathscr{L}_2(\Omega))^n$ whose structures are evident. The Kth lateral diagonal of $\hat{B}_{IK}$ from $\hat{B}_{KO}$ to $\hat{B}_{OK}$ consists of the operators $\hat{A}_k$ with $k = K+1$ for $K \leqslant n-1$ and the remaining elements are zeros. In the matrix $\hat{H}_{IK}$, all the elements of the first row and the first column, except the element $\hat{H}_{00} = -\hat{A}_0$ at their intersection, are zeros; the remaining part of the matrix $\hat{H}_{IK}$ coincides with the matrix $\hat{B}_{IK}$ with the first column and the last row (or the first row and the last column) erased, so all the properties like (3.7) hold.

The generalization of the bilinear action (2.4), the variational principles (2.5) and (2.5') and the bilinear external current (2.6) on the n-component formalism ($n \geqslant 2$) is straightforward. By analogy with the preceding, the conservation laws in the n-component formalism take place in the

same form as in (3.8a), where (3.8b) is the n-component inner product, with the only difference that $I,K=0,1,\ldots,n-1$; the conserved functionals (3.8b) and (3.8c) are equivalent to the first two of those in (2.12) [for the inner product (3.8b) see also (2.14)] after the identifications (3.1) and (3.9). Concerning the remaining $(n+1)-2=n-1$ functionals (2.12), they are obtainable by more complicated ways with the use of (3.1) and (3.9).

A remark should be made that, in the n-component formalism, any functional of the form

$$\sum_{p,q}\left\langle\left(-i\frac{\partial}{\partial t}\right)^{p}\Phi^{*},\hat{F}_{(p)(q)}\left(i\frac{\partial}{\partial t}\right)^{q}\Psi\right\rangle,$$

e.g., mentioned (2.12), with $p,q=0,1,\ldots,n-1$ and $\hat{F}_{(p)(q)}$ a set of some Hermitian operators ($\hat{F}^{\dagger}_{(p)(q)}=\hat{F}_{(p)(q)}$), can be associated with an operator matrix $\hat{F}_{PQ}$; for functionals with the property (2.13) the matrix $\hat{F}_{PQ}$ has the properties like (3.7a) and (3.7b).

It is worth emphasizing that if we considered the set of Eqs. (3.9) and (3.10) instead of the rearranged set (3.11) (cf. Ref. 17), then the corresponding matrices in $(\mathscr{L}_2(\Omega))^n$ possessing (3.7a) would not possess (3.7b), thus the conserved functionals could not be derived by the way it was done.

Let $\hat{F}_{PQ}$ be a self-conjugate, in general, time-dependent operator matrix [i.e., (3.7c) holds only suggesting that the matrix elements may be not Hermitian]. Define its expectation value like (2.17):

$$\hat{F}=(\langle\Psi^{*}_{L},\hat{B}_{LM}\Psi_{M}\rangle)^{-1}\tfrac{1}{2}\langle\Psi^{*}_{I},\{\hat{F}_{IP},\hat{B}_{PK}\}\Psi_{K}\rangle; \tag{3.12}$$

the necessary and sufficient conditions for (3.12) to be conserved in time are

$$\left\{\hat{B}_{IL},i\frac{\partial}{\partial t}\hat{F}_{LK}\right\}=2[\hat{H}_{IL},\hat{F}_{LK}],\quad[\hat{B}_{IL},\hat{F}_{LK}]=0, \tag{3.13}$$

where we imply the matrix anticommutator and commutator as follows:

$$\{\hat{B}_{IL},\hat{F}_{LK}\}=\hat{B}_{IL}\hat{F}_{LK}+\hat{F}_{IL}\hat{B}_{LK},$$
$$[\hat{B}_{IL},\hat{F}_{LK}]=\hat{B}_{IL}\hat{F}_{LK}-\hat{F}_{IL}\hat{B}_{LK}.$$

Note that due to the self-conjugateness of $\hat{F}_{PQ}$ and $\hat{B}_{IK}$ the anticommutator in (3.12) is self-conjugate leading to $\left(\overline{\overline{\hat{F}}}\right)=\left(\overline{\overline{\hat{F}}}\right)^{*}$.

If we compare the two formalisms, it is obvious that the set of n first-order equations (3.11) is equivalent to the nth-order equation (2.2). Moreover, we see the identity of the inner products and the second conserved (energy) functionals in both the formalisms, but as to the expectation values there is a subtlety.

If the operator $\hat{F}_{PQ}$ is such that

$$\hat{F}_{PQ}=I_{PQ}\hat{f},$$

where I_{PQ} is the unit $n\times n$ matrix, then the expectation value (3.12) agrees with (2.17) in the monocomponent formalism. However, in components, the conservation conditions (3.13) include all the equalities (2.18) except (2.18.2) which is replaced by the two conditions

$$[\hat{A}_{1},\hat{f}]=0,\quad\left[\hat{A}_{2},i\frac{\partial}{\partial t}\hat{f}\right]=0,$$

imposing the stronger restrictions upon $\hat{f}$. The cause is that the conditions (3.13) have been derived under the assumption that all Ψ_K (in particular, Ψ_0 and Ψ_1) are treated as independent variables. If the identifications (3.1) and (3.9) were made before deriving the conservation, then we get just (2.18.2).

IV. STATIONARY STATES: THE RAYLEIGH VARIATIONAL PRINCIPLE

Given a stationary wave function

$$\Psi_{s}(t,\mathbf{x})=\exp(-i\lambda t)\psi(\mathbf{x}) \tag{4.1}$$

($\psi(\mathbf{x})\in\mathscr{L}^{r}_{2}(\Omega)$), the complex conjugate being

$$\Psi^{*}_{s}(t,\mathbf{x})=\exp(+i\lambda^{*}t)\psi^{*}(\mathbf{x}), \tag{4.1'}$$

the problem of finding the general solution to (2.2) is reduced to an eigenvalue problem with nonlinear dependence on the eigenvalues:

$$\hat{A}_{(n)}(\lambda)\psi(\mathbf{x})=0, \tag{4.2}$$

where

$$\hat{A}_{(n)}(\lambda)=\sum_{k=0}^{n}\lambda^{k}\hat{A}_{k}. \tag{4.3}$$

After the substitution of (4.1') into (2.2') the conjugate problem is

$$\hat{A}^{*}_{(n)}(\lambda^{*})\psi^{*}(\mathbf{x})=0, \tag{4.2'}$$

with the conjugate operator

$$\hat{A}^{*}_{(n)}(\lambda^{*})=\sum_{k=0}^{n}\lambda^{*k}\hat{A}^{*}_{k}. \tag{4.3'}$$

The operator (4.3) is Hermitian, i.e., $\hat{A}^{T}_{(n)}=\hat{A}^{*}_{(n)}$, if and only if λ is a real number. We concentrate our attention here and below to real λ.

For real λ, the functionals (2.12) with equal arguments automatically lose time dependence:

$$C^{m}_{(n)}[\{\Psi^{*}_{s}\},\{\Psi_{s}\}]$$
$$=C^{m}_{(n)}[\psi^{*},\psi]$$
$$=\lambda^{m}\langle\psi^{*},\hat{A}'_{(n)}(\lambda)\psi\rangle-m\lambda^{m-1}\langle\psi^{*},\hat{A}_{(n)}(\lambda)\psi\rangle, \tag{4.4}$$

and $\hat{A}'_{(n)}(\lambda)\equiv d\hat{A}_{(n)}(\lambda)/d\lambda$ is always Hermitian. Evidently, among the functionals (4.4) only any two are independent without assuming that Eq. (4.2) is fulfilled. This is connected with the facts that the time derivatives of Ψ_s are no longer independent and are proportional to the functions Ψ_s and that only the two equalities (2.10) were used. After fulfilling Eq. (4.2), $C^{m}_{(n)}=\lambda^{m}C^{0}_{(n)}$ so there is one independent functional, e.g., $C^{0}_{(n)}$ defining as before the norm square:

$$\mathscr{N}^{2}=C^{0}_{(n)}[\psi^{*},\psi]$$
$$=\langle\psi^{*},\hat{A}'_{(n)}(\lambda)\psi\rangle$$
$$=\sum_{k=1}^{n}k\lambda^{k-1}\langle\psi^{*},\hat{A}_{k}\psi\rangle. \tag{4.5}$$

Assume that the operator (4.3) is compact in $\mathscr{L}_2(\Omega)$ in order for its set of eigenvalues to be discrete.[6] To find this discrete set of eigenvalues, a known way is to determine all the extrema of the Rayleigh functionals.[15]

Here, we introduce the functionals to be used further:

$$J_{(n)} = T^{-1}\bar{J}_{(n)}, \tag{4.6}$$

where $\bar{J}_{(n)}$ is the bilinear action (2.4) and

$$J'_{(n)} = T^{-1}\int_0^T dt\, C^0_{(n)}$$
$$= T^{-1}\int_0^T dt \sum_{k=1}^{n}\sum_{l=0}^{k-1}\left\langle\left(-i\frac{\partial}{\partial t}\right)^l \Psi^*, \hat{A}_k\left(i\frac{\partial}{\partial t}\right)^{k-l-1}\Psi\right\rangle. \tag{4.6'}$$

It is easily verified that

$$J_{(n)}[\{\Psi_s^*\},\{\Psi_s\}] = J_{(n)}[\psi^*,\psi;\lambda] = \langle\psi^*,\hat{A}_{(n)}(\lambda)\psi\rangle, \tag{4.7}$$

$$J'_{(n)}[\{\Psi_s^*\},\{\Psi_s\}] = J'_{(n)}[\psi^*,\psi;\lambda] = \langle\psi^*,\hat{A}'_{(n)}(\lambda)\psi\rangle. \tag{4.7'}$$

Compared with (4.7), the prime in (4.7') has the functional meaning: $J'_{(n)}(\lambda) = dJ_{(n)}(\lambda)/d\lambda$ and $J'_{(n)}$ is merely another notation for the norm square (4.5).

Let the following two conditions be satisfied. (a) The algebraic (in general, transcendental for $n=\infty$) equation with respect to S,

$$J_{(n)}(S) = \langle\psi^*,\hat{A}_{(n)}(S)\psi\rangle = \sum_{k=0}^{n} S^k\langle\psi^*,\hat{A}_k\psi\rangle = 0, \tag{4.8}$$

has no zero roots, and the functional $S^{\cdot\mathscr{M}}$ known as Rayleigh's

$$S^{\cdot\mathscr{M}} = S^{\cdot\mathscr{M}}[\psi^*,\psi] \neq 0$$

is an $\mathscr{M}$th (nonzero) root of (4.8), where $\mathscr{M} = 1,2,\ldots,p\leqslant n$ enumerates the roots in the increasing order.

(b) For all the roots of (4.8), the functional (4.7') is nonzero:

$$J'_{(n)}(S^{\cdot\mathscr{M}}) = \langle\psi^*,\hat{A}'_{(n)}(S^{\cdot\mathscr{M}})\psi\rangle = \sum_{k=1}^{n} k(S^{\cdot\mathscr{M}})^{k-1}\langle\psi^*,\hat{A}_k\psi\rangle \neq 0, \tag{4.8'}$$

unless $\psi\neq 0$, thus implying that the roots cannot be multiple. (We consider only real roots, $S^{\cdot\mathscr{M}} = S^{\cdot\mathscr{M}*}$, although it is not necessary for the statements below.)

Then: (i) An extremum condition of $S^{\cdot\mathscr{M}}$,

$$_{(3)}\frac{\delta S^{\cdot\mathscr{M}}}{\delta\psi^*} = 0, \tag{4.9}$$

is equivalent to the requirement that Eq. (4.2) is valid on this functional, i.e.,

$$\hat{A}_{(n)}(S^{\cdot\mathscr{M}})\psi(\mathbf{x}) = 0,$$

the $\mathscr{J}_{\cdot\mathscr{M}}$ extremum value of $S^{\cdot\mathscr{M}}$, $S^{\cdot\mathscr{M}}_{\mathscr{J}_{\cdot\mathscr{M}}}$, coincides with the $\mathscr{J}_{\cdot\mathscr{M}}$ eigenvalue $\lambda_{\mathscr{J}_{\cdot\mathscr{M}}}$ (from $S^{\cdot\mathscr{M}}\neq 0$ it follows $\lambda_{\mathscr{J}_{\cdot\mathscr{M}}}\neq 0$) belonging to some $\mathscr{M}$ subset of the eigenvalue set,

$$\lambda_{\mathscr{J}_{\cdot\mathscr{M}}} = S^{\cdot\mathscr{M}}_{\mathscr{J}_{\cdot\mathscr{M}}} = S^{\cdot\mathscr{M}}[\psi^*_{\mathscr{J}_{\cdot\mathscr{M}}},\psi_{\mathscr{J}_{\cdot\mathscr{M}}}], \tag{4.10}$$

and the argument $\psi_{\mathscr{J}_{\cdot\mathscr{M}}}$ supplying $S^{\cdot\mathscr{M}}$ with the $\mathscr{J}_{\cdot\mathscr{M}}$ extremum value coincides with the $\mathscr{J}_{\cdot\mathscr{M}}$ eigenfunction, the index $\mathscr{J}_{\cdot\mathscr{M}}$ enumerating simultaneously the extrema of $S^{\cdot\mathscr{M}}$ and the eigenvalues in the order of increasing as well as the arguments of $S^{\cdot\mathscr{M}}$ at the extremum points and the eigenfunctions, $\mathscr{J}_{\cdot\mathscr{M}} = 0_{\cdot\mathscr{M}}, 1_{\cdot\mathscr{M}},\ldots$. (The complex conjugate statements are also valid.)

Indeed, the substitution of $S^{\cdot\mathscr{M}}$ back into (4.8) gives the identity

$$J_{(n)}(S^{\cdot\mathscr{M}}[\psi^*,\psi]) = \langle\psi^*,\hat{A}_{(n)}(S^{\cdot\mathscr{M}}[\psi^*,\psi])\psi\rangle \equiv 0.$$

Differentiating it implicitly with respect to ψ^* and taking into account (4.8'), one gets

$$_{(3)}\frac{\delta S^{\cdot\mathscr{M}}}{\delta\psi^*} = -(J'_{(n)}(S^{\cdot\mathscr{M}}))^{-1}\hat{A}_{(n)}(S^{\cdot\mathscr{M}})\psi = 0, \tag{4.11}$$

and the above statements become evident.

In our opinion, because the extrema of the functionals $S^{\cdot\mathscr{M}}$ gives the equations, although stationary and restricted by the eigenvalue problem, it is fair to call the $S^{\cdot\mathscr{M}}$ quantum Rayleigh eigenactions.

It is important to note that, for the eigenfunctions $\psi_{\mathscr{J}_{\cdot\mathscr{M}}}$, the condition (4.8'), $J'_{(n)}[\psi^*_{\mathscr{J}_{\cdot\mathscr{M}}},\psi_{\mathscr{J}_{\cdot\mathscr{M}}};\lambda_{\mathscr{J}_{\cdot\mathscr{M}}}]\neq 0$, means that their norm is nonzero and that, in turn, is equivalent to the absence of associated vectors in $\mathscr{L}_2(\Omega)$.[4,6]

Recalling the dependence of $\hat{A}_k$ on $\alpha(\mathbf{x})$, define the Rayleigh external α current, $Q^{\cdot\mathscr{M}}_{(\alpha)}(\mathbf{x})$, as follows

$$Q^{\cdot\mathscr{M}}_{(\alpha)} = -_{(3)}\frac{\delta S^{\cdot\mathscr{M}}}{\delta\alpha}. \tag{4.12a}$$

Then, (ii)[1] (4.12a) is equal to the bilinear α current (2.6) divided by the norm square on the function $\Psi^{\mathscr{M}}_s = \exp(-iS^{\cdot\mathscr{M}}t)\psi(x)$,

$$Q^{\cdot\mathscr{M}}_{(\alpha)} = (\mathscr{N}[\{\psi_s^{\mathscr{M}*}\},\{\psi_s^{\mathscr{M}}\}])^{-2}Q^{\mathrm{bil}}_{(\alpha)}(\{\Psi_s^{\mathscr{M}*}\},\{\Psi_s^{\mathscr{M}}\}). \tag{4.12b}$$

In the extremum points with respect to the variables ψ^* and ψ, not α, (4.12b) gives the stationary current $Q^{\mathrm{bil}}_{(\alpha)}$ on the solutions of the eigenvalue problem divided by the norm square of corresponding eigenfunctions.

In fact, by analogy with (4.11),

$$_{(3)}\frac{\delta S^{\cdot\mathscr{M}}}{\delta\alpha} = -(J'_{(n)}(S^{\cdot\mathscr{M}}))^{-1} \times_{(3)}\frac{\delta}{\delta\alpha}\langle\psi^*,\hat{A}_{(n)}(S^{\cdot\mathscr{M}};\alpha)\psi\rangle. \tag{4.12c}$$

Using the equality

$$_{(4)}\frac{\delta}{\delta\alpha(t,\mathbf{x})}\bar{J}_{(n)}[\{\Psi_s^{\mathscr{M}*}\},\{\Psi_s^{\mathscr{M}}\}]|_{\alpha(t,\mathbf{x})=\alpha(\mathbf{x})} = _{(3)}\frac{\delta}{\delta\alpha(\mathbf{x})}J_{(n)}(S^{\cdot\mathscr{M}})$$

one can see the validity of (4.12b).

If we establish

$$\alpha = (N,N_a,H_{ab}), \tag{4.13}$$

then (4.12b) gives some components of the stationary stress-

energy tensor divided by the norm square (see Sec. VII).

The Schrödinger theory with the eigenvalue problem

$$\hat{A}_{(1)}(\lambda)\psi = (\lambda - \hat{H})\psi = 0 \tag{4.14}$$

[cf. (2.7)] is a trivial particular case. An equation of the type (4.8) is linear in S^{Sch}:

$$S^{\mathrm{Sch}}\langle\psi^*,\psi\rangle - \langle\psi^*,\hat{H}\psi\rangle = 0,$$

and has the unique simple root that is real due to the Hermiticity of $\hat{H}$:

$$S^{\mathrm{Sch}}[\psi^*,\psi] = (\langle\psi^*,\psi\rangle)^{-1}\langle\psi^*,\hat{H}\psi\rangle,$$

and whose range covers the whole set of the eigenvalues; the variational derivative of S^{Sch} with respect to ψ^* is

$$_{(3)}\frac{\delta S^{\mathrm{Sch}}}{\delta\psi^*} = -(\langle\psi^*,\psi\rangle)^{-1}(S^{\mathrm{Sch}}[\psi^*,\psi]\psi - \hat{H}\psi).$$

For the general Klein–Gordon theory with the operator (2.8), the operator (4.3) acquires the form

$$\hat{A}_{(2)}(\lambda) = \lambda^2\hat{A}_2 + \lambda\hat{A}_1 + \hat{A}_0 \tag{4.15}$$

and Eq. (4.8) is now

$$S^2\langle\psi^*,\hat{A}_2\psi\rangle + S\langle\psi^*,\hat{A}_1\psi\rangle + \langle\psi^*,\hat{A}_0\psi\rangle = 0. \tag{4.16}$$

Assume that

$$\langle\psi^*,\hat{A}_2\psi\rangle \neq 0 \tag{4.17}$$

for each $\psi(\mathbf{x})\in\mathscr{L}'_2(\Omega)$ and that the discriminant D of Eq. (4.16) is non-negative;

$$D[\psi^*,\psi] = (\langle\psi^*,\hat{A}_1\psi\rangle)^2 - 4\langle\psi^*,\hat{A}_0\psi\rangle\langle\psi^*,\hat{A}_2\psi\rangle \geqslant 0. \tag{4.18}$$

The condition (4.18) resembles the overdamping condition for heavily damped dynamical systems where, in addition, conditions like $\langle\psi^*,\hat{A}_k\psi\rangle \geqslant 0$ for $k = 0,1,2$ (in our terms) hold.[13]

It is clear that

$$\hat{A}'_{(2)}(S) = 2S\hat{A}_2 + \hat{A}_1,$$

$$\langle\psi^*,\hat{A}'_{(2)}(S)\psi\rangle = 2S\langle\psi^*,\hat{A}_2\psi\rangle + \langle\psi^*,\hat{A}_1\psi\rangle.$$

Owing to the conditions (4.17) and (4.18), Eq. (4.16) has two real roots ($\mathscr{M} = 1,2$):

$$S^{1,2} = (2\langle\psi^*,\hat{A}_2\psi\rangle)^{-1}(-\langle\psi^*,\hat{A}_1\psi\rangle \mp D^{1/2}). \tag{4.19}$$

Substituting the roots (4.19) into the condition (4.8′), we bring the latter to the form:

$$J'_{(2)} = \langle\psi^*,\hat{A}'_{(2)}(S^{1,2}[\psi^*,\psi])\psi\rangle = \mp(D[\psi^*,\psi])^{1/2} \neq 0, \tag{4.20}$$

and it leads to $S^1 \neq S^2$, that is, the ranges of S^1 and S^2 are separated; this can be readily seen from the inequalities

$$S^1 < -(2\langle\psi^*,\hat{A}_2\psi\rangle)^{-1}\langle\psi^*,\hat{A}_1\psi\rangle < S^2. \tag{4.21}$$

The formula (4.11) may be confirmed by a direct calculation.

In the example (2.8b), the operator $\hat{A}^g_0$ is always nonpositive definite (see the explanation in Ref. 1) The number $\langle\psi^*,\hat{A}^g_1\psi\rangle$ is real due to the Hermiticity of $\hat{A}^g_1$, hence its square is non-negative. The operator $\hat{A}^g_2$ is positive definite because $N > 0$. Thus

$$D^g = (\langle\psi^*,\hat{A}^g_1\psi\rangle)^2 - 4\langle\psi^*,\hat{A}^g_0\psi\rangle\langle\psi^*,\hat{A}^g_2\psi\rangle \geqslant 0$$

is always true as a consequence of the physical signature of $g_{\mu\nu}$. By virtue of $\langle\psi^*,\hat{A}^g_0\psi\rangle\cdot\langle\psi^*,\hat{A}^g_2\psi\rangle \leqslant 0$ and (4.20), the roots S^1 and S^2 have different signs and the inequalities (4.21) can be made more precise: if $\langle\psi^*,\hat{A}^g_1\psi\rangle > 0$, then

$$S^1 < -(\langle\psi^*,\hat{A}^g_2\psi\rangle)^{-1}\langle\psi^*,\hat{A}^g_1\psi\rangle, \quad S^2 > 0,$$

and if $\langle\psi^*,\hat{A}^g_1\psi\rangle < 0$, then

$$S^1 < 0, \quad S^2 > (\langle\psi^*,\hat{A}^g_2\psi\rangle)^{-1}\langle\psi^*,\hat{A}^g_1\psi\rangle,$$

similar inequalities standing for the corresponding subsets of the eigenvalue set.

In the example (2.8a) which is a particular case of (2.8b), the nonpositive definiteness of $\hat{A}^f_0$ is clearly demonstrated:

$$\langle\psi^*,\hat{A}^f_0\psi\rangle = \langle\psi^*,(\Delta - \mu^2)\psi\rangle = -\int d\Omega(\nabla_a\psi^*\nabla^a\psi + \mu^2\psi^*\psi) \leqslant 0$$

under the boundary conditions for $\psi(\mathbf{x})\in\mathscr{L}'_2(\Omega)$ providing the Hermiticity of $\hat{A}^f_0$. The roots are

$$S^{1,2} = \mp(\langle\psi^*,\psi\rangle)^{-1/2}(\langle\psi^*,(\mu^2 - \Delta)\psi\rangle)^{1/2}$$

so that $S^1 < 0$ and $S^2 > 0$. Note that the quantity S^2 is the original Rayleigh quotient.[12] In this case, the condition (4.20) is equivalent to the absence of zero roots.

Returning to the modified expectation values, for stationary states, the definition (2.17) transforms to

$$\overline{\overline{f}} = \tfrac{1}{2}\langle\psi^*,\{\hat{A}'_{(n)}(\lambda),\hat{f}\}\psi\rangle / \langle\psi^*,\hat{A}'_{(n)}(\lambda)\psi\rangle, \tag{4.22}$$

where by means of (4.8′) the denominator is not equal to zero. In particular, for the operators (2.8a), $\hat{A}^{f'}_{(2)}(\lambda) = 2\lambda$, therefore, irrespective of the sign of λ, the expectation value (4.22) is identical to that in the Schrödinger theory (compare with the end of Sec. II).

It is also worth remembering that the usual Heisenberg uncertainty relations are derived using the positive definiteness of the square, e.g., $\hat{f}^2$, of a Hermitian operator, $\hat{f}^2 = (\gamma\hat{q} - i\hat{p})^2$, where $\hat{q}$ and $\hat{p}$ are canonically conjugate operators, $[\hat{p},\hat{q}] = -i$, and γ is a parameter. However, even if

$$\langle\psi^*,\hat{A}'_{(n)}(\lambda)\psi\rangle > 0, \quad \langle\psi^*,\hat{f}^2\psi\rangle > 0,$$

one cannot conclude the sign definiteness of the anticommutator in (4.22). This fact entails a serious consequence that the modified uncertainty relation

$$\left\{(\overline{\overline{\hat{q}}})^2 - \left(\overline{\overline{\hat{q}}}\right)^2\right\}^{1/2}\cdot\left\{(\overline{\overline{\hat{p}}})^2 - \left(\overline{\overline{\hat{p}}}\right)^2\right\}^{1/2} \geqslant \frac{1}{2}$$

is feasible not for all the functions ψ. Moreover, this is true for the general case of (2.17).

V. STATIONARY STATES: ORTHOGONALITY AND SUPERPOSITION

For Eq. (4.8) to be of the nth degree, suppose

$$\langle \psi^*, \hat{A}_n \psi \rangle \neq 0, \tag{5.1}$$

for all $\psi(\mathbf{x}) \in \mathscr{L}_2'(\Omega)$ and let Eq. (4.8) have $p \leqslant n$ real roots, $\mathscr{M} = 1,2,...,p$. Let $S^{\prime\mathscr{M}'}$ be the set of values where the functional $J_{(n)}(S)$ [see (4.6)] reaches extrema, i.e., where $J'_{(n)}(S) = 0$, which are enumerated in the order of increasing as well, $\mathscr{M}' = 1',2',...,(p-1)'$. By assumption (4.8'), Eq. (4.8) has no multiple roots, hence the ranges of the functionals $S^{\cdot\mathscr{M}}$ (spectral zones[18]) are separated:

$$S^1 < S^{\prime 1'} < S^2 < S^{\prime 2'} < \cdots < S^{\prime(p-1)'} < S^p.$$

Obviously $\lambda_{f_{2'\mathscr{M}_2}} > \lambda_{f_{1'\mathscr{M}_1}}$ for any $f_{2'\mathscr{M}_2}$ and $f_{1'\mathscr{M}_1}$ if $\mathscr{M}_2 > \mathscr{M}_1$.

If Eq. (4.8) has exactly n real roots, the operator (4.3) is called hyperbolic.[14] [The definition of a hyperbolic operator requires also $\langle \psi^*, \hat{A}_n \psi \rangle > 0$, but due to (5.1) and the boundedness requirement for the operator $\hat{A}_n$ (see Sec. II) implying its continuity this quantity has a constant sign and the negative sign may be excluded by changing the sign of the operator $\hat{A}_{(n)}$].

Generally, the functionals $S^{\cdot\mathscr{M}}$, $\mathscr{M} = 1,2,...,n$ (which are not obliged to be real) obey the generalized Viete theorem:

$$\begin{aligned}
&\sum_{\mathscr{M}=1}^{n} S^{\cdot\mathscr{M}} = -(\langle \psi^*, \hat{A}_n \psi \rangle)^{-1} \langle \psi^*, \hat{A}_{n-1} \psi \rangle \equiv -\alpha_{n-1}, \\
&\sum_{\mathscr{M},\mathscr{N}=1(\mathscr{M}<\mathscr{N})}^{n} S^{\cdot\mathscr{M}} S^{\cdot\mathscr{N}} = (\langle \psi^*, \hat{A}_n \psi \rangle)^{-1} \langle \psi^*, \hat{A}_{n-2} \psi \rangle \\
&\qquad \equiv \alpha_{n-2}, \\
&\sum_{\mathscr{M},\mathscr{N},\mathscr{P}=1(\mathscr{M}<\mathscr{N}<\mathscr{P})}^{n} S^{\cdot\mathscr{M}} S^{\cdot\mathscr{N}} S^{\cdot\mathscr{P}} \\
&\quad = -(\langle \psi^*, \hat{A}_n \psi \rangle)^{-1} \langle \psi^*, \hat{A}_{n-3} \psi \rangle \\
&\quad \equiv -\alpha_{n-3}, \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots, \\
&S^1 S^2 \cdots S^n \\
&\quad = (-1)^n (\langle \psi^*, \hat{A}_n \psi \rangle)^{-1} \langle \psi^*, \hat{A}_0 \psi \rangle \\
&\quad \equiv (-1)^n \alpha_0,
\end{aligned} \tag{5.2}$$

where not all the real functionals α_k are zeros identically, $k = 0,1,...,n-1$.

Each discrete eigenvalue $\lambda_{f_{\cdot\mathscr{M}}}$ is related to one eigenfunction or a finite number l of eigenfunctions, i.e., it may be l fold degenerate. The question arises of whether different eigenvalues values $\lambda_{f_{\cdot\mathscr{M}}}$ for different $\mathscr{M}$ may correspond to one and the same eigenfunction, let it be denoted by ψ_*. Allowing $\lambda_{f_{\cdot\mathscr{M}}}$ to acquire complex values and supposing that there are no multiple roots $S^{\cdot\mathscr{M}}$, the answer is given by the following statement.

The necessary and sufficient condition for the eigenfunction ψ_* to be corresponded to n distinct eigenvalues $\lambda^{\cdot\mathscr{M}}$, $\mathscr{M} = 1,2,...,n$, each of them belonging to one and only one $\mathscr{M}$ subset is the following:

$$\hat{A}_k \psi_* = \beta_k^* \hat{A}_n \psi_*, \tag{5.3}$$

with β_k^* real numbers not all of whose are zeros, $k = 0,1,...,n-1$.

First of all, note that each β_k^* in (5.3) can always be presented in the form

$$\beta_k^* = \langle \varphi^*, \hat{A}_k \psi_* \rangle / \langle \varphi^*, \hat{A}_n \psi_* \rangle, \tag{5.4}$$

$$\beta_k^* = \langle \varphi^*, \hat{A}_k \psi_* \rangle / \langle \varphi^*, \hat{A}_n \psi_* \rangle, \tag{5.4}$$

for every function φ for which the denominator in (5.4) is nonzero. In particular, β_k^* may be presented as the functionals α_k in (5.2) on ψ_*:

$$\beta_k^* = \alpha_k [\psi_*^*, \psi_*] \tag{5.5}$$

under the condition (5.1) for ψ_*.

Differentiate each of Eqs. (5.2) with respect to ψ^*,

$$\begin{aligned}
&\sum_{\mathscr{M}=1}^{n} {}_{(3)}\frac{\delta S^{\cdot\mathscr{M}}}{\delta \psi^*} \\
&\quad = -(\langle \psi^*, \hat{A}_n \psi \rangle)^{-1} (\hat{A}_{n-1} \psi - \alpha_{n-1} \hat{A}_n \psi), \\
&\sum_{\substack{\mathscr{M},\mathscr{N}=1 \\ (\mathscr{M}<\mathscr{N})}}^{n} {}_{(3)}\frac{\delta S^{\cdot\mathscr{M}}}{\delta \psi^*} S^{\cdot\mathscr{N}} + S^{\cdot\mathscr{M}} {}_{(3)}\frac{\delta S^{\cdot\mathscr{N}}}{\delta \psi^*} \\
&\quad = (\langle \psi^*, \hat{A}_n \psi \rangle)^{-1} (\hat{A}_{n-2} \psi - \alpha_{n-2} \hat{A}_n \psi), \\
&\sum_{\substack{\mathscr{M},\mathscr{N},\mathscr{P}=1 \\ (\mathscr{M}<\mathscr{N}<\mathscr{P})}}^{n} {}_{(3)}\frac{\delta S^{\cdot\mathscr{M}}}{\delta \psi^*} S^{\cdot\mathscr{N}} S^{\cdot\mathscr{P}} \\
&\qquad + S^{\cdot\mathscr{M}} {}_{(3)}\frac{\delta S^{\cdot\mathscr{N}}}{\delta \psi^*} S^{\cdot\mathscr{P}} + S^{\cdot\mathscr{M}} S^{\cdot\mathscr{N}} {}_{(3)}\frac{\delta S^{\cdot\mathscr{P}}}{\delta \psi^*} \\
&\qquad = -(\langle \psi^*, \hat{A}_n \psi \rangle)^{-1} (\hat{A}_{n-3} \psi - \alpha_{n-3} \hat{A}_n \psi), \\
&\cdots\cdots\cdots\cdots\cdots\cdots\cdots\cdots, \\
&{}_{(3)}\frac{\delta S^1}{\delta \psi^*} S^2 \cdots S^n \\
&\qquad + S^1 {}_{(3)}\frac{\delta S^2}{\delta \psi^*} \cdots S^n + S^1 S^2 \cdots {}_{(3)}\frac{\delta S^n}{\delta \psi^*} \\
&\qquad = (-1)^n (\langle \psi^*, \hat{A}_n \psi \rangle)^{-1} (\hat{A}_0 \psi - \alpha_0 \hat{A}_n \psi),
\end{aligned} \tag{5.6}$$

then our statement becomes obvious. Indeed, proceeding in terms of quantities at extremum points, if ψ_* supplies an extremum with each $S^{\cdot\mathscr{M}}$, i.e., if

$$ {}_{(3)}\frac{\delta S^{\cdot\mathscr{M}}}{\delta \psi^*}\bigg|_{\psi=\psi_*} = 0, \tag{4.9*}$$

then this yields (5.3) with β_k^* from (5.5). Conversely, owing to the assumption that $S^{\cdot\mathscr{M}} \neq S^{\cdot\mathscr{N}}$ for each $\mathscr{M} \neq \mathscr{N}$ and $S^{\cdot\mathscr{M}} \neq 0$ for each $\mathscr{M}$, the set of Eqs. (5.6) is nondegenerate, thus, when the right-hand sides of (5.6) are zeros by (5.3) and (5.5), it has the only solution (4.9*) for each $\mathscr{M}$.

The Klein–Gordon theory with the operators (2.8a) is an example of such a sort. The eigenvalues $\lambda_{f_2} = -\lambda_{f_1} \equiv -\lambda_f = (\mu^2 + k_f^2)^{1/2}$ correspond to the eigenfunction ψ_f which, in turn, is an eigenfunction of the Laplacian in the domain Ω with the eigenvalue $(-k_f^2)$, $\Delta \psi_f = -k_f^2 \psi_f$, so that $\beta_0^f = -(\mu^2 + k_f^2)^{1/2}$ and $\beta_1^f = 0$.

Now turn to stationary states (4.1), Ψ_{s1} and Ψ_{s2}, with eigenvalues and eigenfunctions corresponding to different $\mathscr{M}$ and $f_{\cdot\mathscr{M}}$. To simplify notation, in this section, we put $\lambda_1 = \lambda_{f_{1'\mathscr{M}_2}}$, $\lambda_2 = \lambda_{f_{2'\mathscr{M}_2}}$ and $\psi_1 = \psi_{f_{1'\mathscr{M}_1}}$, $\psi_2 = \psi_{f_{2'\mathscr{M}_2}}$.

Assume that the functions ψ_1 and ψ_2, instead of Eqs. (4.2) and (4,2'), satisfy some weaker integral equalities:

$$\langle \psi_2^*, \hat{A}_{(n)}(\lambda_1)\psi_1 \rangle = \langle \hat{A}^*_{(n)}(\lambda_1)\psi_2^*, \psi_1 \rangle = 0, \quad (5.7a)$$

$$\langle \hat{A}^*_{(n)}(\lambda_2)\psi_2^*, \psi_1 \rangle = \langle \psi_2^*, \hat{A}_{(n)}(\lambda_2)\psi_1 \rangle = 0. \quad (5.7b)$$

Multiplying (5.7a) by ${\lambda_2}^m$ and (5.7b) by ${\lambda_1}^m$, subtracting the second result from the first one, we actually repeat the derivation giving the conservation laws (2.11) and get the set of equalities equivalent to (2.11):

$$(\lambda_2 - \lambda_1)\mathscr{C}^m_{(n)}[\psi_2^*, \psi_1] = 0, \quad \mathscr{C}^m_{(n)} = \sum_{k=0}^{n} \mathscr{D}^m_k[\psi_2^*, \psi_1],$$

with the correspondence to the functionals (2.12):

$$C^m_{(n)}, D^m_k[\Psi^*_{s2}, \Psi_{s1}] = \exp(i(\lambda_2 - \lambda_1)t)\mathscr{C}^m_{(n)}, \mathscr{D}^m_k,$$

where

$$\mathscr{D}^m_k = \sum_{l=0}^{k-m-1} {\lambda_2}^{l+m}{\lambda_1}^{k-l-1}\langle \psi_2^*, \hat{A}_k\psi_1 \rangle, \quad m \leqslant k-1;$$

$$\mathscr{D}^m_k = 0, \quad m = k;$$

$$\mathscr{D}^m_k = -\sum_{l=0}^{m-k-1} {\lambda_2}^{l+k}{\lambda_1}^{m-l-1}\langle \psi_2^*, \hat{A}_k\psi_1 \rangle,$$

$$m \geqslant k+1.$$

For noncoinciding λ_2 and λ_1, $\lambda_2 \neq \lambda_1$ ($\mathscr{J}_{1''_1} \neq \mathscr{J}_{2''_2}$ even if $\mathscr{M}_2 = \mathscr{M}_1$), we can see the relations

$$\sum_{k=0}^{n} \mathscr{D}^m_k[\psi_2^*, \psi_1] = 0,$$

which make sense of the orthogonality properties. Note that again among them only any two are linearly independent as a consequence of the only two initial equalities (5.7) used. [Of course after Eq. (4.2) is taken into account there is only one independent orthogonality property.] For the two orthogonality properties, let us choose those with $m = 0$ and $m = 1$ and rewrite them in detail:

$$\sum_{k=1}^{n}\sum_{l=0}^{k-1} {\lambda_2}^l{\lambda_1}^{k-l-1}\langle \psi_2^*, \hat{A}_k\psi_1 \rangle = 0, \quad (5.8a)$$

$$\sum_{k=2}^{n}\sum_{l=0}^{k-2} {\lambda_2}^{l+1}{\lambda_1}^{k-l-1}\langle \psi_2^*, \hat{A}_k\psi_1 \rangle - \langle \psi_2^*, \hat{A}_0\psi_1 \rangle = 0. \quad (5.8b)$$

Using the algebraic formula

$${\lambda_2}^p - {\lambda_1}^p = (\lambda_2 - \lambda_1)\sum_{q=0}^{p-1} {\lambda_2}^q{\lambda_1}^{p-q-1}$$

valid for any integer $p \geqslant 1$, we rewrite (5.8a) in the form already given in Ref. 4 in which it holds for nonpolynomial operators as well

$$\left\langle \psi_2^*, \frac{\hat{A}_{(n)}(\lambda_2) - \hat{A}_{(n)}(\lambda_1)}{\lambda_2 - \lambda_1}\psi_1 \right\rangle = 0.$$

The two orthogonality properties (5.8) have an interesting appearance in the case of the quadratic operator (4.15). By (4.17), the relations (5.8) may be presented in the form

$$\lambda_2 + \lambda_1 = -(\langle \psi_2^*, \hat{A}_2\psi_1 \rangle)^{-1}\langle \psi_2^*, \hat{A}_1\psi_1 \rangle,$$

$$\lambda_2\lambda_1 = (\langle \psi_2^*, \hat{A}_2\psi_1 \rangle)^{-1}\langle \psi_2^*, \hat{A}_0\psi_1 \rangle,$$

exciting associations with the Viete theorem for the roots of quadratic equations. However, this analogy does not work for an equation of a degree higher than 2 ($n > 2$), since every quantity $\langle \psi_2^*, \hat{A}_k\psi_1 \rangle$ cannot be expressed via λ_1 and λ_2 by solving only Eqs. (5.7).

In the Schrödinger theory, the orthogonality properties (5.8) are

$$\langle \psi_2^*, \psi_1 \rangle = 0, \quad \langle \psi_2^*, \hat{H}\psi_1 \rangle = 0,$$

thus the first of them is standard, as it should be; the second one is proportional to the first on Eq. (4.14).

Having clarified the modified orthogonality concept, we are able to introduce the orthonormal set of eigenfunctions obeying (5.8a) whose norm square has the unity modulus:

$$\mathscr{N}^2 = C^0_{(n)}[\psi^{\text{ort}*}_{\mathscr{J}''}, \psi^{\text{ort}}_{\mathscr{J}''}] = \pm 1.$$

A solution to Eq. (2.2), which is a superposition of stationary states, is

$$\Psi^{\text{sup}}(t,\mathbf{x}) = \sum_{\mathscr{J}''} c_{\mathscr{J}''} \exp(-i\lambda_{\mathscr{J}''}t)\psi^{\text{ort}}_{\mathscr{J}''}(\mathbf{x}), \quad (5.9)$$

where $c_{\mathscr{J}''}$ are arbitrary complex coefficients. The orthogonality property (5.8a) brings the cross terms to zero; the norm square acquires the value

$$\mathscr{N}^2[\{\Psi^{\text{sup}*}\}, \{\Psi^{\text{sup}}\}] = \sum_{\mathscr{J}''} (\pm 1){c_{\mathscr{J}''}}^* c_{\mathscr{J}''}.$$

Note that the norm square is, as usual, an additive quantity. Keeping in mind the intention to explore nonlinear quantum mechanics with self-interaction where the superposition principle is invalid, we do not concern the completeness concept for the set $\psi^{\text{ort}}_{\mathscr{J}''}$. [See Refs. 6 and 7 where the problem of completeness for eigen- and associated (root) vectors is solved for certain classes of linear operators.] To our purposes it is enough to find a desired variational principle leading to solutions like (5.9).

VI. STATIONARY STATES IN THE *n*-COMPONENT FORMALISM

All the results of Secs. IV and V can be transfered *mutatis mutandi* to the n-component formalism.

The n-component version of a stationary state is given by the stationary n column

$$\Psi_{sK}(t,\mathbf{x}) = \exp(-i\lambda t)\psi_K(\mathbf{x}) \quad (6.1)$$

and the stationary row

$$\Psi^*_{sI}(t,\mathbf{x}) = \exp(+i\lambda t)\psi^*_I(\mathbf{x}) \quad (6.1')$$

($\operatorname{Im}\lambda = 0$). The substitution of (6.1) into (3.6) yields the generalized eigenvalue problem linear in λ, a certain matrix analog of the Schrödinger eigenvalue problem,

$$\lambda\hat{B}_{IK}\psi_K - \hat{H}_{IK}\psi_K = 0. \quad (6.2)$$

The contraction of Eq. (6.2) with ψ^*_I gives an equation like (4.8):

$$\mathscr{S}\langle\psi_I^*,\hat{B}_{IK}\psi_K\rangle - \langle\psi_I^*,\hat{H}_{IK}\psi_K\rangle = 0. \tag{6.3}$$

By analogy with Sec. IV, we assume that for all ψ_I^* and ψ_K satisfying (6.3)

$$\langle\psi_I^*,\hat{B}_{IK}\psi_K\rangle \neq 0,$$

then Eq. (6.3) has the unique root

$$\mathscr{S} = (\langle\psi_I^*,\hat{B}_{IK}\psi_K\rangle)^{-1}\langle\psi_L^*,\hat{H}_{LM}\psi_M\rangle, \tag{6.4}$$

it is real due to (3.7c) and plays the role of a quantum eigenaction.

The extremum condition of (6.4), by analogy with (i),

$$_{(3)}\frac{\delta\mathscr{S}}{\delta\psi_I^*} = -(\langle\psi_L^*,\hat{B}_{LM}\psi_M\rangle)^{-1} \times(\mathscr{S}\hat{B}_{IK}\psi_K - \hat{H}_{IK}\psi_K) = 0,$$

is equivalent to Eq. (6.2) on $\mathscr{S}$ and the set of extremum values of $\mathscr{S}$ coincide with the whole eigenvalue set.

Define the Rayleigh external α current in the n-component formalism

$$\mathscr{Q}_{(\alpha)} = -{}_{(3)}\frac{\delta\mathscr{S}}{\delta\alpha}, \tag{6.5}$$

then with the formula similar to (4.12c),

$$_{(3)}\frac{\delta\mathscr{S}}{\delta\alpha} = -(\langle\psi_L^*,\hat{B}_{LM}\psi_M\rangle)^{-1}\left(\mathscr{S}_{(3)}\frac{\delta}{\delta\alpha}\langle\psi_I^*,\hat{B}_{IK}\psi_K\rangle - {}_{(3)}\frac{\delta}{\delta\alpha}\langle\psi_I^*,\hat{H}_{IK}\psi_K\rangle\right),$$

the statement similar to (ii) can be made.

Concerning the question of the equivalence of the two formalisms, in the particular case of stationary states, the identifications (3.1) and (3.9) are

$$\psi_P^* = \lambda^P\psi^*, \quad \psi_P = \lambda^P\psi, \tag{6.6}$$

and the equivalence of the set of n equations (6.2) on the one hand and Eq. (4.2) on the other hand is easily seen: The equation

$$\lambda\hat{B}_{OK}\psi_K - \hat{H}_{OK}\psi_K = 0$$

($I=0$) is just Eq. (4.2) and the remaining equations ($I=1,2,\ldots,n-1$) are trivial. To get one more corroboration of this equivalence, we show what quantity in the monocomponent formalism the functional $\mathscr{S}$ corresponds to. Expressing it via the stationary states (4.1) of the monocomponent formalism (with real λ) and using (6.6), one obtains

$$\mathscr{S} = (\langle\psi^*,\hat{A}'_{(n)}(\lambda)\psi\rangle)^{-1}[\lambda\langle\psi^*,\hat{A}'_{(n)}(\lambda)\psi\rangle - \langle\psi^*,\hat{A}_{(n)}(\lambda)\psi\rangle] = \lambda|_{(4.2)},$$

thus the functional $\mathscr{S}$ differs from the spectral parameter λ by a term proportional to the left-hand side of Eq. (4.2), hence it coincides, in its extremum points, with the eigenvalues in the monocomponent formalism. Under the identifications (6.6), we have the same expressions (6.5) and (4.12a) in terms of the spectral parameter: for fixed $\mathscr{M}$, $\mathscr{Q}_{(\alpha)}$ coincide with $Q^{\mathscr{M}}_{(\alpha)}$ in each $\mathscr{J}_{\mathscr{M}}$ extremum point.

For different stationary states (6.1′) and (6.1), $\Psi^*_{s2\ I}$ and $\Psi^*_{s1\ K}$, with the eigenfunctions $\psi^*_{2\ I}$ and $\psi_{1\ K}$ and distinct eigenvalues λ_2 and λ_1, respectively, the orthogonality properties ensue from the conservation laws (3.8) (n component):

$$\langle\psi^*_{2\ I},\hat{B}_{IK}\psi_{1\ K}\rangle = 0, \tag{6.7a}$$

$$\langle\psi^*_{2\ I},\hat{H}_{IK}\psi_{1\ K}\rangle = 0, \tag{6.7b}$$

which coincide with (5.8a) and (5.8b), respectively, after the substitution of (6.6); the property (6.7b) is proportional to (6.7a) on Eq. (6.2).

Thus the mono- and n-component formalisms are both usable for the description of a quantum system, involving the set of eigenactions $S^{\cdot\mathscr{M}}$ or the unique eigenaction $\mathscr{S}$, respectively. An essential circumstance is, however, that the choice of $\mathscr{S}$ is appropriated to describe the whole spectrum, while the choice of a concrete eigenaction $S^{\cdot\mathscr{M}}$ may help one to exclude from consideration some "nonphysical" eigenvalues, for example, $\lambda<0$. The problem of making this choice is in the competence of the explorers.

VII. THE VARIATIONAL PRINCIPLES OF AVERAGED EXTREMALS AND THE INCLUSION OF GRAVITATIONAL SELF-INTERACTION

Consider the function

$$\Psi_{\sup}(t,\mathbf{x}) = \sum_{\mathscr{k}}\exp(-i\nu_{\mathscr{k}}t)\psi_{\mathscr{k}}(\mathbf{x}), \tag{7.1}$$

where $\mathscr{k}$ is a nonidentified index as yet, $\nu_{\mathscr{k}}$ are some numbers, and $\psi_{\mathscr{k}}$ are some functions; let all these quantities be not for the present connected with an eigenvalue problem. We want to express the functions $\psi_{\mathscr{k}}$ as functionals of $\Psi_{\sup}$ on the time interval $[0,T]$, it is possible, at least, at the limit $T\to\infty$. The limiting procedure is necessary since if the ratio of any two numbers $\mathscr{T}_{1,2} = 2\pi/\mathscr{k}_{1,2}:\mathscr{T}_1/\mathscr{T}_2$ is an irrational number, then there exists no finite common multiple T for all the $\mathscr{T}$, so that the integral equality we need is valid at the above limit:

$$\lim_{T\to\infty}T^{-1}\int_0^T dt\exp(i(\nu_{\mathscr{k}_1}-\nu_{\mathscr{k}_2})t) = \delta_{12}, \tag{7.2}$$

where δ is a Kronecker symbol. Using (7.2) the desired expression is

$$\psi_{\mathscr{k}}(\mathbf{x}) = \lim_{T\to\infty}T^{-1}\int_0^T dt\exp(i\nu_{\mathscr{k}}t)\Psi_{\sup}(t,\mathbf{x}). \tag{7.3}$$

Consider the functional (2.15) on the function (7.1) without assuming that (7.1) is a solution to Eq. (2.2) and make this functional smoothed over time, then, recalling the definition (4.6′)

$$\lim_{T\to\infty}T^{-1}\int_0^T dt\,\mathscr{N}^2[\{\Psi^*_{\sup}\},\{\Psi_{\sup}\}] = \lim_{T\to\infty}J'_{(n)}[\{\Psi^*_{\sup}\},\{\Psi_{\sup}\}] = \sum_{\mathscr{k}}J'_{(n)}[\psi^*_{\mathscr{k}},\psi_{\mathscr{k}};\nu_{\mathscr{k}}]. \tag{7.4}$$

If (7.4) exists, it means that the smoothed over time norm square for (7.1) is equal to the sum of partial norms square

for $\psi_{\mathcal{k}}$ without reference to the orthogonality properties like (5.8a).

Introduce a definition for the vatiational derivative of ψ_k with respect to Ψ_{sup}. To obtain it directly, the limiting procedure in (7.3) is an obstacle. Define

$$\frac{\delta\psi_{\mathcal{k}(1')}}{\delta\Psi_{\text{sup}}}$$

conditionally so as

$$\lim_{T\to\infty} T^{-1}\int_0^T dt\,{}_{(1')}\frac{\delta\psi_{\mathcal{k}_1}}{\delta\Psi_{\text{sup}}}\cdot\frac{d\Psi_{\text{sup}}}{d\psi_{\mathcal{k}_2}}=\delta_{12}$$

then

$${}_{(1')}\frac{\delta\psi_{\mathcal{k}}}{\delta\Psi_{\text{sup}}}=\exp(i\nu_{\mathcal{k}}t). \tag{7.5a}$$

Let $Y[\psi^*_{\mathcal{k}},\psi_{\mathcal{k}}]$ be some real functional of its arguments $Y=Y^*$, and define the variational derivative:

$${}_{(4')}\frac{\delta Y}{\delta\Psi_{\text{sup}}}=\sum_{\mathcal{k}}{}_{(3)}\frac{\delta Y}{\delta\psi_{\mathcal{k}}}\cdot{}_{(1')}\frac{\delta\psi_{\mathcal{k}}}{\delta\Psi_{\text{sup}}}, \tag{7.5b}$$

where the first multiplier on the right-hand side of (7.5b) was defined in Sec. II, the second one is given by (7.5a). Actually, we use below the functions complex conjugate to those in (7.1) and (7.3) as well as variational derivatives complex conjugate to those in (7.5).

Now we formulate the variational principles promised in Sec. I. These principles state that it is not the variational derivatives themselves but rather some weighted averages of them that should be employed. Let the two conditions (a) and (b) of Sec. IV be satisfied again and let (c) be

$$0\neq\left|\sum_{\mathcal{M},\mathcal{J}_{\mathcal{M}}} J'_{(n)}(S^{\cdot\mathcal{M}})\right|<\infty,$$

then: (iii) the equation

$$\left(\sum_{\mathcal{M},\mathcal{J}_{\mathcal{M}}} J'_{(n)}(S^{\cdot\mathcal{M}})\right)^{-1}\sum_{\mathcal{M},\mathcal{J}_{\mathcal{M}}} J'_{(n)}(S^{\cdot\mathcal{M}})$$
$$\times{}_{(3)}\frac{\delta S^{\cdot\mathcal{M}}}{\delta\psi^*}\exp(-iS^{\cdot\mathcal{M}}t)=0, \tag{7.6}$$

where the sum is taken over all the $\mathcal{M}$ and all the extrema $\mathcal{J}_{\mathcal{M}}$ of $S^{\cdot\mathcal{M}}$ with respect to the variables ψ^* and ψ as in the item (i) (see Sec. IV) is equivalent to Eq. (2.2):

$$\hat{A}_{(n)}\left(i\frac{\partial}{\partial t}\right)\tilde{\Psi}_{\text{sup}}=0,$$

for the function

$$\tilde{\Psi}_{\text{sup}}(t,\mathbf{x})=\sum_{\mathcal{M},\mathcal{J}_{\mathcal{M}}}\exp(-iS^{\cdot\mathcal{M}}t)\psi(\mathbf{x}). \tag{7.7a}$$

As before in (i), all quantities that referred to the extremum points coincide with the corresponding quantities of the eigenvalue problem with (7.8a) the superposition of eigenstates

$$\Psi_{\text{sup}}(t,\mathbf{x})=\sum_{\mathcal{M},\mathcal{J}_{\mathcal{M}}}\exp(-i\lambda_{\mathcal{J}_{\mathcal{M}}}t)\psi_{\mathcal{J}_{\mathcal{M}}}(\mathbf{x}), \tag{7.7b}$$

which is just (7.1) after the identifications $\mathcal{k}=(\mathcal{M},\mathcal{J}_{\mathcal{M}}),\nu_{\mathcal{k}}=\lambda_{\mathcal{J}_{\mathcal{M}}}$.

Some explanations are necessary. The index $\mathcal{J}_{\mathcal{M}}$ enumerating the extrema of $S^{\cdot\mathcal{M}}$ is suppressed at all the quantities in (7.6) in order to emphasize that (iii) does not state that the extrema must be known *a priori*, but states that the weighted average (7.6) of the extrema to be found leads to the solution (7.7b) to Eq. (2.2). One may, certainly, think these extrema independent thus returning to item (i) of Sec. IV. Each function $\psi_{\mathcal{J}_{\mathcal{M}}}$ determined from (iii) [or (i)] represents a one-parametric family of functions, it means it may differ, e.g., from $\psi^{\text{ort}}_{\mathcal{J}_{\mathcal{M}}}$ in Sec. V by an arbitrary complex coefficient $c_{\mathcal{J}_{\mathcal{M}}}$ including zero, giving rise to various functions (7.7b). A part of $\psi_{\mathcal{J}_{\mathcal{M}}}$ may not enter the superposition implying $\psi_{\mathcal{J}_{\mathcal{M}}}=0$ that yields the corresponding $J'_{(n)}(S^{\cdot\mathcal{M}})=0$ (cf. Sec. IV). Thus the number of nonzero functions $\psi_{\mathcal{J}_{\mathcal{M}}}$ is not fixed *a priori*.

To better illustrate (iii), after establishing the correspondence "eigenvalues–extremum points," Eq. (7.6) can be rewritten with the use of the function (7.7b) and the definition (7.5b):

$$\left(\sum_{\mathcal{M},\mathcal{J}_{\mathcal{M}}} J'_{(n)}(S^{\cdot\mathcal{M}})\right)^{-1}\sum_{\mathcal{M},\mathcal{J}_{\mathcal{M}}} J'_{(n)}(S^{\cdot\mathcal{M}})$$
$$\times{}_{(4')}\frac{\delta}{\delta\Psi^*_{\text{sup}}}S^{\cdot\mathcal{M}}[\psi^*_{\mathcal{J}_{\mathcal{M}}},\psi_{\mathcal{J}_{\mathcal{M}}}]\Big|_{\lambda_{\mathcal{J}_{\mathcal{M}}}=S^{\mathcal{M}}_{\mathcal{J}_{\mathcal{M}}}}=0.$$

The principle (iii) is invariant under the transformation (1.1) for the function (7.7), not for each $\psi_{\mathcal{J}_{\mathcal{M}}}$.

To prove (iii), the second multiplier in (7.6) is

$$-\sum_{\mathcal{M},\mathcal{J}_{\mathcal{M}}}\hat{A}_{(n)}(S^{\cdot\mathcal{M}})\psi\exp(-iS^{\cdot\mathcal{M}}t)$$
$$=-\sum_{\mathcal{M},\mathcal{J}_{\mathcal{M}}}\hat{A}_{(n)}\left(i\frac{\partial}{\partial t}\right)\tilde{\Psi}_{\text{sup}}.$$

As to the first multiplier in (7.6), in terms of the function (7.7a) the relation similar to (7.4) holds. Remembering that for every solution to Eq. (2.2), including (7.7b), the norm square is conserved in time, hence the time-smoothed norm square is equal to the nonsmoothed one.

(iv) The expression defined as the above weighted average of the Rayleigh α currents,

$$\left(\sum_{\mathcal{M},\mathcal{J}_{\mathcal{M}}} J'_{(n)}(S^{\cdot\mathcal{M}})\right)^{-1}\sum_{\mathcal{M},\mathcal{J}_{\mathcal{M}}} J'_{(n)}(S^{\cdot\mathcal{M}})Q^{\cdot\mathcal{M}}_{(\alpha)}=Q_{(\alpha)}(\mathbf{x}), \tag{7.8a}$$

where the sum is taken as in (iii), is equal to the time-smoothed bilinear α current (2.6) divided by the time-smoothed norm square on the function (7.7a):

$$Q_{(\alpha)}=(\mathcal{N}[\{\Psi^*_{\text{sup}}\},\{\Psi_{\text{sup}}\}])^{-2}$$
$$\times\lim_{T\to\infty}T^{-1}\int_0^T dt\,Q^{\text{bil}}_{(\alpha)}[\{\Psi^*_{\text{sup}}\},\{\Psi_{\text{sup}}\}]. \tag{7.8b}$$

Evidently, the expression (7.8b) can be reduced to

$$Q_{(\alpha)}=\left(\sum_{\mathcal{M},\mathcal{J}_{\mathcal{M}}} J'_{(n)}(S^{\cdot\mathcal{M}})\right)^{-1}$$
$$\times\sum_{\mathcal{M},\mathcal{J}_{\mathcal{M}}}{}_{(3)}\frac{\delta}{\delta\alpha}\langle\psi^*,\tilde{A}_{(n)}(S^{\cdot\mathcal{M}})\psi\rangle.$$

Using the equality, which is easily obtainable,

$$\lim_{T\to\infty} T^{-1}\int_0^T dt\,{}_{(4)}\frac{\delta}{\delta\alpha(t,\mathbf{x})}\overline{J}_{(n)} \times[\{\Psi^*_{\sup}\},\{\Psi_{\sup}\}]|_{\alpha(t,\mathbf{x})=\alpha(\mathbf{x})} = \sum_{\mathscr{M},\mathscr{J}_{\mathscr{M}}}{}_{(3)}\frac{\delta}{\delta\alpha}J_{(n)}(S^{\mathscr{M}})$$

one can check (7.8b). After establishing the above correspondence, (7.8b) gives the time-smoothed α current (2.6) on the superposition of stationary solutions divided by the norm square for (7.7b), which is the sum of the norm square for eigenfunctions involved.

Generally speaking, the variational principles formulated may include or not include summation over all the $\mathscr{M}$, otherwise they may include summation over a part of the $\mathscr{M}$ if we want to be confined to a physical situation chosen.

The extension of the above principles onto the n-component formalism is evident. Each variational derivative

$${}_{(3)}\frac{\partial\mathscr{S}}{\delta\cdots}$$

where $\mathscr{S}$ is the eigenaction (6.4), should be replaced by

$$\Big(\sum_{\mathscr{J}}\langle\psi^*_I,\hat{B}_{IK}\psi_K\rangle\Big)^{-1}\sum_{\mathscr{J}}\langle\psi^*_L,\hat{B}_{LM}\psi_M\rangle\,{}_{(3)}\frac{\delta\mathscr{S}}{\delta\cdots}.$$

Now we are ready to turn to the Klein–Gordon theory and to start the inclusion of gravitational self-interaction. Equation (4.8) and the condition (4.8′) in this case acquire the form

$$J_{(2)} = S^2\Big\langle\psi^*,\frac{1}{N}\psi\Big\rangle - S\Big\langle\psi^*,i\Big(\frac{N^a}{N}\nabla_a+\nabla_a\frac{N^a}{N}\Big)\psi\Big\rangle + \Big\langle\psi^*,\Big[\nabla_a\Big(NH^{ab}-\frac{N^aN^b}{N}\Big)\nabla_b-\mu^2N\Big]\psi\Big\rangle = 0, \tag{7.9}$$

$$J'_{(2)} = 2S\Big\langle\psi^*,\frac{1}{N}\psi\Big\rangle - \Big\langle\psi^*,i\Big(\frac{N^a}{N}\nabla_a+\nabla_a\frac{N^a}{N}\Big)\psi\Big\rangle \neq 0, \tag{7.9′}$$

a positive root of (7.9) being[19]

$$S^2 = \Big(2\Big\langle\psi^*,\frac{1}{N}\psi\Big\rangle\Big)^{-1}\Big\{\Big\langle\psi^*,i\Big(\frac{N^a}{N}\nabla_a+\nabla_a\frac{N^a}{N}\Big)\psi\Big\rangle + \Big[\Big(\Big\langle\psi^*,i\Big(\frac{N^a}{N}\nabla_a+\nabla_a\frac{N^a}{N}\Big)\psi\Big\rangle\Big)^2 - 4\Big\langle\psi^*,\frac{1}{N}\psi\Big\rangle\cdot\Big\langle\psi^*,\times\Big[\nabla_a\Big(NH^{ab}-\frac{N^aN^b}{N}\Big)\nabla_b-\mu^2N\Big]\psi\Big\rangle\Big]^{1/2}\Big\}.$$

Earlier,[1] the action functional was proposed for a stationary self-gravitating quantum system inside Ω:

$$S^{\text{tot}} = S^g + S^2 \tag{7.10}$$

(outside Ω, if we assume $\Psi=0$, then the action S^g is retained only, implying certain continuity of gravitational quantities on $\partial\Omega$),

$$S^g = -(16\pi k)^{-1}\int d\Omega\,[NP^a_a + N^{-1}\nabla_{(a}N_{b)}\nabla^{(a}N^{b)} - N^{-1}(\nabla^aN_a)^2], \tag{7.11}$$

where k is a gravitational constant, the parentheses denote symmetrization of indices, P_{ab} is the Ricci tensor of a stationary V_3; for other notations see Sec. II. The expression (7.11) represents, modulo a surface term, the usual Einstein–Hilbert action with the integration over time eliminated, so, in usual units, its dimensionality is energy. [Note that the negative root of (7.9) would lead to gravitational repulsion.] The extremum conditions of (7.10) are equivalent to the eigenvalue problem (4.2) with the operators (2.8b) which is restored here as ($S^2\to\lambda$)

$$\Big[\frac{\lambda^2}{N}-i\lambda\Big(\frac{N^a}{N}\nabla_a+\nabla_a\frac{N^a}{N}\Big) + \nabla_a\Big(NH^{ab}-\frac{N^aN^b}{N}\Big)\nabla_b-\mu^2N\Big]\psi=0 \tag{7.12}$$

and the set of the stationary Einstein equations in the 3 + 1 splitting. Henceforth, making the identification (4.13) and recalling the formula (4.12), these equations are

$$-8\pi kH^{-1/2}\,{}_{(3)}\frac{\delta S^g}{\delta N}\equiv N^2\Big(R^{00}+\frac{1}{2}N^{-2}R\Big)=8\pi kQ,$$
$$Q\equiv\mathscr{N}^{-2}\cdot[U(\psi^*,\psi)+V(\psi^*,\psi)+N^{-1}N^aW_a(\psi^*,\psi)];$$
$$8\pi kH^{-1/2}\,{}_{(3)}\frac{\delta S^g}{\delta N^a}\equiv NR^0_a=8\pi kQ_a,$$
$$Q_a\equiv\mathscr{N}^{-2}\cdot[W_a(\psi^*,\psi)+N^{-1}N^bZ_{ab}(\psi^*,\psi)];$$
$$-16\pi kH^{-1/2}N^{-1}\,{}_{(3)}\frac{\delta S^g}{\delta H^{ab}}\equiv R_{ab}-\frac{1}{2}H_{ab}R=8\pi kQ_{ab}, \tag{7.13}$$
$$Q_{ab}\equiv\mathscr{N}^{-2}\cdot\{Z_{ab}(\psi^*,\psi)-H_{ab}[U(\psi^*,\psi)-V(\psi^*,\psi)-N^{-1}N^cW_c(\psi^*,\psi)]\},$$

where

$$U=\tfrac{1}{2}H^{ab}Z_{ab}+\mu^2\psi^*\psi,$$
$$V=N^{-2}(\tfrac{1}{2}N^aN^bZ_{ab}+\lambda^2\psi^*\psi),$$
$$W_a=i\lambda N^{-1}(\psi^*_a\psi-\psi^*\psi_{,a}),$$
$$Z_{ab}=\psi^*_a\psi_{,b}+\psi^*_{,b}\psi_a,$$

R^{00}, R^0_a, and R_{ab} are the components of the Ricci 4-tensor and R is the Ricci scalar, $\mathscr{N}^{-2}$ is the norm (7.9′) inverse square. The right-hand sides of (7.13) are related to the definition (4.12a) and the components of the bilinear stress-energy tensor $T_{\mu\nu}$ as follows[19]

$$Q=-Q^2_{(N)}H^{-1/2}=\mathscr{N}^{-2}N^2T^{00},$$
$$Q_a=Q^2_{(N^a)}H^{-1/2}=\mathscr{N}^{-2}NT^0_a,$$
$$Q_{ab}=-2Q^2_{(H^{ab})}H^{-1/2}N^{-1}=\mathscr{N}^{-2}T_{ab}.$$

The variational principle of extremality of (7.10) describes a particle in a single stationary state and is irrelevant for a particle in a superposition of stationary states. In the latter case the bilinear stress-energy tensor $T_{\mu\nu}$ contains os-

cillating cross terms with the factors like $\exp(i(\lambda_2-\lambda_1)t)$ and thus is nonstationary. That the oscillating terms should disappear after smoothing in time suggests to describe the time-smoothed situation by the approximate action functional[1]

$$S^{\text{tot}} = S^g + \sum_j p_j S^2[\psi_j^*, \psi_j]. \tag{7.14}$$

Here, S^g and S^2 are the above functionals, the numbers p_j play the role of some probabilities given *a priori*, $0 \leqslant p_j \leqslant 1$ and $\Sigma_j p_j = 1$, the index j has an abstract meaning at the present, ψ_j are the functions to be determined. The extremum conditions for (7.14) lead to the same equations as (7.12) for each ψ_j and the set of stationary Einstein equations with the unchanged left-hand sides [cf. (7.13)] but with other sources:

$$N^2\left(R^{00} + \frac{1}{2}N^{-2}R\right) = 8\pi k \sum_j p_j Q(\psi_j^*, \psi_j),$$

$$NR_a^0 = 8\pi k \sum_j p_j Q_a(\psi_j^*, \psi_j), \tag{7.15}$$

$$R_{ab} - \frac{1}{2}H_{ab}R = 8\pi k \sum_j p_j Q_{ab}(\psi_j^*, \psi_j)$$

(each quantity "Q" contains the inverse norm square of ψ_j).

The variational principle of extremality of (7.14) has the drawback that the probabilities are fixed and not connected directly with the solutions ψ_j. Moreover, the action (7.14) and the equations ensued from it are invariant under the range of transformations wider than (1.1):

$$\psi_j \to a_j \psi_j, + \text{ complex conjugate}, \tag{7.16}$$

where a_j are arbitrary complex numbers attached to each ψ_j. The requirement of invariance under (7.16) is unnecessary and is not fulfilled in the usual Schrödinger theory. In connection with that, we include gravitational self-interaction into the variational principle of averaged extremals.

Take as a basis the action (7.10) and extend to it the principles (iii) and (iv). Because S^g is independent explicitly on ψ^* and ψ the principle (iii) for the operators (2.8b) stands as before with the only difference that it must be formulated in common with the following principle.

Let the conditions (a), (b), and (c) be satisfied, then:

(v) the equations

$$\left(\sum_{\mathscr{J}_2} J'_{(2)}(S^2)\right)^{-1} \sum_{\mathscr{J}_2} J'_{(2)}(S^2) \cdot_{(3)} \frac{\delta S^{\text{tot}}}{\delta \alpha}$$
$$= \left(\sum_{\mathscr{J}_2} J'_{(2)}(S^2)\right)^{-1} \sum_{\mathscr{J}_2} J'_{(2)}(S^2)$$
$$\cdot_{(3)} \left(\frac{\delta S^g}{\delta \alpha} +_{(3)} \frac{\delta S^2}{\delta \alpha}\right) = 0,$$

where the sum is taken over all the extrema of S^2 with respect to ψ^* and ψ, are equivalent to the set of stationary Einstein equations with the time-smoothed stress-energy tensor from (iv).

With S^g independent explicitly of ψ^* and ψ

$$_{(3)}\frac{\delta S^g}{\delta \alpha} + \left(\sum_{\mathscr{J}_2} J'_{(2)}(S^2)\right)^{-1} \sum_{\mathscr{J}_2} J'_{(2)}(S^2) \cdot_{(3)} \frac{\delta S^2}{\delta \alpha} = 0. \tag{7.17}$$

Clearly, the set (7.17) is just the set (7.15) after identifying $j = \mathscr{J}_2$ and

$$p_{j_k} = p_{\mathscr{J}_{k'}} = \left(\sum_{\mathscr{J}_2} J'_{(2)}(S^2)\right)^{-1} \cdot J'_{(2)}(S^2{}_{\mathscr{J}_{k'}}).$$

The probabilities $p_{\mathscr{J}_2}$ are now determined as the functionals of $\psi_{\mathscr{J}_2}$, not arbitrary numbers. As it was shown in Sec. IV and Ref. 1, $J'_{(2)}(S^2) \geqslant 0$ for any gravitational field, thus due to (7.9') for each $\mathscr{J}_2$: $p_{\mathscr{J}_2} > 0$ if $\psi_{\mathscr{J}_2} \neq 0$ and $p_{\mathscr{J}_2} = 0$ if $\psi_{\mathscr{J}_2} = 0$; it means that the number of nonzero $p_{\mathscr{J}_2}$ and $\psi_{\mathscr{J}_2}$ is not fixed before finding the solutions. Obviously,

$$\sum_{\mathscr{J}_2} p_{\mathscr{J}_2} = 1.$$

The inclusion of any interaction, e.g., gravitational, brings intrinsic nonlinearity into a quantum system.[20] The superposition principle is invalid now: a linear combination of solutions is no longer a solution. In the presence of self-consistent gravity, the principle (iii) can be also reduced to the principle (i). However, the extrema of S^2 cannot be found individually, but only simultaneously in a common self-consistent metric that enters the operators $\hat{A}_k$. The space of states is now non-Hilbert; it is separated into finite-or infinite-dimensional nonlinear subspaces, depending on the number of the functions $\psi_{\mathscr{J}_2}$ involved. In other words, each of these subspaces contains the states $\psi_{\mathscr{J}_2}$ which are self-consistent solutions (not combinations of those) creating a common time-smoothed gravitational field. Within every subspace, all the results of Sec. II concerning conserved functionals, inner products norms, and expectation values are true. The cause is that the nature of any function $\alpha(\mathbf{x})$ was not specified. Likewise, for eigenfunctions ψ_2 and ψ_1 with distinct eigenvalues $\lambda_2 \neq \lambda_1$, the orthogonality property of the type (5.8a) in the common metric holds:

$$\left\langle \psi_2^*, \left[N^{-2}(\lambda_2 + \lambda_1) - i\left(\frac{N^a}{N}\nabla_a + \nabla_a \frac{N^a}{N}\right)\right]\psi_1 \right\rangle = 0.$$

We do not speculate here about the statistical interpretation, it was done in Ref. 1. Our main assertions remain unchanged after replacing the variational principle of extremality of (7.1) by that of averaged extremals.

In conclusion, first, we should like to add that the Klein–Gordon theory in the two-component formalism may be used to include self-interaction as well, but then the difficulty of the norm nonpositivity arises again. Hence, to our opinion, the monocomponent formalism with the chosen S^2 is more preferable. Nevertheless, this difficulty is less essential in nonlinear quantum mechanics just owing to the failure of the superposition principle; at any rate a "general" solution cannot be decomposed upon individual eigenfunctions. Second, we emphasize the importance of the rescaling invariance (1.1) for a quantum system and ask whether the extra-

polation of the variational principle of extremality from classical theories to the quantum sphere is true at all? Possibly, the variational principle of averaged extremals is more suitable for quantum systems. Otherwise, we must choose between rescaling invariance and the extremality of an action.

ACKNOWLEDGMENT

The author would like to thank V. D. Ivashchuk for helpful discussions.

[1]A. D. Popova, Int. J. Mod. Phys. A **4**, 3229 (1989).
[2]In this paper, we reserve the word "theory" for quantum-mechanical theories only, without second quantization.
[3]H. Feshbach and F. Villars, Rev. Mod. Phys. **30**, 24 (1958).
[4]V. A. Rizov, H. Sazdjian, and I. T. Todorov, Ann. Phys. **165**, 59 (1985).
[5]R. Weder, Helv. Phys. Acta **50**, 105 (1977); J. Funct. Anal. **27**, 100 (1978).
[6]M. V. Keldysh, Math. Nauk. **26** (4), 15 (1971).
[7]V. N. Vizitii and A. S. Markus, Mat. Sbornik **66** (108), 287 (1965), in Russian.
[8]N. Nakanishi, Prog. Theor. Phys. Suppl. **43**, 1 (1969).
[9]R. N. Faustov, Teor. Mat. Fiz. **3**, 240 (1970), in Russian.
[10]T. W. B. Kibble, Commun. Math. Phys. **64**, 73 (1978).
[11]We use the units where $\hbar = c = 1$ and unless otherwise unspecified, the notations and conventions of the book: C. W. Misner, K. S. Thorne, and J. A. Wheeler, *Gravitation* (Freeman and Company, San Francisco, 1973); μ^2 is the mass square; three-space indices are denoted by first several Latin letters, whereas the operator indices and orders of time derivatives are denoted by letters from the middle of the Latin alphabet.
[12]J. W. Rayleigh, Proc. London Math. Soc. **4**, 357 (1873).
[13]R. G. Duffin, J. Rat. Mech. Anal. **4**, 221 (1955).
[14]A. S. Marcus, V. I. Matsayev, and G. I. Russu, Acta Sci. Math. Szeged. **34**, 245 (1973). Note that the operator (4.15) considered in Sec. IV is the hyperbolic operator of the second order.
[15]Yu. Sh. Abramov, *Variational Methods in the Theory of Operator Bundles, Spectral Optimisation* (Leningrad University, Leningrad, 1983), in Russian.
[16]A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1966), Vol. 2.
[17]Yu. A. Pallant, Dokl. Akad. Nauk **141**, 558 (1961), in Russian.
[18]M. B. Orazov and G. V. Radzievskii, Siberian Mat. Zhurn. **26**, 572 (1975), in Russian.
[19]Not to confuse the index 2 and the second power of S, Q^2 means $Q^{\mathscr{M}}$ for $\mathscr{M} = 2$.
[20]B. Mielnik, Commun. Math. Phys. **37**, 221 (1974).

An asymptotic analysis and its application to the nonrelativistic limit of the Pauli–Fierz and a spin-boson model

Asao Arai
Department of Mathematics, Hokkaido University, Sapporo 060, Japan

(Received 23 February 1990; accepted for publication 20 June 1990)

An abstract asymptotic theory of a family of self-adjoint operators $\{H_\kappa\}_{\kappa>0}$ acting in the tensor product of two Hilbert spaces is presented and it is applied to the nonrelativistic limit of the Pauli–Fierz model in quantum electrodynamics and of a spin-boson model. It is proven that the resolvent of H_κ converges strongly as $\kappa\to\infty$ and the limit is a pseudoresolvent, which defines an "effective operator" of H_κ at $\kappa\approx\infty$. As corollaries of this result, some limit theorems for H_κ are obtained, including a theorem on spectral concentration. An asymptotic estimate of the infimum of the spectrum (the ground state energy) of H_κ is also given. The application of the abstract theory to the above models yields some new rigorous results for them.

I. INTRODUCTION

This paper consists of two parts: one is concerned with an abstract asymptotic theory of a family of self-adjoint operators and the other presents its application to the nonrelativistic limit of the Pauli–Fierz[1-9] and a spin-boson model.[10-15]

The Pauli–Fierz model is a model in quantum electrodynamics and describes a nonrelativistic one-electron atom coupled to a quantized radiation field. It is known that the model is a realistic one in the sense that in a nonrelativistic region, it explains well some physical phenomena such as the Lamb shift, although the explanations are usually done by using formal perturbation calculations to which rigorous mathematical basis has not yet been given. Only a few mathematically rigorous results have been obtained for the model.[4,5,7,9] The spin-boson model we consider describes a two-level atom coupled to a quantized Bose field and can be regarded as a simplified version of the Pauli–Fierz model.[8] The nonrelativistic limit we study on these models is a scaling limit of the speed of light at the same time as the coupling constant of the models gets a scale transformation, which, as far as we know, has not been discussed in the literature.

To treat the problem of the nonrelativistic limit of the Pauli–Fierz and the spin-boson model in a unified way, we first present in Sec. II an abstract asymptotic theory of a family of self-adjoint operators $\{H_\kappa\}_{\kappa>0}$ acting in the tensor product of two Hilbert spaces. The self-adjoint operator H_κ is an abstract version of operators *unitarily equivalent* to Hamiltonians of some models of an atom coupled to a quantized radiation field, including the Pauli–Fierz and the spin-boson model. We prove that the resolvent of H_κ converges strongly as $\kappa\to\infty$ and the limit is a pseudoresolvent, which defines an "effective" operator of H_κ at $\kappa\approx\infty$. Introducing a concept of "partial expectation" of operators, we represent the effective operator more explicitly. In applications, partial expectations can be used also to describe "fluctuations" caused by a quantized radiation field on an atom (see Sec. III). Further, we obtain an asymptotic estimate of the infimum of the spectrum (the ground state energy) of H_κ. The abstract theory presented here is closely related to asymptotic theories given in Refs. 16–18. But our class of H_κ is different from the operators considered there in the scaling order with respect to κ. There may be different asymptotic theories depending on the scaling order of κ and the form of the relevant operators. We also discuss the spectral concentration of H_κ.

In Sec. III we discuss the Pauli-Fierz model, which, as mentioned above, describes a one-electron atom coupled to a quantized radiation field. For a mathematical generality, we consider the case where the one-electron atom is placed in the d-dimensional space ($d\geqslant 2$). The total Hamiltonian of the model is given by a self-adjoint operator $H(c,e)$ with parameters $c>0$ and $e\in\mathbb{R}\setminus\{0\}$ denoting the speed of light and the elementary charge (the coupling constant in this model), respectively. The scaled Hamiltonian is defined by $H(\kappa)=H(c(\kappa),e(\kappa))$ with $c(\kappa)=\kappa c$ and $e(\kappa)=\kappa^{3/2}e$. The nonrelativistic limit we study is taken in the sense of the scaling limit $\kappa\to\infty$. Since $|e(\kappa)|\to\infty$ as $\kappa\to\infty$, the nonrelativistic limit is a scaling limit of the speed of light at the same time as the magnitude of the coupling constant becomes infinite. We show that $H(\kappa)$ is unitarily equivalent to an operator $\tilde{H}(\kappa)$, which is of the form of H_κ discussed in Sec. II. Applying the abstract theory in Sec. II to $\tilde{H}(\kappa)$, we find that the effective operator of $\tilde{H}(\kappa)$ is a Schrödinger operator $H_{A,\mathrm{eff}}$. In the case $d=3$, the potential operator of $H_{A,\mathrm{eff}}$ coincides with the effective potential that Welton[3] proposed to calculate some observable effects of the quantized radiation field such as the Lamb shift. In Ref. 3 the effective potential was derived by physical arguments. We derive it as a scaling limit in the sense described above, starting from the total Hamiltonian $H(\kappa)$. This does not only justify rigorously the effective potential of Welton but also clarifies a mathematical meaning of it, in other words, in what sense the effective potential is "effective." Further, we show that the ground state energy of the model is nondecreasing as a function of κ and obtain an estimate of the ground state energy, which, to our knowledge, has not been given so far in the literature. We also prove that the spectrum of $H(\kappa)$ is asymptotically concentrated on the spectrum of $H_{A,\mathrm{eff}}$ "locally" as $\kappa\to\infty$.

In Sec. IV we consider the spin-boson model. The nonrelativistic limit of this model is also a scaling limit of the

 0022-2488/90/112653-11$03.00

speed of light at the same time as the magnitude of the coupling constant becomes infinite, but the scaling order of the coupling constant is different from that of the Pauli–Fierz model. We show that the total Hamiltonian of the model is unitarily equivalent to an operator $\tilde{H}(\kappa)$ of the form of H_κ in Sec. II. We derive the effective operator of $\tilde{H}(\kappa)$. Moreover, we show that the ground state energy is nondecreasing in the scaling parameter κ and obtain an estimate of the ground state energy, which slightly improves that given by Davies.[10] We also give a meaning to the transition probability between the two degenerate ground states of the model without the atom part (cf. Ref. 11). Finally, we prove the existence of a "local" spectral concentration of the total Hamiltonian.

In the last section some remarks are given. We conclude the present paper with an Appendix, where we prove some limit theorems related to the strong convergence of resolvents in which the limiting operator is a pseudoresolvent.

II. AN ABSTRACT ASYMPTOTIC THEORY

In this section we present an abstract asymptotic theory for a class of self-adjoint operators. The theory developed below may be formulated in a more abstract setting using a Banach space as in Ref. 16 and for a more general class of operators. In the present paper, however, we take a Hilbert space formulation and restrict our consideration to a class of self-adjoint operators, which allows us to obtain more concrete and stronger results in some respects.

In what follows, we use the following notation: $(\cdot,\cdot)_{\mathscr{H}}$ and $\|\cdot\|_{\mathscr{H}}$ denote the inner product and the norm of the Hilbert space $\mathscr{H}$, respectively. If there is no danger of confusion, then we omit the subscript $\mathscr{H}$ of them. The domain (resp. range) of an operator T is denoted by $D(T)$ (resp. Ran T). For bounded operators T, we denote by $\|T\|$ the operator norm. By I we denote identity.

Let $\mathscr{H}$ and $\mathscr{K}$ be two Hilbert spaces and

$$\mathscr{L} = \mathscr{H} \otimes \mathscr{K}. \tag{2.1}$$

Let A and B be non-negative self-adjoint operators in $\mathscr{H}$ and $\mathscr{K}$, respectively. We assume that

$$\text{Ker } B \neq \{0\}. \tag{2.2}$$

Let $\{C_\kappa\}_{\kappa>0}$ be a family of symmetric operators in $\mathscr{L}$ satisfying the following conditions.

(i) For all $\kappa>0$, $D(A\otimes I)\subset D(C_\kappa)$ and $C_\kappa(A\otimes I+\lambda)^{-1}$ is bounded for all $\lambda>0$ with

$$\lim_{\lambda\to\infty}\|C_\kappa(A\otimes I+\lambda)^{-1}\|=0,$$

where the convergence is uniform in $\kappa\geqslant\kappa_0$ for some $\kappa_0>0$.

(ii) For all $\lambda>0$, $C_\kappa(A\otimes I+\lambda)^{-1}$ is strongly continuous in $\kappa>0$.

(iii) There exists a symmetric operator C in $\mathscr{L}$ such that $D(A\otimes I)\subset D(C)$ and for all $\lambda>0$,

$$\text{s-}\lim_{\kappa\to\infty} C_\kappa(A\otimes I+\lambda)^{-1}=C(A\otimes I+\lambda)^{-1},$$

where s-lim means strong limit.

For each $\kappa>0$, we define

$$H_{0,\kappa}=A\otimes I+\kappa I\otimes B. \tag{2.3}$$

The above property (i) of C_κ implies that for every $\epsilon>0$, there exists a constant $\lambda_0=\lambda_0(\epsilon,\kappa_0)>0$, independent of $\kappa\geqslant\kappa_0$, such that for all $\lambda\geqslant\lambda_0$,

$$\|C_\kappa\Psi\|<\epsilon\|(A\otimes I+\lambda)\Psi\|,\quad \Psi\in D(A\otimes I). \tag{2.4}$$

Since $I\otimes B$ is non-negative and commutes with $A\otimes I$, it follows that for $\lambda\geqslant\lambda_0$,

$$\|C_\kappa\Psi\|<\epsilon\|(H_{0,\kappa}+\lambda)\Psi\|,$$
$$\Psi\in D(H_{0,\kappa})=D(A\otimes I)\cap D(I\otimes B)\equiv D_{A,B}. \tag{2.5}$$

Hence, C_κ with $\kappa\geqslant\kappa_0$ is infinitesimally small with respect to $H_{0,\kappa}$. Therefore, by the Kato–Rellich theorem (e.g., Refs. 19 and 20), the operator

$$H_\kappa=H_{0,\kappa}+C_\kappa \tag{2.6}$$

with $\kappa\geqslant\kappa_0$ is self-adjoint on $D_{A,B}$ and essentially self-adjoint on every core of $H_{0,\kappa}$. Further, H_κ is bounded from below with

$$H_\kappa\geqslant-\epsilon\lambda/(1-\epsilon), \tag{2.7}$$

where $0<\epsilon<1$ and $\lambda\geqslant\lambda_0(\epsilon,\kappa_0)$. The ground state energy E_κ of H_κ is defined by

$$E_\kappa=\inf\sigma(H_\kappa), \tag{2.8}$$

where $\sigma(T)$ denotes the spectrum of operator T. By (2.7), E_κ is bounded from below uniformly in $\kappa\geqslant\kappa_0$:

$$\epsilon_0\equiv\inf_{\kappa>\kappa_0}E_\kappa>-\infty. \tag{2.9}$$

Our aim is to consider the limit $\kappa\to\infty$ of H_κ and to give an asymptotic estimate of E_κ for large κ.

Let P_0 be the orthogonal projection from $\mathscr{K}$ onto Ker B. Then it follows from property (iii) of C_κ and (2.4) that $(I\otimes P_0)C(I\otimes P_0)$ is infinitesimally small with respect to $A\otimes I$. Hence, by the Kato–Rellich theorem again, the operator

$$H_\infty=A\otimes I+(I\otimes P_0)C(I\otimes P_0) \tag{2.10}$$

is self-adjoint on $D(A\otimes I)$ and bounded from below. It is easy to see that the resolvent of H_∞ commutes with $I\otimes P_0$. Hence, H_∞ is reduced by Ran $I\otimes P_0=\mathscr{H}\otimes\text{Ker }B$. We define

$$E_\infty=\inf\sigma(H_\infty\restriction\mathscr{H}\otimes\text{Ker }B). \tag{2.11}$$

The first of the main results in this section is the following.

Theorem 2.1: For all $z\in\mathbb{C}$ with Im $z\neq0$ or for $z<0$ with $|z|$ sufficiently large, $(H_\kappa-z)^{-1}$ is strongly continuous in $\kappa\geqslant\kappa_0$ and

$$\text{s-}\lim_{\kappa\to\infty}(H_\kappa-z)^{-1}=(H_\infty-z)^{-1}(I\otimes P_0). \tag{2.12}$$

Further,

$$\overline{\lim_{\kappa\to\infty}}E_\kappa\leqslant E_\infty. \tag{2.13}$$

Proof: Let $\lambda>0$ be sufficiently large so that $-\lambda\in\rho(H_\kappa)\cap\rho(H_\infty)\cap\rho(H_{0,\kappa})$ for all $\kappa\geqslant\kappa_0$, where $\rho(T)$ denotes the resolvent set of T. Iterating the second resolvent formula with respect to the pair $(H_\kappa,H_{0,\kappa})$, we have

$$(H_\kappa+\lambda)^{-1}=\sum_{n=0}^{N}(-1)^n(H_{0,\kappa}+\lambda)^{-1}T_\kappa^n+R_N(\kappa),$$

where

$$T_\kappa = C_\kappa (H_{0,\kappa} + \lambda)^{-1},$$

and

$$R_N(\kappa) = (-1)^{N+1}(H_\kappa + \lambda)^{-1} T_\kappa^{N+1}.$$

It follows from (2.5) that

$$\|R_N(\kappa)\| < (\epsilon_0 + \lambda)^{-1} \epsilon^{N+1}.$$

Hence, taking $\epsilon < 1$, we see that for $\lambda > 0$ sufficiently large

$$(H_\kappa + \lambda)^{-1} = \sum_{n=0}^{\infty} (-1)^n (H_{0,\kappa} + \lambda)^{-1} T_\kappa^n \tag{2.14}$$

is norm convergent uniformly in $\kappa \geqslant \kappa_0$. It is easy to see that

$$\text{s-}\lim_{\kappa \to \infty} (H_{0,\kappa} + \lambda)^{-1} = (A \otimes I + \lambda)^{-1} I \otimes P_0.$$

Further, by property (iii) of C_κ, we have

$$\text{s-}\lim_{\kappa \to \infty} T_\kappa = C(A \otimes I + \lambda)^{-1}(I \otimes P_0).$$

By the uniform convergence of the series on the right-hand side (rhs) of (2.14), we can interchange the limit $\kappa \to \infty$ and the summation Σ_n to obtain

$$\text{s-}\lim_{\kappa \to \infty} (H_\kappa + \lambda)^{-1} = \sum_{n=0}^{\infty} (-1)^n (A \otimes I + \lambda)^{-1} \times \{\tilde{C}(A \otimes I + \lambda)^{-1}\}^n (I \otimes P_0), \tag{2.15}$$

where

$$\tilde{C} = (I \otimes P_0) C (I \otimes P_0),$$

and we have used the fact that $I \otimes P_0$ is a projection. The rhs of (2.15) is equal to

$$(H_\infty + \lambda)^{-1}(I \otimes P_0).$$

Thus (2.12) with $z = -\lambda$ follows. Once (2.12) is proved for some $z = -\lambda \in \mathbb{R} \cap \rho(H_\kappa) \cap \rho(H_\infty)$, it can be extended to the case $\operatorname{Im} z \neq 0$ by mimicking a standard argument for resolvents (e.g., the proof of Theorem VIII.19 in Ref. 21). The strong continuity of $(H_\kappa - z)^{-1}$ in κ follows similarly. Inequality (2.13) follows from an application of Theorem A.1 in the Appendix. ■

Theorems 2.1 can be generalized by the following theorem.

Theorem 2.2: Denote by $C_\infty(\mathbb{R})$ the space of continuous functions on $\mathbb{R}$ vanishing at ∞. Then, for all $F \in C_\infty(\mathbb{R})$,

$$\text{s-}\lim_{\kappa \to \infty} F(H_\kappa) = F(H_\infty) I \otimes P_0.$$

Proof: This follows from Theorem 2.1 and an application of Theorem A.1 in the Appendix. ■

We next consider the asymptotic behavior of the ground state energy E_κ. Concerning this problem, we have been able to obtain a result only in the case where C_κ and C are bounded. Let

$$H_\infty(\kappa) = A \otimes I + (I \otimes P_0) C_\kappa (I \otimes P_0) \tag{2.16}$$

and

$$E_\infty(\kappa) = \inf \sigma(H_\infty(\kappa) \upharpoonright \mathscr{H} \otimes \operatorname{Ker} B). \tag{2.17}$$

Lemma 2.3: Let C_κ and C be bounded. Then, for all $\kappa > 0$,

$$|E_\infty - E_\infty(\kappa)| \leqslant \|C_\kappa - C\|. \tag{2.18}$$

In particular, if $\|C_\kappa - C\| \to 0$ as $\kappa \to \infty$, then

$$\lim_{\kappa \to \infty} E_\infty(\kappa) = E_\infty.$$

Proof: For $\Psi \in D(A \otimes P_0)$ with $\|\Psi\| = 1$, we have

$$|(\Psi, H_\infty(\kappa)\Psi) - (\Psi, H_\infty \Psi)| \leqslant \|C_\kappa - C\|,$$

which, combined with the variational principle, gives (2.18). ■

An estimate of the ground state energy E_κ is given by the following theorem.

Theorem 2.4: Let C_κ and C be bounded. Suppose that $B \upharpoonright (\operatorname{Ker} B)^\perp \geqslant b$ with some constant $b > 0$. Then, for all $\kappa > (E_\infty(\kappa) + \|C_\kappa\|)/b$,

$$E_\infty(\kappa) - \|C_\kappa\| \eta_\kappa (1 + \sqrt{1 + \eta_\kappa^2})^{-1} \leqslant E_\kappa \leqslant E_\infty(\kappa), \tag{2.19}$$

where

$$\eta_\kappa = 2\|C_\kappa\| / [b\kappa - E_\infty(\kappa) - \|C_\kappa\|].$$

In particular, if $\|C_\kappa - C\| \to 0$ as $\kappa \to \infty$, then,

$$\lim_{\kappa \to \infty} E_\kappa = E_\infty. \tag{2.20}$$

Proof: We have

$$\mathscr{H} = \mathscr{H}_1 \oplus \mathscr{H}_2,$$

where

$$\mathscr{H}_1 = \mathscr{H} \otimes \operatorname{Ker} B, \quad \mathscr{H}_2 = \mathscr{H} \otimes (\operatorname{Ker} B)^\perp.$$

Let P_j $(j = 1, 2)$ be the orthogonal projection from $\mathscr{H}$ onto $\mathscr{H}_j$. It is easy to see that

$$P_1 = I \otimes P_0, \quad P_2 = I \otimes (I - P_0).$$

We can write

$$H_\kappa = H_\infty(\kappa) + \kappa I \otimes B(I - P_0) + P_1 C_\kappa P_2 + P_2 C_\kappa P_1 + P_2 C_\kappa P_2.$$

For all $\Psi \in \mathscr{H}_1 \cap D_{A,B}$ with $\|\Psi\| = 1$, we have

$$(\Psi, H_\kappa \Psi) = (\Psi, H_\infty(\kappa)\Psi).$$

Hence, it follows from the variational principle that

$$E_\kappa \leqslant (\Psi, H_\infty(\kappa)\Psi),$$

which implies the second inequality of (2.19). [Note that $D_{A,B}$ is a core of $H_\infty(\kappa)$.]

To prove the first inequality of (2.19), we write $\Psi \in D_{A,B}$ with $\|\Psi\| = 1$ as

$$\Psi = \Psi_1 + \Psi_2,$$

with $\Psi_j \in \mathscr{H}_j$ $(j = 1,2)$. Then, using the Schwarz inequality and the fact that A is non-negative and $B \upharpoonright (\operatorname{Ker} B)^\perp \geqslant b > 0$, we have

$$(\Psi, H_\kappa \Psi) \geqslant E_\infty(\kappa)\|\Psi_1\|^2 + (b\kappa - \|C_\kappa\|)\|\Psi_2\|^2 - 2\|C_\kappa\| \, \|\Psi_1\| \, \|\Psi_2\|.$$

Since

$$\|\Psi_2\|^2 = 1 - \|\Psi_1\|^2,$$

it follows that

$$E_\kappa \geqslant E_\infty(\kappa) - \sup_{0<x<1} \lambda(x),$$

where

$$\lambda(x) = (b\kappa - E_\infty(\kappa) - \|C_\kappa\|)x^2 + 2\|C_\kappa\| x\sqrt{1-x^2} - (b\kappa - \|C_\kappa\| - E_\infty(\kappa)).$$

It is easy to show that the inequality

$$\alpha x^2 + \beta x\sqrt{1-x^2} - \alpha \leqslant \tfrac{1}{2}(-\alpha + \sqrt{\alpha^2+\beta^2}) = \frac{\beta}{2}\left(\frac{\beta}{\alpha}\right)\left(1+\sqrt{1+\left(\frac{\beta}{\alpha}\right)^2}\right)^{-1}$$

holds for all $\alpha > 0, \beta \geqslant 0$ and $0 \leqslant x \leqslant 1$. Applying this inequality with $\alpha = b\kappa - E_\infty(\kappa) - \|C_\kappa\| > 0$ and $\beta = 2\|C_\kappa\|$, we obtain the first inequality of (2.19). Formula (2.20) follows from (2.19), Lemma 2.3, and the fact that $\eta_\kappa \to 0$ $(\kappa \to \infty)$. ∎

Remark: As the above proof shows, the second inequality of (2.19) holds also for the case where C_κ and C are not bounded.

In order to write H_∞ in a more explicit way, we introduce a concept of "partial expectation" for linear operators. For $S \in \mathbb{B}(\mathscr{H})$ (the space of all bounded linear operators on $\mathscr{H}$) and $f,g \in \mathscr{K}$, we define the sesquilinear form $q_{f,g}(\cdot,\cdot)$ on $\mathscr{H} \times \mathscr{H}$ by

$$q_{f,g}(u,v) = (u \otimes f, S(v \otimes g))_{\mathscr{H}}, \quad u,v \in \mathscr{H},$$

which is bounded with

$$|q_{f,g}(u,v)| \leqslant \|S\| \, \|f\| \, \|g\| \, \|u\| \, \|v\|.$$

Therefore, by the Riesz lemma, there exists a unique $E_{f,g}(S) \in \mathbb{B}(\mathscr{H})$ such that

$$(u \otimes f, S(v \otimes g))_{\mathscr{H}} = (u, E_{f,g}(S)v)_{\mathscr{H}}$$

and

$$\|E_{f,g}(S)\| \leqslant \|f\| \, \|g\| \, \|S\|.$$

We also define $E_f(S) \in \mathbb{B}(\mathscr{H})$ by

$$E_f(S) = E_{f,f}(S).$$

We call the operator $E_{f,g}(S)$ [resp. $E_f(S)$] the *partial expectation* of S with respect to $\{f,g\}$ (resp. f). Note that, in the case $S = L \otimes M$ with $L \in \mathbb{B}(\mathscr{H})$ and $M \in \mathbb{B}(\mathscr{K})$, we have

$$E_{f,g}(L \otimes M) = (f, Mg)_{\mathscr{K}} L.$$

Some elementary facts of $E_{f,g}(S)$ are summarized in the following proposition, whose proof is left to the reader.

Proposition 2.5: (i) For all $f,g,h \in \mathscr{K}$, $\alpha,\beta \in \mathbb{C}$, and $S \in \mathbb{B}(\mathscr{H})$,

$$E_{h,\alpha f+\beta g}(S) = \alpha E_{h,f}(S) + \beta E_{h,g}(S).$$

(ii) For all $f,g \in \mathscr{K}$, $\alpha,\beta \in \mathbb{C}$, and $S,T \in \mathbb{B}(\mathscr{H})$,

$$E_{f,g}(\alpha S + \beta T) = \alpha E_{f,g}(S) + \beta E_{f,g}(T).$$

(iii) For all $f,g \in \mathscr{K}$ and $S \in \mathbb{B}(\mathscr{H})$,

$$E_{f,g}(S)^* = E_{g,f}(S^*).$$

The following continuity properties of the map $: S \to E_{f,g}(S)$ can also be easily proved.

Proposition 2.6: Let $S, S_n \in \mathbb{B}(\mathscr{H})$, $n \geqslant 1$, and $f,g \in \mathscr{K}$. Then:

(i) If $S_n \to S (n \to \infty)$ in operator norm, then $E_{f,g}(S_n) \to E_{f,g}(S) (n \to \infty)$ in operator norm.

(ii) If $S_n \to S (n \to \infty)$ strongly, then $E_{f,g}(S_n) \to E_{f,g}(S) (n \to \infty)$ strongly.

(iii) If $S_n \to S (n \to \infty)$ weakly, then $E_{f,g}(S_n) \to E_{f,g}(S) (n \to \infty)$ weakly.

Lemma 2.7: Let $P \in \mathbb{B}(\mathscr{K})$ be an orthogonal projection with $\dim \operatorname{Ran} P = n < \infty$. Let $\{f_j\}_{j=1}^n$ be an orthonormal basis of $\operatorname{Ran} P$. Then, for all $S \in \mathbb{B}(\mathscr{H})$,

$$(I \otimes P)S(I \otimes P) = \sum_{j,k=1}^{n} E_{f_j f_k}(S) \otimes P_{kj}, \tag{2.21}$$

where $P_{kj} \in \mathbb{B}(\mathscr{K})$ is defined by

$$P_{kj} f = (f_k, f)_{\mathscr{K}} f_j. \tag{2.22}$$

In particular, if $n = 1$ and $\operatorname{Ran} P = \{\alpha f_0 \mid \alpha \in \mathbb{C}\}$ with $\|f_0\| = 1$, then

$$(I \otimes P)S(I \otimes P) = E_{f_0}(S) \otimes P. \tag{2.23}$$

Proof: Let $u,v \in \mathscr{H}$ and $f,g \in \mathscr{K}$. Then we have

$$\begin{aligned}(u \otimes f, (I \otimes P)S(I \otimes P)v \otimes g)_{\mathscr{H}} &= \sum_{j,k=1}^{n} (f, f_j)_{\mathscr{K}} (f_k, g)_{\mathscr{K}} (u \otimes f_j, S(v \otimes f_k))_{\mathscr{H}} \\ &= \sum_{j,k=1}^{n} (f, P_{kj} g)_{\mathscr{K}} (u, E_{f_j f_k}(S)v)_{\mathscr{H}} \\ &= \left(u \otimes f, \left\{\sum_{j,k=1}^{n} E_{f_j f_k}(S) \otimes P_{kj}\right\} v \otimes g\right)_{\mathscr{H}}.\end{aligned}$$

Thus (2.21) follows. ∎

We next define the partial expectation for unbounded operators. For this purpose, we introduce a class of linear operators in $\mathscr{H}$.

Definition 2.8: We say that a densely defined linear operator S in $\mathscr{H}$ is in $\mathbb{E}(\mathscr{H})$ if and only if there exist subspaces $D_{\mathscr{H}}(S)$ and $D_{\mathscr{K}}(S)$ dense in $\mathscr{H}$ and $\mathscr{K}$, respectively, such that

$$D_{\mathscr{H}}(S) \hat{\otimes} D_{\mathscr{K}}(S) \subset D(S),$$

where $\hat{\otimes}$ denotes algebraic tensor product.

Let $S \in \mathbb{E}(\mathscr{H})$. Then, for all $f \in \mathscr{K}$, $g \in D_{\mathscr{K}}(S)$, and $v \in D_{\mathscr{K}}(S)$, the conjugate linear functional

$$L(u) = (u \otimes f, S(v \otimes g))x, \quad u \in \mathscr{H},$$

on $\mathscr{H}$ is bounded with

$$|L(u)| \leqslant \|f\| \, \|S(v \otimes g)\| \, \|u\|.$$

Therefore, by the Riesz lemma, there exists a unique vector $E_{f,g}(S)v \in \mathscr{H}$ such that

$$L(u) = (u, E_{f,g}(S)v)_{\mathscr{H}}$$

and

$$\|E_{f,g}(S)v\| \leqslant \|f\| \, \|S(v \otimes g)\|.$$

The map $: v \to E_{f,g}(S)v \in \mathscr{H}$ is linear. Hence, $E_{f,g}(S)$ gives a densely defined linear operator in $\mathscr{H}$ with $D(E_{f,g}(S)) = D_{\mathscr{H}}(S)$. We remark that $E_{f,g}(S)$ may depend on the choice of the pair of the subspaces $D_{\mathscr{H}}(S)$ and $D_{\mathscr{K}}(S)$. A criterion for the closability of $E_{f,g}(S)$ is given by the following proposition.

Proposition 2.9: Let $S \in \mathbb{E}(\mathscr{H})$. Suppose that S^* is in

$\mathbb{E}(\mathscr{H})$. Then, for all $f\in D_{\mathscr{H}}(S^*), g\in D_{\mathscr{H}}(S)$, $E_{f,g}(S)$ is closable and

$$E_{g,f}(S^*)\subset E_{f,g}(S)^*.$$

Proof: It is straightforward to see that for all $u\in D_{\mathscr{H}}(S^*)$, $v\in D_{\mathscr{H}}(S)$, $f\in D_{\mathscr{H}}(S^*)$, and $g\in D_{\mathscr{H}}(S)$,

$$(u,E_{f,g}(S)v)_{\mathscr{H}}=(E_{g,f}(S^*)u,v)_{\mathscr{H}},$$

which implies the desired result. ■

Lemma 2.7 is translated into the present case as follows.

Lemma 2.10: Let P and $\{f_j\}_{j=1}^n$ be as in Lemma 2.7. Suppose that $S\in\mathbb{E}(\mathscr{H})$ with $\mathrm{Ran}\,P\subset D_{\mathscr{H}}(S)$. Then, the same conclusion as in Lemma 2.7 holds for S.

The above lemma and (2.10) immediately give the following result.

Proposition 2.11: Suppose that $\dim \mathrm{Ker}\,B=n<\infty$ and C is in $\mathbb{E}(\mathscr{H})$ with $D_{\mathscr{H}}(C)\supset \mathrm{Ker}\,B$. Let $\{f_j\}_{j=1}^n$ be an orthonormal basis of $\mathrm{Ker}\,B$. Then,

$$H_\infty=A\otimes I+\sum_{j,k=1}^{n}E_{f_jf_k}(C)\otimes(P_0)_{kj}.$$

In particular, if $\mathrm{Ker}\,B=\{\alpha f_0\,|\,\alpha\in\mathbb{C}\}$ with $\|f_0\|=1$, then

$$H_\infty=H_{\mathrm{eff}}\otimes P_0+A\otimes(I-P_0),$$

where

$$H_{\mathrm{eff}}=A+E_{f_0}(C). \tag{2.24}$$

The following fact easily follows from Theorems 2.1, 2.2, and (2.24).

Theorem 2.12: Let C be as in Proposition 2.11 and $\mathrm{Ker}\,B=\{\alpha f_0\,|\,\alpha\in\mathbb{C}\}$ with $\|f_0\|=1$. Then: (i) Let $z\in\mathbb{C}$ be as in Theorem 2.1. Then

$$\text{s-}\lim_{\kappa\to\infty}(H_\kappa-z)^{-1}=(H_{\mathrm{eff}}-z)^{-1}\otimes P_0.$$

(ii) For all $F\in C_\infty(\mathbb{R})$,

$$\text{s-}\lim_{\kappa\to\infty}F(H_\kappa)=F(H_{\mathrm{eff}})\otimes P_0.$$

Under the assumption of Theorem 2.12, the self-adjoint operator H_{eff} may be regarded as an "effective" operator, in the asymptotic region $\kappa\approx\infty$, of H_κ restricted to the subspace $\mathscr{H}\otimes\mathrm{Ker}\,B$.

We next consider the relation between the spectrum of H_κ and of H_{eff}.

Theorem 2.13: Under the assumption of Theorem 2.12, we have:

(i) If $a,b\in\mathbb{R}$, $a<b$, and $(a,b)\cap\sigma(H_\kappa)=0$ for all large κ, then $(a,b)\cap\sigma(H_{\mathrm{eff}})=0$.

(ii) Let $\{E_\lambda(H_\kappa)\}$ and $\{E_\lambda(H_{\mathrm{eff}})\}$ be the spectral family of H_κ and of H_{eff}, respectively. Let $a,b\in\mathbb{R}$, $a<b$, and $a,b\notin\sigma_{\mathrm{pp}}(H_{\mathrm{eff}})$, where $\sigma_{\mathrm{pp}}(H_{\mathrm{eff}})$ denotes the pure point spectrum of H_{eff}. Then,

$$\text{s-}\lim_{\kappa\to\infty}E_{(a,b)}(H_\kappa)=E_{(a,b)}(H_{\mathrm{eff}})\otimes P_0.$$

Proof: This follows from Theorem 2.12 and an application of Theorem A.2 in the Appendix. ■

Let A_1 be a symmetric operator in $\mathscr{H}$ such that $A+A_1$ has a discrete spectrum. We may write H_κ as

$$H_\kappa=\tilde{H}_{0,\kappa}+H_I(\kappa),$$

where

$$\tilde{H}_{0,\kappa}=(A+A_1)\otimes I+\kappa I\otimes B$$

and

$$H_I(\kappa)=C_\kappa-A_1\otimes I.$$

If the spectrum of B is of the form $[0,\infty)$, then all the eigenvalues of $\tilde{H}_{0,\kappa}$ are embedded in the continuous spectrum of $\tilde{H}_{0,\kappa}$ and hence H_κ gives an example for perturbation problem of embedded eigenvalues. In general, embedded eigenvalues may be unstable under perturbations, i.e., they may disappear under perturbations (e.g., Refs. 6, 9, and 22); H_κ may have no eigenvalues more than the ground state energy. On the other hand, the effective operator H_{eff} may be regarded as the unperturbed operator of H_κ in the sense of Theorem 2.12 and its eigenvalues may be discrete (see Secs. III and IV). It is well known that one of the concepts to handle such a situation in perturbation problems is *spectral concentration* (e.g., Chap. VIII, Sec. 5 in Ref. 20 and Sec. XII.5 in Ref. 22). We recall the following definition.

Definition: Let T_n be a family of self-adjoint operators and $E_\lambda(T_n)$ be the spectral family of T_n. Let $\{\Lambda_n\}_{n=1}^\infty$ and Λ be subsets of $\mathbb{R}$. We say that *the part of the spectrum of T_n in Λ is asymptotically concentrated on Λ_n as $n\to\infty$* if and only if

$$\text{s-}\lim_{n\to\infty}E_{\Lambda\cap\Lambda_n^c}(T_n)=0,$$

where $\Lambda_n^c=\mathbb{R}-\Lambda_n$.

Theorem 2.14: Let C and $\mathrm{Ker}\,B$ be as in Theorem 2.12. Let $R>0$ and Λ be the union of a finite number of mutually disjoint, bounded open intervals of $\mathbb{R}$ such that $[-R,R]\cap\sigma(H_{\mathrm{eff}})\subset\Lambda$. Then, the part of the spectrum of H_κ in $[-R,R]$ is asymptotically concentrated on Λ as $\kappa\to\infty$.

Proof: We write

$$\Lambda=\cup_{j=1}^n(a_j,b_j).$$

It suffices to consider the case where $\alpha_1<-R<b_1<a_2<b_2\cdots<a_n<R<b_n$. Then we have

$$\Lambda^c(R)\equiv[-R,R]\cap\Lambda^c=\cup_{j=1}^{n-1}[b_j,a_{j+1}].$$

For all $j=1,\ldots,n-1$, the interval $[b_j,a_{j+1}]$ is included in the resolvent set $\rho(H_{\mathrm{eff}})$. Hence, for each $j-1,\ldots,n-1$, there exist constants a_j' and b_j' such that $[b_j,a_{j+1}]\subset(b_j',a_{j+1}')\subset\rho(H_{\mathrm{eff}})$ and $b_j',a_{j+1}'\notin\sigma(H_{\mathrm{eff}})$. Hence, by Theorem 2.13(ii), we have

$$E_{(b_j',a_{j+1}')}(H_\kappa)\to E_{(b_j',a_{j+1}')}(H_{\mathrm{eff}})\otimes P_0=0,$$

strongly as $\kappa\to\infty$. Since

$$\Lambda^c(R)\subset\cup_{j=1}^{n-1}(b_j',a_{j+1}')$$

and hence

$$E_{\Lambda^c(R)}(H_\kappa)\leqslant\sum_{j=1}^{n-1}E_{(b_j',a_{j+1}')}(H_\kappa),$$

we obtain

$$E_{\Lambda^c(R)}(H_\kappa)\to 0,$$

strongly as $\kappa\to\infty$. Thus the desired result follows. ■

Remark: The above result is weaker than the standard result on spectral concentration (e.g., Chap. VIII, Sec. 5,

Theorem 5.1 in Ref. 20). This is due to the fact that the strong resolvent convergence of H_κ is different from the usual strong resolvent convergence where the limiting operator is also a resolvent. Theorem 2.14 may be interpreted as a *local* spectral concentration of H_κ on the spectrum of H_{eff}.

III. THE PAULI–FIERZ MODEL

In this section we apply the abstract theory in the last section to the Pauli–Fierz model to study its nonrelativistic limit.

A. Definition of the model and some fundamental facts

The model describes a quantum system of a one-electron atom coupled to a quantized radiation field[1-4] (cf. also Refs. 5–9). For a mathematical generality, we assume that the one-electron atom is placed in the d-dimensional space $\mathbb{R}^d$ ($d \geqslant 2$). We shall denote by $\hbar$ (resp. m, c) the Planck constant divided by 2π (resp. the electronic mass, the speed of light), regarding them as positive parameters. In what follows, the differential operators $\partial/\partial x_j$, $j = 1,\ldots, d$, $x = (x_1,\ldots,x_d) \in \mathbb{R}^d$, are taken in the generalized sense. We set

$$p = \left(-i\hbar \frac{\partial}{\partial x_1}, \ldots, -i\hbar \frac{\partial}{\partial x_d} \right). \tag{3.1}$$

We take the potential $V(x)$ of the atom to be a real-valued measurable function on $\mathbb{R}^d$ which satisfies:

(V-1) $D(p^2) \subset D(V)$ and for all $\lambda > 0$, $V(p^2 + \lambda)^{-1}$ is bounded with

$$\lim_{\lambda \to \infty} \| V(p^2 + \lambda)^{-1} \| = 0.$$

(V-2) For all $t > 0$ and $x \in \mathbb{R}^d$,

$$\int_{\mathbb{R}^d} e^{-t|x-y|^2} |V(y)| dy < \infty.$$

Condition (V-1) implies that V is infinitesimally small with respect to p^2 and hence the Hamiltonian of the atom

$$H_A = (1/2m)p^2 + V \tag{3.2}$$

is self-adjoint on $D(p^2)$ and bounded from below.

Remark: If V is a Phillips perturbation of p^2, then V satisfies (V-1) (see Refs. 23 and 24). It was proved in Ref. 23 that if

$$V \in L^q(\mathbb{R}^d) + L^\infty(\mathbb{R}^d),$$

with $q > d/2$ and $q \geqslant 2$, then V is a Phillips perturbation of p^2. In particular, it follows that the Coulomb potential in the case $d = 3$ satisfies (V-1) and (V-2).

We use the Coulomb gauge in quantizing the radiation field. The Hilbert space of state vectors for the quantized radiation field is then defined by the boson Fock space:

$$\mathscr{F}_{\text{EM}} = \oplus_{n=0}^{\infty} \otimes_s^n W \tag{3.3}$$

over the Hilbert space

$$W = \underbrace{L^2(\mathbb{R}^d) \oplus \cdots \oplus L^2(\mathbb{R}^d)}_{d-1 \text{ times}}, \tag{3.4}$$

where $\otimes_s^n W$ denotes the n-fold symmetric tensor product of W with convention $\otimes_s^{n=0} W = \mathbb{C}$. We denote by $a(F)$, $F \in W$, the annihilation operator in $\mathscr{F}_{\text{EM}}$. For $r = 1,\ldots, d-1$ and $f \in L^2(\mathbb{R}^d)$, we define $f_r \in W$ by $f_r = (0,\ldots,f,\ldots,0)$ (the rth component is equal to f and the other components are zero). The map $f \to a(f_r)$ defines an operator-valued distribution on $\mathbb{R}^d$. We denote the distribution kernel by $a_r(k)$, $r = 1,\ldots, d-1$, $k \in \mathbb{R}^d$. Then the following canonical commutation relations hold in the sense of operator-valued distribution:

$$[a_r(k), a_q(k')] = [a_r(k)^*, a_q(k')^*] = 0,$$

$$[a_r(k), a_q(k')^*] = \delta_{rq}\delta(k - k'), \quad r,q = 1,\ldots,d-1.$$

Let $e_r(k)$ be an $\mathbb{R}^d$-valued measurable function on $\mathbb{R}^d$ such that

$$k \cdot e_r(k) = 0, \quad e_r(k) \cdot e_q(k) = \delta_{rq},$$

$$a.e.\, k \in \mathbb{R}^d, \quad r,q = 1,\ldots,d-1.$$

The vectors $e_r(k)$, $r = 1,\ldots,d-1$, serve as polarization vectors of "photon."

The free Hamiltonian of the quantized radiation field is defined by

$$H_F = \hbar c \sum_{r=1}^{d-1} \int dk\, \omega(k) a_r(k)^* a_r(k). \tag{3.5}$$

Here, $\omega(k)$ is a non-negative measurable function on $\mathbb{R}^d$ with $\omega \in L^2_{\text{loc}}(\mathbb{R}^d)$ which depends only on $|k|$. The physical choice for $\omega(k)$ is given by $\omega(k) = |k|$.

The Hilbert space $\mathscr{F}$ of state vectors for the interacting system of the atom and the radiation field is taken to be the tensor product of $L^2(\mathbb{R}^d)$ and $\mathscr{F}_{\text{EM}}$:

$$\mathscr{F} = L^2(\mathbb{R}^d) \otimes \mathscr{F}_{\text{EM}}. \tag{3.6}$$

To define the interaction between the atom and the radiation field as an operator in $\mathscr{F}$, we have to introduce a cutoff for photon momenta: Let $\rho(x)$ be a real distribution on $\mathbb{R}^d$ such that its Fourier transform

$$\hat{\rho}(k) = \frac{1}{(2\pi)^{d/2}} \int dx\, \rho(x) e^{-ikx} \tag{3.7}$$

is a measurable function and depends only on $|k|$ with

$$\int dk \frac{|\hat{\rho}(k)|^2}{\omega(k)^3} < \infty, \quad \int dk \frac{|\hat{\rho}(k)|^2}{\omega(k)} < \infty. \tag{3.8}$$

Then we define the time-zero radiation field with cutoff ρ by

$$A(x;\rho) = \sum_{r=1}^{d-1} \int dk \frac{\sqrt{\hbar c}}{\sqrt{2\omega(k)}} e_r(k) \times \{\hat{\rho}(k)^* a_r(k)^* e^{-ikx} + \hat{\rho}(k) a_r(k) e^{ikx}\}. \tag{3.9}$$

The total Hamiltonian of the coupled system of the atom and the radiation field with the full minimal interaction reads:

$$H = (1/2m)(p - (e/c)A(x;\rho))^2 + H_F + V, \tag{3.10}$$

where $e \in \mathbb{R} \setminus \{0\}$ is a coupling parameter denoting the elementary charge. In the present paper, however, we take as the total Hamiltonian of the coupled system a version of H simplified in the following way: (i) We use the dipole approximation, i.e., we replace $A(x;\rho)$ by $A(0;\rho)$; (ii) We neglect the term $A(x;\rho)^2$.

Further, we take the mass renormalization of the electron into account, i.e., we introduce the "bare mass" m_0 of

the electron by

$$\frac{1}{m_0}=\frac{1}{m}+\frac{(d-1)}{d}\left(\frac{e}{mc}\right)^2\int dk\,\frac{|\hat{\rho}(k)|^2}{\omega(k)^2}, \tag{3.11}$$

and define the "renormalized" atom Hamiltonian $H_A^{\rm ren}$ by

$$H_A^{\rm ren}=(1/2m_0)p^2+V. \tag{3.12}$$

Thus the total Hamiltonian of our model is defined by

$$H(c,e)=H_A^{\rm ren}\otimes I+I\otimes H_F+H_I, \tag{3.13}$$

where

$$H_I=-(e/mc)p\otimes A(0;\rho).$$

For $\kappa>0$, we introduce

$$c(\kappa)=\kappa c,\quad e(\kappa)=\kappa^{3/2}e,$$

which are regarded as a scaled speed of light and a scaled elementary charge, respectively. Then we define the scaled Hamiltonian $H(\kappa)$ by

$$H(\kappa)\equiv H(c(\kappa),e(\kappa))$$

$$=\left(\frac{1}{2m(\kappa)}p^2+V\right)\otimes I+\kappa I\otimes H_F+\kappa H_I, \tag{3.14}$$

where $m(\kappa)$ is defined by

$$\frac{1}{m(\kappa)}=\frac{1}{m}+\kappa\frac{(d-1)}{d}\left(\frac{e}{mc}\right)^2\int dk\,\frac{|\hat{\rho}(k)|^2}{\omega(k)^2}.$$

We want to consider the scaling limit $\kappa\to\infty$ of the model in terms of $H(\kappa)$. Obviously $c(\kappa),|e(\kappa)|\to\infty$ as $\kappa\to\infty$. In this sense, the scaling limit $\kappa\to\infty$ in $H(\kappa)$ corresponds to the nonrelativistic limit at the same time as the magnitude of the coupling charge becomes infinite. Note also that the "scaled bare mass" $m(\kappa)\to 0$ as $\kappa\to\infty$.

Before stating the main results on the scaling limit, we give some known facts. We denote by Ω the Fock vacuum in $\mathscr{F}_{\rm EM}$:

$$\Omega=\{1,0,0,\ldots\}. \tag{3.15}$$

Let $\mathscr{F}_{\rm EM,0}$ be the dense subspace in $\mathscr{F}_{\rm EM}$ spanned by vectors of the form

$$a(F_1)^*\cdots a(F_n)^*\Omega,\ \Omega,\quad F_j\in W,\ j=1,\ldots,n,\ n\geqslant 1$$

and

$$\mathscr{S}_0(\mathbb{R}^d)=\{f\in\mathscr{S}(\mathbb{R}^d)\,|\,\hat{f}\in C_0^\infty(\mathbb{R}^d)\}, \tag{3.16}$$

where $\mathscr{S}(\mathbb{R}^d)$ is the Schwartz test function space of rapidly decreasing C^∞-functions on $\mathbb{R}^d$ and $C_0^\infty(\mathbb{R}^d)$ denotes the space of C^∞-functions on $\mathbb{R}^d$ with compact support. Then the subspace

$$\mathscr{F}_0=\mathscr{S}_0(\mathbb{R}^d)\,\hat{\otimes}\,\mathscr{F}_{\rm EM,0} \tag{3.17}$$

is dense in $\mathscr{F}$.

Let

$$T=i\frac{e}{mc}\sum_{r=1}^{d-1}\int dk\,\frac{1}{\omega(k)\sqrt{2\hbar c\omega(k)}}p$$
$$\cdot e_r(k)\{\hat{\rho}(k)^*a_r(k)^*-\hat{\rho}(k)a_r(k)\}. \tag{3.18}$$

Then we can show that $\mathscr{F}_0$ is a set of analytic vectors of T and hence T is essentially self-adjoint on it.[5] We denote the unique self-adjoint extension of $T\restriction\mathscr{F}_0$ by the same symbol.

Let

$$C(V)=e^{iT}(V\otimes I)e^{-iT}. \tag{3.19}$$

Since $\exp(-tp^2\otimes I)$ commutes with $\exp(i\lambda T)\,(\lambda\in\mathbb{R})$ and $\exp(i\lambda T)$ is unitary, it follows from (V-1) that $C(V)$ is infinitesimally small with respect to $p^2\otimes I$ with

$$\lim_{\lambda\to\infty}\|C(V)(p^2\otimes I+\lambda)^{-1}\|=0, \tag{3.20}$$

which implies that $C(V)$ is infinitesimally small with respect to $(p^2\otimes I)/2m+\kappa I\otimes H_F$ with

$$\lim_{\lambda\to\infty}\|C(V)[(1/2m)p^2\otimes I+\kappa I\otimes H_F+\lambda]^{-1}\|=0 \tag{3.21}$$

uniformly in κ. Therefore, the operator

$$\tilde{H}(\kappa)=(1/2m)p^2\otimes I+\kappa I\otimes H_F+C(V) \tag{3.22}$$

is self-adjoint on

$$D_0=D(p^2\otimes I)\cap D(I\otimes H_F) \tag{3.23}$$

and bounded from below. We have

$$\tilde{H}(\kappa)\geqslant\inf\sigma(H_A). \tag{3.24}$$

This follows from the non-negativity of H_F and the fact that $p^2\otimes I$ commutes with $\exp(\pm iT)$.

A fundamental fact concerning our model $H(\kappa)$ is the following lemma.

Lemma 3.1: The unitary operator e^{iT} maps D_0 onto D_0 and for all $\kappa>0$,

$$e^{iT}H(\kappa)e^{-iT}=\tilde{H}(\kappa)$$

on D_0. In particular, $H(\kappa)$ is self-adjoint on D_0 and bounded from below with

$$H(\kappa)\geqslant\inf\sigma(H_A).$$

Proof: See Ref. 5 (cf. also Ref. 4). ∎

Proposition 3.2: Let

$$E(\kappa)=\inf\sigma(H(\kappa)). \tag{3.25}$$

Then, $E(\kappa)$ is nondecreasing in κ.

Proof: Since H_F is non-negative, we have from (3.22)

$$\tilde{H}(\kappa)\geqslant\tilde{H}(\kappa'),$$

for all $\kappa>\kappa'>0$. By Lemma 3.1, we have

$$E(\kappa)=\inf\sigma(\tilde{H}(\kappa)).$$

Hence, $E(\kappa)\geqslant E(\kappa')$, for $\kappa>\kappa'>0$. ∎

B. Convergence of the Hamiltonian, effective potential, and an estimate of the ground state energy

Lemma 3.3: The operator $C(V)$ is in $\mathbb{E}(\mathscr{F})$ (see Definition 2.8) with $D_{L^2(\mathbb{R}^d)}(C(V))=D(p^2)$ and $D_{\mathscr{F}_{\rm EM}}(C(V))=D(H_F)$. Further, the partial expectation $E_\Omega(C(V))$ of $C(V)$ with respect to Ω is given by

$$E_\Omega(C(V))=V_{\rm eff}\ \text{ on }\ D(p^2), \tag{3.26}$$

where $V_{\rm eff}$ is the multiplication operator associaed with the function

$$V_{\rm eff}(x)=(2\pi C(\rho))^{-d/2}\int dy\,e^{-|x-y|^2/2C(\rho)}V(y), \tag{3.27}$$

with

$$C(\rho)=\frac{(d-1)}{2d}\left(\frac{\hbar}{mc}\right)^2\frac{e^2}{\hbar c}\int dk\,\frac{|\hat{\rho}(k)|^2}{\omega(k)^3}.$$

Proof: By Lemma 3.1 and condition (V-1), $D(p^2)\hat{\otimes}D(H_F)\subset D_0\subset D(C(V))$. Hence, $C(V)$ is in $\mathbb{E}(\mathscr{F})$. To prove (3.26), we first consider the case where $V\in\mathscr{S}(\mathbb{R}^d)$. Then it follows that for all $f,g\in L^2(\mathbb{R}^d)$,

$$(f,E_\Omega(C(V))g)_{L^2}=\frac{1}{(2\pi)^{d/2}}\int d\xi\,\hat{V}(\xi)(f\otimes\Omega,e^{i\xi x(T)}g\otimes\Omega)_{\mathscr{F}},$$

where

$$x(T)=e^{iT}x\otimes Ie^{-iT}. \tag{3.28}$$

Let

$$X=i\frac{e}{mc}\sqrt{\frac{\hbar}{2c}}\times\sum_{r=1}^{d-1}\int\frac{e_r(k)}{\omega(k)^{3/2}}\{\hat{\rho}(k)^*a_r(k)^*-\hat{\rho}(k)a_r(k)\}.$$

Then, it is not so difficult to see that

$$e^{iT}D(x_\mu\otimes I)=D(x_\mu\otimes I+I\otimes X_\mu),$$

$$x(T)_\mu=x_\mu\otimes I+I\otimes X_\mu,\quad \mu=1,\ldots,d.$$

Hence, we have

$$(f\otimes\Omega,e^{i\xi x(T)}g\otimes\Omega)_{\mathscr{F}}=(f,e^{i\xi x}g)_{L^2}(\Omega,e^{i\xi X}\Omega)_{\mathscr{F}_{\mathrm{EM}}}.$$

By the standard Fock space calculus, we find

$$(\Omega,e^{i\xi X}\Omega)_{\mathscr{F}_{\mathrm{EM}}}=e^{-|\xi|^2C(\rho)/2}.$$

Thus (3.26) follows.

We next consider the case where V is bounded, but, not in $\mathscr{S}(\mathbb{R}^d)$. In this case, we approximate V by a sequence $\{V_n\}_n\subset\mathscr{S}(\mathbb{R}^d)$ in the sense of strong convergence in $L^2(\mathbb{R}^d)$. Then, by Proposition 2.6 (ii), we have

$$E_\Omega(C(V_n))\to E_\Omega(C(V))\,(n\to\infty)$$

strongly. On the other hand, we have

$$(V_n)_{\mathrm{eff}}(x)\to V_{\mathrm{eff}}(x)\,(n\to\infty),$$

for all $x\in\mathbb{R}^d$. Thus we obtain (3.26).

Finally, let V satisfy (V-1) and (V-2). Denoting by χ_n the characteristic function of $[0,n]$, $n\in\mathbb{N}$, we define

$$V_n(x)=\chi_n(|x|)V(x).$$

Then V_n is bounded and hence (3.26) holds with V replaced by V_n. It is easy to see that for all $\Psi\in D_0$,

$$C(V_n)\Psi\to C(V)\Psi$$

strongly. Hence, for all $f\in C_0^\infty(\mathbb{R}^d)$ and g in $D(p^2)$,

$$(f,E_\Omega(C(V_n))g)\to(f,E_\Omega(C(V))g). \tag{3.29}$$

It follows from condition (V-2) that $|V|_{\mathrm{eff}}$ is a continuous function on $\mathbb{R}^d$. Hence, by using the dominated convergence theorem, we have

$$(f,(V_n)_{\mathrm{eff}}g)\to(f,V_{\mathrm{eff}}g),$$

which, combined with (3.29), gives

$$(f,E_\Omega(C(V))g)=(f,V_{\mathrm{eff}}g).$$

Thus (3.26) follows. ∎

By (3.26) and the fact that $C(V)$ is infinitesimally small with respect to $p^2\otimes I$, it follows that V_{eff} is infinitesimally small with respect to $p^2/2m$. Hence,

$$H_{A,\mathrm{eff}}=(1/2m)p^2+V_{\mathrm{eff}} \tag{3.30}$$

is self-adjoint on $D(p^2)$ and bounded from below. Let

$$E_{A,\mathrm{eff}}=\inf\sigma(H_{A,\mathrm{eff}}). \tag{3.31}$$

We denote by P_0 the orthogonal projection from $\mathscr{F}_{\mathrm{EM}}$ onto the Fock vacuum sector $\{\alpha\Omega|\alpha\in\mathbb{C}\}$.

Theorem 3.4: For all $z\in\mathbb{C}$ with $\mathrm{Im}\,z\neq0$ or $z\in\mathbb{R}$ with $z<\min\{\inf\sigma(H_A),\inf\sigma(H_{A,\mathrm{eff}})\}$,

$$\text{s-}\lim_{\kappa\to\infty}(H(\kappa)-z)^{-1}=e^{-iT}\{(H_{A,\mathrm{eff}}-z)^{-1}\otimes P_0\}e^{iT}. \tag{3.32}$$

Proof: By Lemma 3.1, we need only to consider the scaling limit $\kappa\to\infty$ of $(\tilde{H}(\kappa)-z)^{-1}$. Note that $\tilde{H}(\kappa)$ is just of the form of the operator H_κ considered in Theorem 2.1 with the following identifications:

$$\mathscr{H}=L^2(\mathbb{R}^d),\quad \mathscr{K}=\mathscr{F}_{\mathrm{EM}},$$

$$A=p^2/2m,\quad B=H_F,\quad C_\kappa=C=C(V). \tag{3.33}$$

We have $\mathrm{Ker}\,H_F=\{\alpha\Omega|\alpha\in\mathbb{C}\}$. Hence we obtain from Theorem 2.12(i)

$$\text{s-}\lim_{\kappa\to\infty}(\tilde{H}(\kappa)-z)^{-1}=(p^2/2m+E_\Omega(C(V))-z)^{-1}\otimes P_0,$$

which, together with (3.26), gives (3.32). ∎

Remarks: (i) In the case $d=3$, V_{eff} coincides with the effective potential given by Welton,[3] who derived it by physical arguments to calculate some observable effects of the quantized radiation field such as the Lamb shift. Theorem 3.4 shows that the effective potential can be derived as a scaling limit of the total Hamiltonian $H(\kappa)$. This does not only justify rigorously the effective potential but also clarifies a mathematical meaning of it.

(ii) As (3.27) shows, the effective potential V_{eff} is a Gaussian transformation of the original potential V. The functional $C(\rho)$ of ρ, which characterizes the Gaussian transformation, has a mathematical meaning: let $x(T)$ be defined by (3.28). Then it follows from the proof of Lemma 3.3 that

$$E_\Omega[(x(T)-x\otimes I)^2]=C(\rho)I. \tag{3.34}$$

Hence, $C(\rho)$ can be identified with the partial expectation of the square of $\Delta x\equiv x(T)-x\otimes I$ with respect to the Fock vacuum Ω. In Ref. 3, Δx was regarded as a fluctuation in position of a free electron and $C(\rho)$ was interpreted as the mean-square fluctuation in position of a free electron. In this sense, (3.34) suggests that mean fluctuations of observables under the action of a quantized radiation field may be formulated in terms of the notion of partial expectation.

Theorem 3.5: Suppose that V is bounded and

$$\inf_{\kappa\in\mathbb{R}^d}\omega(k)\geqslant\omega_0, \tag{3.35}$$

with a constant $\omega_0>0$. Then, for all $\kappa>(E_{A,\mathrm{eff}}+\|V\|)/\hbar c\omega_0$,

$$E_{A,\mathrm{eff}}-\|V\|\nu_\kappa(1+\sqrt{1+\nu_\kappa^2})^{-1}\leqslant E(\kappa)\leqslant E_{A,\mathrm{eff}},$$

where

$$\nu_\kappa = 2\|V\|/(\hbar c\omega_0\kappa - E_{A,\text{eff}} - \|V\|).$$

In particular,

$$\lim_{\kappa\to\infty} E(\kappa) = E_{A,\text{eff}}.$$

Proof: Since V is bounded in the present case, we have

$$\|C(V)\| = \|V\|.$$

Under condition (3.35), $H_F \upharpoonright (\text{Ker } H_F)^\perp \geqslant \hbar c\omega_0$. Hence, an application of Theorem 2.4 to the present case with the identifications (3.33) and $b = \hbar c\omega_0$ yields the desired result. ■

C. Spectral concentration

In the physical case $\omega(k) = |k|$, all the eigenvalues of the unperturbed Hamiltonian $H_A \otimes I + \kappa I \otimes H_F$ are embedded in its continuous spectrum. Hence, as already remarked in Sec. II, they may be unstable under the perturbation $\kappa H_I +$ (the self-energy term), i.e., they may disappear under the perturbation (for an example, see Ref. 6). On the other hand, eigenvalues of $H_{A,\text{eff}}$ may be discrete. The concept of spectral concentration may be helpful to handle such a situation in perturbation problems. We have the following result on the spectral concentration of $H(\kappa)$.

Theorem 3.6: Let $R > 0$ and Λ be the union of a finite number of mutually disjoint, bounded open intervals of $\mathbb{R}$ such that $[-R,R] \cap \sigma(H_{A,\text{eff}}) \subset \Lambda$. Then, the part of the spectrum of $H(\kappa)$ in $[-R,R]$ is asymptotically concentrated on Λ as $\kappa \to \infty$.

Proof: Let $E_\lambda(H(\kappa))$ be the spectral family of $H(\kappa)$. Then, by Lemma 3.1,

$$\exp(iT)E_\lambda(H(\kappa))\exp(-iT) \equiv E_\lambda(\tilde{H}(\kappa))$$

is the spectral family of $\tilde{H}(\kappa)$ By Theorems 3.4 and 2.14, we have

$$E_{\Lambda^c(R)}(\tilde{H}(\kappa)) \to 0$$

strongly as $\kappa \to \infty$. Hence,

$$E_{\Lambda^c(R)}(H(\kappa)) \to 0$$

strongly as $\kappa \to \infty$. ■

IV. THE SPIN-BOSON MODEL

The spin-boson model we are going to discuss describes a two-level atom coupled to a quantized Bose field (a simplified version of a quantized radiation field) (e.g., Ref. 13 for a review and Refs. 10–12, 14, and 15 for some rigorous results). We denote by $\mu > 0$ the half of the gap of the two energy levels of the unperturbed atom. The total Hamiltonian of the model is given as follows:

$$\begin{aligned} H &= H(c,\lambda) \\ &= I \otimes H_b + \mu\sigma_1 \otimes I + \sigma_3 \\ &\quad \otimes \int dk\{\lambda(k)^* a(k)^* + \lambda(k)a(k)\} - E_0(c,\lambda). \end{aligned} \tag{4.1}$$

Here, the Hilbert space in which H acts is

$$\mathcal{F} = \mathbb{C}^2 \otimes \mathcal{F}_B(L^2(\mathbb{R}^d)) = \mathcal{F}_B(L^2(\mathbb{R}^d)) \oplus \mathcal{F}_B(L^2(\mathbb{R}^d)), \tag{4.2}$$

where $\mathcal{F}_B(L^2(\mathbb{R}^d))$ denotes the boson Fock space over $L^2(\mathbb{R}^d)$ $(d \geqslant 1)$. The 2×2 matrices σ_1 and σ_3 are the standard Pauli matrices and $a(k)$ is the operator-valued distribution kernel of the boson annihilation operator acting in $\mathcal{F}_B(L^2(\mathbb{R}^d))$. The operator H_b is the free boson Hamiltonian:

$$H_b = c\int dk\, \omega(k)a(k)^*a(k), \tag{4.3}$$

where $\omega(k)$ is a non-negative measurable function on $\mathbb{R}^d$ with $\omega \in L^2_{\text{loc}}(\mathbb{R}^d)$. We assume that $\lambda(k)$ is a measurable function on $\mathbb{R}^d$ satisfying the following conditions:

$$\int dk\, |\lambda(k)|^2 < \infty, \quad \int dk\, \frac{|\lambda(k)|^2}{\omega(k)^2} < \infty. \tag{4.4}$$

The functional $E_0(c,\lambda): \lambda \to \mathbb{C}$ is defined by

$$E_0(c,\lambda) = -\frac{1}{c}\int dk\, \frac{|\lambda(k)|^2}{\omega(k)}, \tag{4.5}$$

which physically means the ground state energy of the bosonic Hamiltonian

$$\begin{aligned} H_b(c,\lambda) &= I \otimes H_b + \sigma_3 \\ &\quad \otimes \int dk\{\lambda(k)^* a(k)^* + \lambda(k)a(k)\}. \end{aligned}$$

We have set $\hbar = 1$.

It is known (or easy to see by applying the Kato–Rellich theorem) that H is self-adjoint with $D(H) = D(I \otimes H_b)$ and bounded from below.

For $\kappa > 0$, we set

$$H(s) = H(\kappa c, \kappa\lambda), \tag{4.6}$$

which is the scaled Hamiltonian we are going to study in the asymptotic region $\kappa \approx \infty$. The operators

$$T_\pm = \pm i\frac{1}{c}\int dk\, \frac{1}{\omega(k)}\{\lambda(k)^* a(k)^* - \lambda(k)a(k)\} \tag{4.7}$$

are self-adjoint in $\mathcal{F}_B$. Hence, we can define the unitary operators

$$U_\pm = e^{iT_\pm} \tag{4.8}$$

on $\mathcal{F}_B$. Set

$$U = \begin{pmatrix} U_+ & 0 \\ 0 & U_- \end{pmatrix}, \tag{4.9}$$

which is unitary on $\mathcal{F}$. The following fact is easily proved.

Lemma 4.1: For all $\kappa > 0$,

$$\tilde{H}(\kappa) \equiv U^{-1}H(\kappa)U = \kappa I \otimes H_b + W, \tag{4.10}$$

where

$$W = \mu\begin{pmatrix} 0 & U_-^2 \\ U_+^2 & 0 \end{pmatrix}. \tag{4.11}$$

Let

$$E(\kappa) = \inf \sigma(H(\kappa)). \tag{4.12}$$

Proposition 4.2: The ground state energy $E(\kappa)$ is nondecreasing in $\kappa > 0$ and

$$\inf_{\kappa>0} E(\kappa) \geqslant -\mu. \tag{4.13}$$

Proof: By (4.10), we have

$$\tilde{H}(\kappa) \geqslant \tilde{H}(\kappa'),$$

for all $\kappa > \kappa' > 0$. Since we have

$$E(\kappa) = \inf \sigma(\tilde{H}(\kappa)),$$

the nondecreasingness of $E(\kappa)$ follows. Note that $W \geqslant -\mu$, which, combined with (4.10) and the non-negativity of H_b, gives (4.13). ∎

Lemma 4.1 shows that $H(\kappa)$ is unitarily equivalent to an operator of the form of H_κ discussed in Sec. II. Hence, we can apply the main theorems in Sec. II to the present case. Let us compute the partial expectation of W with respect to the vectors in Ker H_b first. We have Ker $H_b = \{\alpha\Omega | \alpha \in \mathbb{C}\}$, where Ω is the Fock vacuum in $\mathscr{F}_B(L^2(\mathbb{R}^d))$.

Lemma 4.3: Let

$$F(c,\lambda) = \exp\left(-\frac{2}{c^2} \int dk \frac{|\lambda(k)|^2}{\omega(k)^2} \right). \tag{4.14}$$

Then,

$$E_\Omega(W) = \mu F(c,\lambda)\sigma_1. \tag{4.15}$$

Proof: By computing the inner product $(u \otimes \Omega, W(v \otimes \Omega))$ for

$$u = (z_1, z_2), v = (w_1, w_2) \in \mathbb{C}^2,$$

we see that

$$E_\Omega(W) = \mu \begin{pmatrix} 0 & (\Omega, U_-^2 \Omega) \\ (\Omega, U_+^2 \Omega) & 0 \end{pmatrix}.$$

It is straightforward to show that

$$(\Omega, U_\pm^2 \Omega) = F(c,\lambda).$$

Thus (4.15) follows. ∎

Theorem 4.4: For all $z \in \mathbb{C} \setminus [-\mu, \infty)$,

$$\text{s-}\lim_{\kappa\to\infty} (H(\kappa) - z)^{-1} = U(\mu F(c,\lambda)\sigma_1 - z)^{-1} \otimes P_0 U^{-1}. \tag{4.16}$$

Proof: We need only to use (4.10) and apply Theorem 2.12 with the following identifications:

$$\mathscr{H} = \mathbb{C}^2, \quad \mathscr{K} = \mathscr{F}_B(L^2(\mathbb{R}^d)),$$
$$A = 0, \quad B = H_b, \quad C_\kappa = C = W. \tag{4.17}$$

[Note that $\mu F(c,\lambda)\sigma_1 \geqslant -\mu F(c,\lambda) \geqslant -\mu$.] ∎

The estimate (4.13) of the ground state energy $E(\kappa)$ is improved as follows.

Theorem 4.5: Suppose that $\omega(k)$ satisfies (3.35). Then, for all $\kappa > \mu(1 - F(c,\lambda))/c\omega_0$, we have

$$-\mu F(c,\lambda) - \mu d_\kappa (1 + \sqrt{1 + d_\kappa^2})^{-1} \leqslant E(\kappa) \leqslant -\mu F(c,\lambda), \tag{4.18}$$

where

$$d_\kappa = 2\mu / [c\omega_0 \kappa - \mu(1 - F(c,\lambda))].$$

Proof: Under condition (3.35), $H_b \upharpoonright (\text{Ker } H_b)^\perp \geqslant c\omega_0$. Further, we have

$$\|W\| = \mu, \quad \inf \sigma(\mu F(c,\lambda)\sigma_1) = -\mu F(c,\lambda).$$

Hence, applying Theorem 2.4 with the identifications (4.17), we obtain (4.18). ∎

Remarks: (i) The ground state of the model $H(\kappa)$ with $\mu = 0$ is twofold degenerate. We note that $F(c,\lambda)^2$ is equal to the transition probability between the two ground states at zero temperature and in the case $\kappa = 1$ (cf. Ref. 11).

(ii) Inequality (4.18) gives a nonperturbative estimate of the ground state energy of $H(\kappa)$ with respect to both parameters μ and λ and slightly improves the estimate given by Davies.[10]

As for the spectral concentration of $H(\kappa)$, we have the following result.

Theorem 4.6: Let $R > 0$ and $\epsilon > 0$. Then, the part of the spectrum of $H(\kappa)$ in $[-R,R]$ is asymptotically concentrated on

$$(-\mu F(c,\lambda) - \epsilon, -\mu F(c,\lambda) + \epsilon) \cup (\mu F(c,\lambda) - \epsilon, \mu F(c,\lambda) + \epsilon)$$

as $\kappa \to \infty$.

Proof: We first note that the spectrum of the effective operator $\mu F(c,\lambda)\sigma_1$ is equal to $\{\pm \mu F(c,\lambda)\}$. Then, in the same way as in the proof of Theorem 3.6, we obtain the desired result. ∎

V. CONCLUDING REMARKS

In this paper we have developed an abstract asymptotic theory of a family of self-adjoint operators, which allows us to study in a unified way the nonrelativistic limit of the Pauli–Fierz and a spin-boson model. We have obtained some new rigorous results for the models, including an asymptotic estimate of their ground state energy and the existence of "local" spectral concentration.

Our method can be applied to other quantum field models whose Hamiltonians can be transformed by unitary transformations ("dressing transformations") to operators of the form of H_κ discussed in Sec. II.

In the present paper we have considered only the case where the quantum system under consideration is at zero temperature. In the case of finite temperature, we have to reformulate the asymptotic theory in Sec. II in terms of correlation functions of a KMS state associated with H_κ.

The following topics may be worth being studied as a continuation of the present work.

(i) Extension of the abstract theory in Sec. II to the case where the operator C_κ is more singular and/or the scaling order of C_κ in κ is different.

(ii) The nonrelativistic limit of the Pauli–Fierz model (3.13) without the dipole approximation. In this case, a formal perturbation calculation suggests that we should have an effective potential different from V_{eff} given by (3.27) (cf. Ref. 3).

(iii) The nonrelativistic limit of the model whose Hamiltonian is given by (3.10). In the case of the dipole approximation, we may use the results in Ref. 7.

APPENDIX: SOME LIMIT THEOREMS

In Sec. II we encounter a strong convergence of resolvent that is different from the usual strong resolvent convergence in that the limiting operator is not the resolvent of an operator. In this Appendix we present some limit theorems related to such a strong resolvent convergence, which are variants of the standard limit theorems of the usual strong resolvent convergence (e.g., Refs. 20 and 21). We first consider a general case.

Theorem A.1: Let T_n, $n\in\mathbb{N}$, and T be self-adjoint operators in a Hilbert space $\mathscr{H}$, and Q be an orthogonal projection on $\mathscr{H}$. Suppose that Q commutes with the resolvent of T and for all $z\in\mathbb{C}\setminus\mathbb{R}$,

$$\text{s-}\lim_{s\to\infty}(T_n - z)^{-1} = (T-z)^{-1}Q. \tag{A1}$$

Then, for all $F\in C_\infty(\mathbb{R})$ (the space of continuous functions on $\mathbb{R}$ vanishing at ∞), $F(T)$ commutes with Q and

$$\text{s-}\lim_{n\to\infty} F(T_n) = F(T)Q. \tag{A2}$$

Further, if T_n is bounded from below uniformly in n, then T is bounded from below and

$$\overline{\lim_{n\to\infty}} E_n \leqslant E, \tag{A3}$$

where

$$E_n = \inf\sigma(T_n), \quad E = \inf\sigma(T\restriction \mathrm{Ran}\, Q).$$

Remark: The commutativity of the resolvent of T with Q implies that T is reduced by Ran Q, so that $T\restriction$ Ran Q is also self-adjoint.

Proof: The proof of (A2) can be done in the same way as in the proof of Theorem VIII.20(a) in Ref. 21. The point that we are careful about in the present case is that the limiting operator $(T-z)^{-1}Q$ is not a resolvent. We can show that for all polynomials $P = P(x,y)$ in two variables x, y:

$$P((T_n+i)^{-1},(T_n-i)^{-1}) \to P((T+i)^{-1},(T-i)^{-1})Q$$

strongly as $n\to\infty$, where we have used the assumption that Q is an orthogonal projection commuting with the resolvent of T. We then see that (A2) follows from a slight modification of the proof of Theorem VIII.20(a) in Ref. 21.

If T_n is bounded from below uniformly in n, then a standard method shows that T is bounded from below and (A1) holds for $z<z_0\equiv\min\{\inf_n E_n, E\}$. We fix a real number $\lambda<z_0$. Then, a standard formula on the strong convergence of bounded linear operators gives

$$\|(T-\lambda)^{-1}Q\| \leqslant \varliminf_{n\to\infty}\|(T_n-\lambda)^{-1}\|.$$

Note that

$$\|(T-\lambda)^{-1}Q\| = \|(T\restriction \mathrm{Ran}\, Q-\lambda)^{-1}\|.$$

Thus (A3) follows. ∎

We next consider operators in the Hilbert space $\mathscr{K}$ given by (2.1).

Theorem A.2: Let T_n, $n\in\mathbb{N}$, be self-adjoint operators in $\mathscr{K}$ and S be a self-adjoint operator in $\mathscr{H}$. Suppose that for all $z\in\mathbb{C}\setminus\mathbb{R}$,

$$\text{s-}\lim_{n\to\infty}(T_n - z)^{-1} = (S-z)^{-1}\otimes P$$

with an orthogonal projection P on $\mathscr{K}$. Denote by $E_\lambda(T_n)$ and $E_\lambda(S)$ the spectral family of T_n and S, respectively. Then: (i) If $a,b\in\mathbb{R}$, $a<b$, and $(a,b)\cap\sigma(T_n)=0$ for all n, then $(a,b)\cap\sigma(S)=0$. (ii) If $a,b\in\mathbb{R}$, $a<b$, and $a,b\notin\sigma_{\mathrm{pp}}(S)$ (the pure point spectrum of S), then $E_{(a,b)}(T_n)\to E_{(a,b)}(S)\otimes P$ strongly as $n\to\infty$.

Proof: The proof of part (i) is similar to that of Theorem VIII.24(a) in Ref. 21; we need only to note that

$$\|(S-z)^{-1}\| = \|(S-z)^{-1}\otimes P\|.$$

To prove part (ii), we note that Theorem A.1 applied to the present case gives

$$F(T_n)\to F(S)\otimes P$$

strongly for all $F\in C_\infty(\mathbb{R})$. Then the method of the proof of Theorem VIII.24(b) in Ref. 21 works and the desired result follows. ∎

Remark: Under the assumption of Theorem A.2 and the condition that $P\neq I$, $\{\sigma(T_n)\}_{n\in\mathbb{N}}$ cannot be bounded in n.

[1] W. Pauli, and M. Fierz, "Zur Theorie der Emission Langwelliger Lichtquanten," Nuovo Cimento **15**, 167 (1938).
[2] H. A. Bethe, "The electromagnetic shift of energy levels," Phys. Rev. **72**, 339 (1947).
[3] T. A. Welton, "Some observable effects of the quantum-mechanical fluctuations of the electromagnetic field," Phys. Rev. **74**, 1157 (1948).
[4] P. Blanchard, "Discussion mathématique du modéle de Pauli et Fierz relatif à la catastrophie infrarouge," Commun. Math. Phys. **15**, 156 (1969).
[5] A. Arai, "Self-adjointness and spectrum of Hamiltonians in nonrelativistic quantum electrodynamics," J. Math. Phys. **22**, 534 (1981).
[6] A. Arai, "Rigorous theory of spectra and radiation for a model in quantum electrodynamics," J. Math. Phys. **24**, 1896 (1983).
[7] A. Arai, "A note on scattering theory in non-relativistic quantum electrodynamics," J. Phys. A: Math. Gen. **16**, 49 (1983).
[8] *Foundations of Radiation Theory and Quantum Electrodynamics*, edited by A. O. Barut (Plenum, New York, 1980).
[9] T. Okamoto, and K. Yajima, "Complex scaling technique in non-relativistic QED," Ann. Inst. Henri Poincaré A **42**, 311 (1985).
[10] E. B. Davies, "Symmetry breaking for molecular open systems," Ann. Inst. H. Poincaré A **35**, 149 (1981).
[11] M. Fannes, B. Nachtergaele, and A. Verbeure, "Tunnelling in the equilibrium state of a spin-boson model," J. Phys. A: Math. Gen. **21**, 1759 (1988).
[12] M. Fannes, B. Nachtergaele, and A. Verbeure, "The equilibrium states of the spin-boson model," Commun. Math. Phys. **114**, 537 (1988).
[13] A. J. Leggett, S. Chakravarty, A. T. Dorsey, M. P. A. Fisher, A. Garg, and W. Zwerger, "Dynamics of the dissipative two-state system," Rev. Mod. Phys. **59**, 1 (1987).
[14] H. Spohn, and R. Dümcke, "Quantum tunneling with dissipation and the Ising model over R," J. Stat. Phys. **41**, 389 (1985).
[15] H. Spohn, "Ground state(s) of the spin-boson Hamiltonian," Commun. Math. Phys. **123**, 277 (1989).
[16] E. B. Davies, "Asymptotic analysis of some abstract evolution equations," J. Funct. Anal. **25**, 81 (1977).
[17] E. B. Davies, "Particle–boson interactions and the weak coupling limit," J. Math. Phys. **20**, 345 (1979).
[18] A. Arai, "Scaling limit for quantum systems of nonrelativistic particles interacting with a Bose field," Hokkaido Univ. Preprint Ser. in Math. No. 59, 1989.
[19] M. Reed, and B. Simon, *Methods of Modern Mathematical Physics II: Fourier Analysis, Self-adjointness* (Academic, New York, 1975).
[20] T. Kato, *Perturbation Theory for Linear Operators* (Springer, Berlin, 1976), 2nd ed.
[21] M. Reed and B. Simon, *Methods of Modern Mathematical Physics I: Functional Analysis* (Academic, New York, 1972).
[22] M. Reed and B. Simon, *Methods of Modern Mathematical Physics IV: Analysis of Operators* (Academic, New York, 1978).
[23] E. B. Davies, "Properties of the Green's functions of some Schrödinger operators," J. London Math. Soc. **7**, 483 (1973).
[24] E. Seiler and B. Simon, "Nelson's symmetry and all that in the Yukawa$_2$ and $(\phi^4)_3$ field theories," Ann. Phys. (NY) **97**, 470 (1976).

Moyal quantization with compact symmetry groups and noncommutative harmonic analysis

Héctor Figueroa, José M. Gracia-Bondía, and Joseph C. Várilly
Escuela de Matemática, Universidad de Costa Rica, San José, Costa Rica

(Received 4 January 1990; accepted for publication 16 May 1990)

The phase-space approach to quantization of systems whose symmetry group is compact and semisimple is developed from two basic principles: covariance and traciality. This generalizes results and methods already implemented for spin systems [J. C. Várilly and J. M. Gracia-Bondía, Ann. Phys. **190**, 107 (1989)]. The twisted product of phase-space functions is shown to be the image of group convolution in the context of a novel Fourier theory on the coadjoint orbits.

I. INTRODUCTION

The abstract nature of the expansions of functions on compact Lie groups in terms of equivalence classes of unitary representations has been the source of some dissatisfaction.[1-3] One of the purposes of this paper is to reformulate the harmonic analysis on compact groups so that it closely parallels standard Fourier analysis on commutative Lie groups. Interestingly enough, this arose from a program of research on the phase-space formulation of quantum mechanics.

The phase-space approach to quantum mechanics started from Moyal's observation[4] that the mapping from density matrices to Wigner functions[5] of the phase-space variables served to invert the Weyl correspondence[6] between phase-space functions and operators on Hilbert space. The subject was appreciably generalized and revived a decade ago in the "deformation" paradigm of Bayen *et al.*,[7] who, in fact, proposed to develop quantum theory directly from phase-space functions composed via the twisted or "star" product, by-passing any reference to Hilbert-space operators. In "deformation quantization," the systems under study always possess an intrinsic group of symmetries. It turns out that, even allowing only for quantizations compatible with the action of that group, in most cases a large variety of "star-products" appear; in C^*-algebraic versions of deformation theory, this nonuniqueness remains, as has been shown recently by Rieffel.[8]

It would seem that a supplementary condition is needed in order to select a (preferably unique) star-product ensuring the physical equivalence of the quantization with the ordinary brand of quantum mechanics, obtaining thus a true generalization of the formalism that Moyal originally developed for nonrelativistic spinless particles. Recently, two of us have indeed shown how this can be done, for the physically relevant cases of pure-spin systems,[9] Galilean particles of arbitrary spin,[10] and (together with Cariñena) relativistic theories also.[11]

In all systems so far considered, the phase space is a coadjoint orbit of a symmetry group, the relevant objects on this phase space transform covariantly under the group action and satisfy a *tracial property* that allows one to compute expected values of observables as integrals over the orbit. This tracial property not only allows physical interpretation of the phase-space functions, but also provides a rule to single out the "correct" symbol calculus from among the many that could be proposed. Just as in ordinary Moyal theory, a *quantizer* or operator kernel is found that implements the transfer between phase-space functions and operators in a two-way fashion.

In the aforementioned particular cases, direct methods sufficed; here, however, we seek a general prescription. This article shows that the Moyal quantization exists and is essentially unique for all elementary systems governed by compact connected semisimple Lie groups. As a first by-product of the construction, we obtain the twisted product of functions on coadjoint orbits, which allows the development of an autonomous phase-space theory. A second important by-product is the novel Fourier analysis theory alluded to at the beginning of this introduction.

The organization of the paper is as follows. In Sec. II we identify the coadjoint orbits and review the construction of coherent states for compact simply connected Lie groups. This technical prelude serves mainly to fix the notation. The only result of note here is a theorem of Wildberger[3] that establishes the faithfulness of the Berezin "covariant symbol" calculus[12] in the compact case.

In Sec. III we construct the *Stratonovich–Weyl quantizer*, which implements the correspondence between functions on the coadjoint orbits and operators on the representation spaces of the group. We show how it may be explicitly computed in terms of a set of special functions on the coadjoint orbits generalizing the usual spherical harmonics. As a side-effect, we obtain the explicit form of the Berezin "contravariant symbols" for any compact simply connected group.

In Sec. IV we write down the twisted product of functions on coadjoint orbits. We redefine the Fourier transformation for a compact simply connected group as a transformation between function spaces by means of a *scalar* "Fourier kernel" which, in fact, arises from the Stratonovich–Weyl quantizer and is the exact analog of the exponentials for commutative Fourier analysis. The dual $\hat{G}$ of the group G under consideration is exhibited as a G-manifold with a countably infinite set of connected components, carrying a naturally defined measure m so that the Fourier transformation is an isometry of $L^2(G)$ into $L^2(\hat{G},dm)$.

 0022-2488/90/112664-08$03.00

The tracial property of the Stratonovich–Weyl symbols is what makes the Fourier theory work. The twisted product is then shown to be the Fourier image of the group convolution.

II. COADJOINT ORBITS AND COHERENT STATES

The Lie groups G we initially consider will be compact and simply connected, therefore semisimple. An *elementary classical system* is a symplectic manifold that is homogeneous for a symplectic action of G. Such manifolds are isomorphic to coadjoint orbits of G (these are simply connected). The first task is then to identify the coadjoint orbits of G, which are the phase spaces of our theory.

We use the following notation: $\mathfrak{g}$ denotes the Lie algebra of the Lie group G and $\mathfrak{g}^*$ its dual vector space. The coadjoint action $u \mapsto g\cdot u$ of G on $\mathfrak{g}^*$ is defined by

$$\langle g\cdot u, X\rangle := \langle u, \mathrm{Ad}(g^{-1})X\rangle \quad (g\in G, X\in\mathfrak{g}),$$

where Ad denotes the adjoint action of G on $\mathfrak{g}$. We write $\mathcal{O}_u$ for the coadjoint orbit of $u\in\mathfrak{g}^*$. The isotropy subgroup of u is denoted G_u; its tangent space is the Lie subalgebra $\mathfrak{g}_u$ of $\mathfrak{g}$.

Now

$$\mathfrak{g}_u = \{X\in\mathfrak{g}: \langle u, [X,Y]\rangle = 0, \text{ for all } Y\in\mathfrak{g}\}.$$

Since $\mathcal{O}_u \approx G/G_u$, we have that the tangent space $T_u(\mathcal{O}_u) \cong \mathfrak{g}/\mathfrak{g}_u$ and $(X,Y)\mapsto\langle u,[X,Y]\rangle$ induces a nondegenerate alternating bilinear form on $\mathfrak{g}/\mathfrak{g}_u$ which gives rise to a G invariant symplectic two-form ω_u on $\mathcal{O}_u$.[13] The exterior product of $\frac{1}{2}(\dim \mathcal{O}_u)$ copies of ω_u gives a G invariant volume form m_u (Liouville measure) on $\mathcal{O}_u$.

The coadjoint orbits may be explicitly determined by the following argument. Let ad be the differential of Ad. Since G is semisimple, the negative Killing form $\langle X,Y\rangle_K := -\mathrm{Tr}[\mathrm{ad}(X)\mathrm{ad}(Y)]$ is definite and allows us to identify $\mathfrak{g}^*$ with $\mathfrak{g}$; let $U\in\mathfrak{g}$ correspond to $u\in\mathfrak{g}^*$. Then

$$\langle u,[X,Y]\rangle = \langle U,[X,Y]\rangle_K = \langle [U,X],Y\rangle_K$$

and so $\mathfrak{g}_u = \{X\in\mathfrak{g}: \mathrm{ad}(X)U = 0\}$. Thus

$$\begin{aligned} G_u &= \{g\in G: \mathrm{Ad}(g)U = U\} \\ &= \{g\in G: g\exp(tU)g^{-1} = \exp(tU), \text{ for all } t\in\mathbb{R}\} \\ &= Z_G(T_1), \end{aligned}$$

where $Z_G(T_1)$ is the centralizer of the torus T_1 which is the closure of $\{\exp(tU): t\in\mathbb{R}\}$. Since G is a compact connected group, $Z_G(T_1)$ is the union of the maximal tori of G which contain T_1.[14] Choose a Weyl chamber K (and thus fix a set of positive roots of G) so that $u\in\overline{K}$. The torus T_1 is not maximal if and only if u lies in a wall of K. The generic coadjoint orbits thus correspond to the case $u\in K$, are of maximal dimension, and form a single stratum; indeed, since all maximal tori are conjugate, the maximal-dimensional orbits are all of the form $G_u \cong G/T$, where T is a fixed maximal torus of G.

We say that the orbit $\mathcal{O}_u$ is *integral* if ω_u is integral, i.e., has integer-valued integrals over integral two-cycles. By Kostant's theorem,[15] this occurs when the real-linear form $2\pi i u: \mathfrak{g}_u \to i\mathbb{R}$ lifts to a character $\chi_u: G_u \to U(1)$ defined by $\chi_u(\exp X) := e^{2\pi i\langle u,X\rangle}$ for $X\in\mathfrak{g}_u$. These are the orbits that correspond to unitary irreducible representations of G via holomorphic induction of χ_u. More precisely, the complexified Lie algebra $\mathfrak{g}^{\mathbb{C}}$ has a complex subalgebra $\mathfrak{p}$ containing the complexification of $\mathfrak{g}_u$ and also the root spaces of the positive roots of G. If P is the subgroup of G with Lie algebra $\mathfrak{p}$ (parabolic subgroup), then χ_u extends to a one-dimensional holomorphic representation of P. This gives rise to a holomorphic representation of the complexified group $G^{\mathbb{C}}$ by induction of χ_u; its restriction to G is a unitary irreducible representation ρ_u of G. Furthermore, the Borel–Weil theorem[16,17] says that every unitary irrep of G is obtained in this way, and that ρ_u and ρ_v are unitarily equivalent if and only if u and v belong to the same integral coadjoint orbit. Thus, if $\mathcal{O}_u$ is an integral orbit, then

$$\mathcal{O}_u \approx G/Z_G(T_1) \approx G^{\mathbb{C}}/P$$

is a compact Kähler manifold.

We now fix a unitary irreducible representation ρ of G. Fix a set of positive roots for G. Let $\mathfrak{t}$ be the Lie algebra of the maximal torus T and let $\lambda\in\mathfrak{t}^*$ be the corresponding highest (real) weight of ρ (we will write ρ_λ for ρ whenever more than one unitary irrep must be considered). Since G is semisimple and we have identified $\mathfrak{g}^*$ with $\mathfrak{g}$, we can regard $\mathfrak{t}^*$ as a subset of $\mathfrak{g}^*$, so that $\lambda\in\mathfrak{g}^*$. Let $\mathcal{O}_\lambda$ be the coadjoint orbit of G through λ. The orbit $\mathcal{O}_\lambda$ is integral since λ is an integral linear form on $\mathfrak{g}$,[14] and the correspondence $\rho_\lambda \mapsto \mathcal{O}_\lambda$ is reciprocal to the holomorphic-induction method of passing from integral coadjoint orbits to unitary irreps of G.

Remark: It is more fashionable nowadays to associate ρ_λ with $\mathcal{O}_{\lambda+\delta}$, where δ is half the sum of the positive real roots; for instance, if one wishes to exploit Kirillov's character formula.[18] (Since G is simply connected, this is again an integral orbit.) In this way, one avoids dealing with nongeneric orbits at all; but for our purposes here, the earlier correspondence is perhaps more natural.

What is needed to effect a phase-space quantization is the construction of a "quantizer," which maps functions on the phase space $\mathcal{O}_\lambda$ to operators on the representation space of ρ_λ. The simplest method of proceeding to prove *existence* of the quantizer seems to be to use the coherent-state formalism,[3,19] which we outline as follows.

Let V_λ be the representation space of ρ, which is a Hilbert space of finite dimension d_λ. We shall employ Dirac's bra-ket notation for its elements whenever convenient, and otherwise the inner product will be denoted by $(\cdot,\cdot)$. Let $|\lambda\rangle\in V_\lambda$ be a weight vector of λ of norm 1 (the choice of $|\lambda\rangle$ is unique up to a phase factor).

For any unit vector $v\in V_\lambda$ we may define $\Phi(v)\in\mathfrak{g}^*$ by

$$\langle\Phi(v),X\rangle := (1/2\pi i)(v, d\rho(X)v);$$

then the momentum mapping Φ is equivariant, i.e., $\Phi(\rho(g)v) = g\cdot\Phi(v)$; $\Phi(|\lambda\rangle) = \lambda$, and (since the weight space of λ is one-dimensional)

$$\Phi^{-1}(\lambda) = \{z|\lambda\rangle : z\in\mathbb{C}, |z| = 1\}.$$

Thus

$$\Phi^{-1}(\mathcal{O}_\lambda) = \{\rho(g)|\lambda\rangle : g\in G\},$$

since $\Phi(v) = g\cdot\lambda$ iff $\Phi(\rho(g^{-1})v) = \lambda$ iff $v = \rho(g)z|\lambda\rangle = \rho(g)\rho(h)|\lambda\rangle$ for some $h\in G_\lambda$, since λ is a real weight of ρ.

For each $u\in\mathcal{O}_\lambda$, choose $g_u\in G$ so that $g_u\cdot\lambda = u$

and $g_\lambda = 1$ (and so that $u \mapsto g_u$ is a measurable section of $G \to \mathcal{O}_\lambda$). We then define $|u\rangle := \rho(g_u)|\lambda\rangle \in V_\lambda$. The equivariance of Φ shows that $\Phi(|u\rangle) = u$ for any $u \in \mathcal{O}_\lambda$, so that $|u\rangle$ is determined up to a phase factor by u.

Definition 1: The *covariant symbol*[12] or *Q-symbol* of an operator $A \in \mathcal{L}(V_\lambda)$ is defined as the function on the coadjoint orbit $\mathcal{O}_\lambda$ given by

$$Q_A(u) := \langle u|A|u\rangle. \tag{1}$$

Since u determines $|u\rangle$ up to a phase factor, $Q_A(u)$ depends only on u; in fact, Q_A is a continuous function on $\mathcal{O}_\lambda$. Let $\mathcal{S}_\lambda$ denote the (finite-dimensional) space of functions $\{Q_A : A \in \mathcal{L}(V_\lambda)\}$.

The essential point of the coherent state construction is that no information is lost on passing from operators $A \in \mathcal{L}(V_\lambda)$ to their symbols Q_A. This can be verified by analytic continuation arguments, using the Kähler structure of $\mathcal{O}_\lambda$. However, a direct algebraic proof for general compact semisimple groups was first given in an unpublished article of Wildberger.[3] We give our version of his argument here since it provides an important stepping stone to the Stratonovich–Weyl quantizer.

Theorem 1: (Wildberger) The symbol map $Q : \mathcal{L}(V_\lambda) \to \mathcal{S}_\lambda$ is one-to-one.

Proof: For each positive root α of G we can choose root vectors X_α, $Y_\alpha \in \mathfrak{g}$ for the roots α, $-\alpha$, respectively, so that

$$(v, d\rho(X_\alpha)w) = (d\rho(Y_\alpha)v, w), \quad \text{for all } v,w \in V_\lambda.$$

We partially order the (complex) weights of ρ by $\mu > \nu$ iff $\mu - \nu = c_1\alpha_1 + \cdots + c_k\alpha_k$, where $\{\alpha_1,\ldots,\alpha_k\}$ are the simple roots and $c_1,\ldots,c_k$ are non-negative integers not all zero. There is a greatest element Λ for this ordering, namely the extension to $\mathfrak{t}^{\mathbb{C}}$ of the functional $2\pi i\lambda$ on $\mathfrak{t}$. We order pairs of weights lexicographically by $(\mu,\mu') > (\nu,\nu')$ if $\mu > \nu$ or $\mu = \nu$ and $\mu' > \nu'$. Let H_μ denote the weight space of μ; then

$$\mathcal{L}(V_\lambda) = \bigoplus_{\mu,\nu} L_{\mu,\nu},$$

where $L_{\mu,\nu}$ is the vector space spanned by $\{|v\rangle\langle w| : v \in H_\mu, w \in H_\nu\}$. For each $A \in \mathcal{L}(V_\lambda)$, let $A = \Sigma_{\mu,\nu} A_{\mu,\nu}$ be the corresponding decomposition of A, with $A_{\mu,\nu} \in L_{\mu,\nu}$.

The representation $\pi := \rho \otimes \bar{\rho}$ of G acts on $\mathcal{L}(V_\lambda)$ by $\pi(g)A = \rho(g)A\rho(g^{-1})$. For $X \in \mathfrak{g}^{\mathbb{C}}$, we have $d\pi(X)A = d\rho(X)A - A\,d\rho(X)$. In particular,

$$d\pi(X_\alpha)|v\rangle\langle w| = |d\rho(X_\alpha)v\rangle\langle w| - |v\rangle\langle d\rho(Y_\alpha)\omega|,$$

$$d\pi(Y_\alpha)|v\rangle\langle w| = |d\rho(Y_\alpha)v\rangle\langle w| - |v\rangle\langle d\rho(X_\alpha)w|.$$

Suppose now that $A \in \mathcal{L}(V_\lambda)$ with $A \neq 0$, but $Q_A = 0$ in $\mathcal{S}_\lambda$. Then $\langle u|\pi(g)A|u\rangle = \langle g^{-1}\cdot u|A|g^{-1}\cdot u\rangle = 0$ and so $\langle u|d\pi(X)A|u\rangle = 0$ for $u \in \mathcal{O}_\lambda$, $X \in \mathfrak{g}$.

Let (ξ, η) be a maximal pair of weights for which we can find an A with $Q_A = 0$ and $A_{\xi,\eta} \neq 0$. If $\xi < \Lambda$, then $A_{\xi,\eta} = \Sigma_{i=1}^m |v_i\rangle\langle w_i|$, where $\{w_1,\ldots,w_m\}$ is a basis for H_η and $v_i \in H_\xi$ with, say, $v_1 \neq 0$. Since $\eta < \Lambda$ and ρ is irreducible, $d\rho(X_\alpha)v_1 \neq 0$ for some positive root α. But then $d\pi(X_\alpha)A$ has Q-symbol 0 and a nonzero $(\xi + \alpha, \eta)$-component, contradicting the maximality of ξ. Thus $\xi = \Lambda$.

Next, if $\eta < \Lambda$ then $A_{\Lambda,\eta} = |v\rangle\langle w|$ with $v \in H_\Lambda$, $w \in H_\eta$, and $w \neq 0$; so for some positive root α, we have that $d\pi(Y_\alpha)A$ has Q-symbol 0 and a nonzero $(\Lambda, \eta + \alpha)$ component, contradicting the maximality of η. Thus $\eta = \Lambda$, and so $A_{\Lambda,\Lambda}$ is a nonzero scalar.

Now $|\lambda\rangle \in H_\Lambda$, so

$$Q_A(\lambda) = \langle\lambda|A|\lambda\rangle = A_{\Lambda,\Lambda}\langle\lambda|\lambda\rangle \neq 0,$$

contradicting $Q_A = 0$. This shows that $Q_A = 0$ only if $A = 0$. ∎

III. MOYAL QUANTIZATION

We now have the basic ingredients to construct a *Stratonovich–Weyl quantizer.* We follow the path set forth in Ref. 9, Sec. 4, for the group SU(2).

It turns out that the Q-symbols are not satisfactory for this task, as they lack the all-important "tracial property" (see below). Another class of symbols, called "contravariant"[12] or "P-symbols," may be defined by duality. To see this, note that $\mathcal{L}(V_\lambda)$ is a Hilbert space under the inner product $(A,B) := \mathrm{Tr}[A^\dagger B]$, but that $A \mapsto Q_A$ is not an isometry from this Hilbert space into $L^2(\mathcal{O}_\lambda)$. Nonetheless, the linear form on $\mathcal{S}_\lambda$ given by $Q_B \mapsto \mathrm{Tr}[A^\dagger B]$ corresponds by the Riesz theorem to a unique $P_A \in \mathcal{S}_\lambda$ such that

$$(P_A|Q_B) := \int P_A^*(u)Q_B(u)\,dm_\lambda(u) = \mathrm{Tr}[A^\dagger B]. \tag{2}$$

(Here,* denotes the complex conjugate function and † denotes the adjoint operator.) Thus $A \mapsto P_A$ is a linear bijection of $\mathcal{L}(V_\lambda)$ onto $\mathcal{S}_\lambda$, since both spaces have the same finite dimension by Theorem 1. In order that P_I be the constant function 1, we must renormalize the Liouville measure on the orbit so that $m_\lambda(\mathcal{O}_\lambda) = d_\lambda$; we maintain this normalization hereinafter.

It is straightforward to check the *covariance* of the Q- and P-symbols:

$$Q_{\rho(g)A\rho(g^{-1})}(u) = Q_A(g^{-1}\cdot u),$$

$$P_{\rho(g)A\rho(g^{-1})}(u) = P_A(g^{-1}\cdot u), \tag{3}$$

for all $u \in \mathcal{O}_\lambda$. If we denote by τ the quasiregular representation of G on $L^2(\mathcal{O}_\lambda)$ restricted to $\mathcal{S}_\lambda$, i.e., $[\tau(g)f](u) = f(g^{-1}\cdot u)$, the covariance property (3) may be rephrased as

$$Q_{\pi(g)A} = \tau(g)Q_A, \quad P_{\pi(g)A} = \tau(g)P_A.$$

Theorem 2: $K : Q_A \mapsto P_A$ is a positive invertible operator on $\mathcal{S}_\lambda$, which commutes with $\{\tau(g) : g \in G\}$.

Proof: K is bijective since $Q_A \mapsto A$, $A \mapsto P_A$ are bijections. The positivity of K is seen, using (2), from

$$(KQ_A|Q_A) = (P_A|Q_A) = \mathrm{Tr}[A^\dagger A] \geqslant 0.$$

The fact that $\tau(g)KQ_A = K\tau(g)Q_A$ is just a restatement of (3). ∎

We choose an orthonormal basis for $\mathcal{S}_\lambda$ for which any operator commuting with $\tau(G)$ has a matrix which is a direct sum of scalar blocks, one for each irreducible component of τ. As a consequence of Theorem 2, K and its positive square root $K^{1/2}$ are of this form. We can thus introduce a symbol that is "halfway" between the Q- and P-symbols.

Definition 2: For $A \in \mathcal{L}(V_\lambda)$, we define the *Stratonovich–Weyl symbol* $W_A \in \mathcal{S}_\lambda$ by

$$W_A := K^{1/2}Q_A = K^{-1/2}P_A. \tag{4}$$

Theorem 3: The symbol map $A \mapsto W_A$ has the following properties:

(i) it is a linear one-to-one map from $\mathcal{L}(V_\lambda)$ onto $\mathcal{S}_\lambda$;

(ii) $W_{A\dagger} = W_A^*$;

(iii) W_I is the constant function 1;

$$\text{(iv)}\ W_{\rho(g)A\rho(g^{-1})}(u) = W_A(g^{-1}\cdot u), \quad \text{for } u \in \mathcal{O}_\lambda, g \in G; \tag{5}$$

$$\text{(v)} \int_{\mathcal{O}_\lambda} W_A(u) W_B(u) dm_\lambda(u) = \mathrm{Tr}[AB]. \tag{6}$$

Proof: The bijectivity of $A \mapsto W_A$ follows from the established bijectivity of the Q-symbol map and the invertibility of K.

It follows from (1) and (2) that $Q_{A\dagger} = Q_A^*$ and $P_{A\dagger} = P_A^*$ for all A; thus K and hence $K^{1/2}$ commutes with complex conjugation, and so

$$W_{A\dagger} = K^{1/2}(Q_A^*) = (K^{1/2}Q_A)^* = W_A^*.$$

Since $Q_I = P_I = 1$, we have $W_I = K^{1/2}(1) = 1$ also.

The covariance property (5) is an immediate consequence of (3) and (4).

The *tracial property* (6) follows from (2), since

$$\begin{aligned}\int_{\mathcal{O}_\lambda} & W_A(u) W_B(u) dm_\lambda(u) \\ &= (W_A^* | W_B) = (K^{1/2}Q_{A\dagger} | K^{1/2}Q_B) \\ &= (KQ_{A\dagger} | Q_B) = (P_{A\dagger} | Q_B) = \mathrm{Tr}[AB]. \end{aligned}$$ ■

The inverse map $W_A \mapsto A$ generalizes the Weyl correspondence.[6] We have baptized it the *Stratonovich–Weyl correspondence*, because conditions essentially equivalent to those of Theorem 3 were first outlined in a remarkable paper by Stratonovich.[20]

Remark 1: The importance of the property (v) cannot be overestimated. It tells us that quantum and classically calculated averages coincide, thus ensuring that we are building a genuine Moyal theory. In fact, Theorem 3 says that the Stratonovich–Weyl correspondence is a unitary equivalence of the representation τ of G on $\mathcal{S}_\lambda$, regarded as a subspace of $L^2(\mathcal{O}_\lambda, dm_\lambda)$, and the representation π of G on $\mathcal{L}(V_\lambda)$. In other words, $\mathcal{S}_\lambda$ is a representation space of the representation $\pi = \rho \otimes \bar{\rho}$. Thus one can decompose $\mathcal{S}_\lambda$ as

$$\mathcal{S}_\lambda = \oplus_\gamma n_\lambda(\gamma) V_\gamma,$$

where the sum runs over the highest weights γ of the unitary irreducible subrepresentations of $\rho \otimes \bar{\rho}$, and $n_\lambda(\gamma)$ is the multiplicity of the corresponding subrepresentation. In particular, since ρ is irreducible, Schur's lemma shows that

$$n_\lambda(0) = \dim(\{A \in \mathcal{L}(V_\lambda) : \pi(g)A = A, \text{ for all } g \in G\}) = 1.$$

In retrospect, we see that one of the main points of the previous theorems is to establish that $L^2(\mathcal{O}_\lambda)$ accomodates a representation equivalent to $\rho \otimes \bar{\rho}$. This was recognized some time ago as a *necessary* condition to have a star-product on $\mathcal{O}_\lambda$ with complex conjugation as an involution.[21]

Remark 2: Suppose that $A \mapsto \tilde{W}_A$ is any symbol map satisfying the conditions of Theorem 3, its support space being a d_λ^2-dimensional subspace of $L^2(\mathcal{O}_\lambda)$ which we can take to be $\mathcal{S}_\lambda$. The intertwining operator $T: W_A \mapsto \tilde{W}_A$ is then a unitary operator on $\mathcal{S}_\lambda$ which commutes with $\tau(G)$. Thus the matrix of T (with respect to the basis which diagonalizes K) is a direct sum of unitary blocks of size $d_\gamma \times d_\gamma$, and each block commutes with the complex-conjugation operator. Moreover, since $T(1) = 1$, the 1×1 block corresponding to $\gamma = 0$ is 1. Conversely, any matrix with these properties defines an operator T so that $A \mapsto T(W_A)$ satisfies the conditions of Theorem 3.

For instance, if $\rho \otimes \bar{\rho}$ is simply reducible [i.e., $n_\lambda(\gamma) = 1$ for each γ], then T has a diagonal matrix with diagonal entries ± 1, the sign being positive for the $(\gamma = 0)$ entry.

In summary, the Stratonovich–Weyl correspondence for compact semisimple groups always exists, but need not be unique: The conditions given for the intertwining operator T provide the exact measure of nonuniqueness. In other words, the correspondence K between covariant and contravariant symbols has in general several square roots which are reality-preserving, one-preserving, covariant in the sense of (5), and tracial. Among these we have selected the *positive* square root (4), to obtain a standard recipe for the Stratonovich–Weyl correspondence.

So far, our definition of the Stratonovich–Weyl symbols has been entirely nonconstructive. We now give an explicit computational construction of the Stratonovich–Weyl symbols; as a bonus, we derive by the same procedure an explicit formula for the Berezin P-symbols. In order to do so, we find the diagonal elements of the matrix of K.

The reproducing kernel of the Hilbert space $\mathcal{S}_\lambda$ may be written as

$$I_\lambda(u,v) := \sum_{\gamma,r,k} Y^\lambda_{\gamma rk}(u) Y^{\lambda *}_{\gamma rk}(v),$$

where $\{Y^\lambda_{\gamma rk} : k = 1,\ldots,d_\gamma\}$ is an orthonormal basis of the rth copy of the unitary irrep ρ_γ (as a subrepresentation of τ), for $r = 1,\ldots,n_\lambda(\gamma)$. Note that $n_\lambda(\gamma) \leqslant d_\lambda$ since the projection $G \to G/G_\lambda \approx \mathcal{O}_\lambda$ induces a one-to-one mapping of $L^2(\mathcal{O}_\lambda, dm_\lambda)$ into $L^2(G, dg)$, where, as usual, $\int_G dg = 1$. Thus we can choose an orthonormal basis for $\mathcal{S}_\lambda$ by the recipe

$$Y^\lambda_{\gamma rk}(u) := \sqrt{d_\gamma / d_\lambda}\, \mathcal{D}^\gamma_{kr}(g_u),$$

where $\mathcal{D}^\gamma_{kr}$ are matrix elements of the unitary irrep ρ_γ. (The $Y^\lambda_{\gamma rk}$ may be thought of as a kind of generalized spherical harmonics on compact coadjoint orbits.) Thus

$$I_\lambda(u,v) := \sum_\gamma \frac{d_\gamma}{d_\lambda} \sum_{r=1}^{n_\lambda(\gamma)} \sum_{k=1}^{d_\gamma} \mathcal{D}^\gamma_{kr}(g_u) \mathcal{D}^{\gamma *}_{kr}(g_v). \tag{7}$$

We determine the kernel corresponding to the operator K in the following way. The Clebsch–Gordan coefficients corresponding to the decomposition $\rho_\lambda \otimes \rho_\mu = \oplus_\gamma n_{\lambda\mu}(\gamma) \rho_\gamma$ are computed from the matrix elements of ρ_λ, ρ_μ, and ρ_γ by integrating over G. It can be proved that

$$\begin{aligned}\left\langle \begin{matrix} \lambda & \mu \\ i & j \end{matrix} \middle| \begin{matrix} \gamma, r \\ k \end{matrix} \right\rangle & \left\langle \begin{matrix} \gamma, r \\ k' \end{matrix} \middle| \begin{matrix} \lambda & \mu \\ i' & j' \end{matrix} \right\rangle \\ &= d_\gamma \int_G \mathcal{D}^{\gamma *}_{kk'}(g) \mathcal{D}^\lambda_{ii'}(g) \mathcal{D}^\mu_{jj'}(g) dg,\end{aligned}$$

where $i,i' = 1,\ldots,d_\lambda$; $j,j' = 1,\ldots,d_\lambda$; $k,k' = 1,\ldots,d_\gamma$, with an obvious notation for the kets in the several representation spaces. Let $\bar{\lambda}$ denote the highest weight of $\overline{\rho_\lambda}$; of course, $\mathscr{D}^{\bar{\lambda}}_{ij}(g) = \mathscr{D}^{\lambda *}_{ij}(g)$ since $\rho_{\bar{\lambda}} = \overline{\rho_\lambda}$ is the conjugate representation of ρ_λ. Thus, in particular, we have

$$\left\langle \begin{matrix} \lambda & \bar{\lambda} \\ i & j \end{matrix} \middle| \begin{matrix} \gamma,r \\ k \end{matrix} \right\rangle \left\langle \begin{matrix} \gamma,r \\ r \end{matrix} \middle| \begin{matrix} \lambda & \bar{\lambda} \\ \lambda & \bar{\lambda} \end{matrix} \right\rangle = d_\gamma \int_G \mathscr{D}^{\gamma *}_{kr}(g)\mathscr{D}^{\lambda}_{i\lambda}(g)\mathscr{D}^{\lambda *}_{j\lambda}(g)dg, \tag{8}$$

where, of course, $|{}^{\lambda}_{\lambda}\rangle$ now denotes the highest weight state. We have selected the phase factor for these indices so that

$$\left\langle \begin{matrix} \gamma,r \\ r \end{matrix} \middle| \begin{matrix} \lambda & \bar{\lambda} \\ \lambda & \bar{\lambda} \end{matrix} \right\rangle = \left\langle \begin{matrix} \lambda & \bar{\lambda} \\ \lambda & \bar{\lambda} \end{matrix} \middle| \begin{matrix} \gamma,r \\ r \end{matrix} \right\rangle > 0.$$

It follows from the Peter–Weyl theorem and (8) that

$$\mathscr{D}^{\lambda}_{i\lambda}(g)\mathscr{D}^{\lambda *}_{j\lambda}(g) = \sum_{\gamma,r,k} \left\langle \begin{matrix} \lambda & \bar{\lambda} \\ i & j \end{matrix} \middle| \begin{matrix} \gamma,r \\ k \end{matrix} \right\rangle \left\langle \begin{matrix} \gamma,r \\ r \end{matrix} \middle| \begin{matrix} \lambda & \bar{\lambda} \\ \lambda & \bar{\lambda} \end{matrix} \right\rangle \mathscr{D}^{\gamma}_{kr}(g).$$

From (2) we obtain

$$Q_A(u) = \langle u|A|u\rangle = \left\langle u \middle| \int_{\mathscr{O}_\lambda} P_A(v)|v\rangle\langle v|dm_\lambda(v) \middle| u \right\rangle = \int_{\mathscr{O}_\lambda} P_A(v)|\langle u|v\rangle|^2 dm_\lambda(v),$$

so that $|\langle u|v\rangle|^2$ is the kernel of K^{-1}. Now

$$\begin{aligned} |\langle u|v\rangle|^2 &= |\langle g_u\cdot\lambda\,|g_v\cdot\lambda\rangle|^2 \\ &= |\langle\lambda\,|\rho(g_u^{-1})\rho(g_v)|\lambda\rangle|^2 \\ &= \left|\sum_{j=1}^{d_\lambda} \mathscr{D}^{\lambda *}_{j\lambda}(g_u)\mathscr{D}^{\lambda}_{j\lambda}(g_v)\right|^2 \\ &= \sum_{\substack{\gamma,r,k \\ \gamma',r',k'}} \sum_{i,j=1}^{d_\lambda} \left\langle \begin{matrix} \lambda & \bar{\lambda} \\ i & j \end{matrix} \middle| \begin{matrix} \gamma,r \\ k \end{matrix} \right\rangle \left\langle \begin{matrix} \gamma,r \\ r \end{matrix} \middle| \begin{matrix} \lambda & \bar{\lambda} \\ \lambda & \bar{\lambda} \end{matrix} \right\rangle \left\langle \begin{matrix} \lambda & \bar{\lambda} \\ \lambda & \bar{\lambda} \end{matrix} \middle| \begin{matrix} \gamma',r' \\ r' \end{matrix} \right\rangle \left\langle \begin{matrix} \gamma',r' \\ k' \end{matrix} \middle| \begin{matrix} \lambda & \bar{\lambda} \\ i & j \end{matrix} \right\rangle \mathscr{D}^{\gamma}_{kr}(g_u)\mathscr{D}^{\gamma *}_{k'r'}(g_v) \\ &= \sum_{\gamma,r,k} \left\langle \begin{matrix} \lambda & \bar{\lambda} \\ \lambda & \bar{\lambda} \end{matrix} \middle| \begin{matrix} \gamma,r \\ r \end{matrix} \right\rangle^2 \mathscr{D}^{\gamma}_{kr}(g_u)\mathscr{D}^{\gamma *}_{kr}(g_v), \end{aligned}$$

using the orthogonality relations for the Clebsch–Gordan coefficients. Comparing this kernel with (7), we see that the orthonormal basis $\{Y^{\lambda}_{\gamma rk}\}$ diagonalizes K^{-1}. It is now clear that each $d_\gamma \times d_\gamma$ block of K is a scalar matrix with eigenvalue

$$\frac{d_\gamma}{d_\lambda} \left\langle \begin{matrix} \lambda & \bar{\lambda} \\ \lambda & \bar{\lambda} \end{matrix} \middle| \begin{matrix} \gamma,r \\ r \end{matrix} \right\rangle^{-2} = \left\langle \begin{matrix} \lambda & \bar{\gamma},r \\ \lambda & r \end{matrix} \middle| \begin{matrix} \lambda \\ \lambda \end{matrix} \right\rangle^{-2}.$$

Thus we arrive at last at an explicit formula for the Stratonovich–Weyl symbols:

$$W_A(u) = \sum_{\gamma,r,k} \left\langle \begin{matrix} \lambda & \bar{\gamma},r \\ \lambda & r \end{matrix} \middle| \begin{matrix} \lambda \\ \lambda \end{matrix} \right\rangle^{-1} \times \int_{\mathscr{O}_\lambda} Q_A(v) Y^{\lambda *}_{\gamma rk}(v) dm_\lambda(v) Y^{\lambda}_{\gamma rk}(u) \tag{9}$$

(choosing the phase factor so that

$$\left\langle \begin{matrix} \lambda & \bar{\gamma},r \\ \lambda & r \end{matrix} \middle| \begin{matrix} \lambda \\ \lambda \end{matrix} \right\rangle$$

is positive).

It is worth noting that the same arguments yield an explicit expression for Berezin's contravariant symbol:

$$P_A(u) = \sum_{\gamma,r,k} \left\langle \begin{matrix} \lambda & \bar{\gamma},r \\ \lambda & r \end{matrix} \middle| \begin{matrix} \lambda \\ \lambda \end{matrix} \right\rangle^{-2} \times \int_{\mathscr{O}_\lambda} Q_A(v) Y^{\lambda *}_{\gamma rk}(v) dm_\lambda(v) Y^{\lambda}_{\gamma rk}(u), \tag{10}$$

which generalizes the P-symbol given in Ref. 9 for the SU(2) case. As far as we know, this expression has not hitherto been derived. We remark that the formulas (9) and (10) may be gainfully used to compute quantum partition functions.

The symbol calculus is now reformulated by introducing a family of self-adjoint operators indexed by the points of phase space, namely, the Stratonovich–Weyl quantizer:

$$\Omega_\lambda(u) := \sum_{\gamma,r,k} \left\langle \begin{matrix} \lambda & \bar{\gamma},r \\ \lambda & r \end{matrix} \middle| \begin{matrix} \lambda \\ \lambda \end{matrix} \right\rangle^{-1} \times \int_{\mathscr{O}_\lambda} Y^{\lambda *}_{\gamma rk}(v)|v\rangle\langle v|dm_\lambda(v) Y^{\lambda}_{\gamma rk}(u), \tag{11}$$

for which

$$W_A(u) = \mathrm{Tr}[A\Omega_\lambda(u)], \tag{12a}$$

$$A = \int_{\mathscr{O}_\lambda} W_A(u)\Omega_\lambda(u)dm_\lambda(u). \tag{12b}$$

Indeed, (12b) is simply a reformulation of the tracial property (6). This property is thus equivalent to the possibility of quantizing $W_A \mapsto A$ and "dequantizing" $A \mapsto W_A$ with the *same* operator kernel Ω_λ: this is characteristic of Moyal quantization. It should be clear that the operators $\Omega_\lambda(u)$ do not depend on the particular choice of basis of $\mathscr{S}_\lambda$.

Note also that (12) is equivalent to

$$W_{\Omega_\lambda(u)}(v) = \mathrm{Tr}[\Omega_\lambda(u)\Omega_\lambda(v)] = I_\lambda(u,v). \tag{13}$$

Define now $\hat{Y}^{\lambda}_{\gamma rk}$ as the operator corresponding to the symbol $Y^{\lambda *}_{\gamma rk}$ by (12b). One has by this definition:

$$\Omega_\lambda(u) = \sum_{\gamma,r,k} \hat{Y}^\lambda_{\gamma rk}\, Y^\lambda_{\gamma rk}(u). \tag{14}$$

We may call the $\hat{Y}^\lambda_{\gamma rk}$ *generalized Wigner operators*, because they coincide with the ordinary Wigner operators[22] for the SU(2) case. Also, we have the suggestive formula

$$\langle \lambda i | \hat{Y}^\lambda_{\gamma rk} | \lambda j \rangle = \begin{pmatrix} \lambda & \bar{\lambda} & \gamma,r \\ i & j & k \end{pmatrix}. \tag{15}$$

The formulas (12)–(15) display the intrinsic simplicity and elegance of the SW-symbols, which is somewhat hidden in the coherent-state format. The corresponding formulas for the operators whose Q- and P-symbols are $Y^{\lambda *}_{\gamma rk}$ are more cumbersome; for instance, the (i,j) matrix element of the P-symbol operator is

$$\text{(some phase factor)} \sqrt{\frac{d_\gamma}{d_\lambda}} \begin{pmatrix} \lambda & \bar{\gamma},r \\ i & k \end{pmatrix} \begin{pmatrix} \lambda & \bar{\gamma},r \\ \lambda & r \end{pmatrix} \begin{pmatrix} \lambda \\ j \end{pmatrix}.$$

(An equivalent formula is given without proof in Perelomov's book.[19])

To avoid minor technical problems, the theory has been developed so far for simply connected groups; however, since any compact semisimple Lie group is isomorphic to the quotient of its simply connected covering group by a finite central subgroup, the formalism goes through for that larger class of groups; it suffices to select the appropriate subsets of representations and orbits.

To illustrate the theory, we give a fully explicit formula for Ω_λ in the SU(2) case. Here, $\lambda = j$ is a non-negative half-integer, the coadjoint orbits for $j > 0$ are spheres $\mathbf{S}^2$; and we adopt the conventional normalization of the spherical area measure: $\int_{\mathbf{S}^2} d\mathbf{n} = 4\pi$, in order to use the *ordinary* spherical harmonics Y_{lm} in the formula. Thus we have, for $j > 0$:

$$\Omega_j(\mathbf{n}) = \sum_{r,s=-j}^{j} Z^j_{sr}(\mathbf{n}) |jr\rangle\langle js|,$$

where

$$Z^j_{sr}(\mathbf{n}) = \frac{\sqrt{4\pi}}{2j+1} \sum_{l=0}^{2j} \sqrt{2l+1} \begin{pmatrix} j & l & j \\ r & s-r & s \end{pmatrix} Y_{l,s-r}(\mathbf{n}) \tag{16}$$

are the matrix elements of the quantizer.

IV. THE SCALAR FOURIER TRANSFORM FOR COMPACT SEMISIMPLE GROUPS

The Stratonovich–Weyl quantizer (11) provides a royal road to Fourier analysis for compact connected Lie groups as a function-space calculus. One wishes to consider the Fourier transform of a function on the group G as a function on some "dual" manifold $\hat{G}$ with appropriate symmetry properties. First, suppose that G is semisimple. We declare $\hat{G}$ to be the union of the *integral coadjoint orbits*, which is parametrized by (λ,u), with λ a dominant integral weight of G and $u \in \mathcal{O}_\lambda$. (Note that if one adopts the correspondence $\rho_\lambda \leftrightarrow \mathcal{O}_{\lambda+\delta}$, then $\hat{G}$ is alternatively given by $\Lambda \times F$, where Λ is the set of all dominant integral weights and F is the flag manifold G/T.)

The *twisted product* or *Moyal product* of two functions on $\hat{G}$ is defined by

$$(f \times h)(\lambda,u) := \int_{\mathcal{O}_\lambda} \int_{\mathcal{O}_\lambda} \mathrm{Tr}[\Omega_\lambda(u)\Omega_\lambda(v)\Omega_\lambda(w)] \times f(\lambda,v) h(\lambda,w) dm_\lambda(v) dm_\lambda(w).$$

The tracial property (6) of the Stratonovich–Weyl symbols is equivalent to the *tracial identity* for the twisted product:

$$\int_{\mathcal{O}_\lambda} (f \times h)(\lambda,u) dm_\lambda(u) = \int_{\mathcal{O}_\lambda} f(\lambda,u) h(\lambda,u) dm_\lambda(u), \tag{17}$$

since the left-hand side equals

$$\begin{aligned} &\int_{\mathcal{O}_\lambda} \int_{\mathcal{O}_\lambda} f(\lambda,v) h(\lambda,w) \int_{\mathcal{O}_\lambda} W_{\Omega_\lambda(v)\Omega_\lambda(w)}(u) \\ &\quad \times dm_\lambda(u) dm_\lambda(v) dm_\lambda(w) \\ &= \int_{\mathcal{O}_\lambda} \int_{\mathcal{O}_\lambda} f(\lambda,v) h(\lambda,w) \mathrm{Tr}[\Omega_\lambda(v)\Omega_\lambda(w)] \\ &\quad \times dm_\lambda(v) dm_\lambda(w) \\ &= \int_{\mathcal{O}_\lambda} f(\lambda,v) h(\lambda,v) dm_\lambda(v) \end{aligned}$$

on using (13).

Remark: Note that the restriction of $f \times h$ to each orbit $\mathcal{O}_\lambda$ belongs to the finite-dimensional space $\mathcal{S}_\lambda$, even though f, h may be arbitrary integrable functions on each orbit; we therefore lose little by restricting the twisted product to the space of functions on $\hat{G}$ which belong to $\mathcal{S}_\lambda$ on each orbit $\mathcal{O}_\lambda$. The definition corresponds to demanding that $W_A \times W_B = W_{AB}$ for $A, B \in \mathcal{L}(V_\lambda)$. In view of (5) the Moyal bracket $-i(f \times g - g \times f)$ and the Poisson bracket of a general element of $\mathcal{S}_\lambda$ and a symbol of the type $d\rho_\lambda(\mathfrak{g})$ are proportional, so the Moyal product is certainly a deformation product; the proportionality factor, however, may vary between orbits.

The covariance property (5) of the Stratonovich–Weyl symbols yields *equivariance* of the twisted product: if $f^g(\lambda,u) := f(\lambda, g^{-1} \cdot u)$, then $(f \times h)^g = f^g \times h^g$. This occurs because the twisted product of functions on the phase space $\hat{G}$ is the Fourier image of the convolution of functions on the group. This is also true for the "ordinary" Moyal product where G is the Heisenberg group and the Fourier transform is the familiar one.[23] For compact semisimple groups, we must first explain what the Fourier transform is; indeed, what happens is that the natural link between Moyal quantization and harmonic analysis determines the explicit form of the Fourier transform for such groups.

The *Fourier kernel* is defined as the function on $G \times \hat{G}$ given by

$$E(g;\lambda,u) := \mathrm{Tr}[\rho_\lambda(g)\Omega_\lambda(u)] = W_{\rho_\lambda(g)}(u). \tag{18}$$

This represents a departure from Wildberger's original suggestion[3] to use $Q_{\rho_\lambda(g)}(u)$ as the Fourier kernel; what distinguishes (18) from this kernel is the tracial property (6), which allows us to obtain a Fourier inversion theorem.

Definition 3: If $f \in L^1(G)$, we define its Fourier transform $\mathcal{F}f$ on $\hat{G}$ by

$$\mathcal{F}f(\lambda,u) := \int_G E(g;\lambda,u) f(g) dg.$$

We proceed somewhat formally in what follows.

Let us write $e_g(\lambda,u) := E(g;\lambda,u)$; then, applying (12b) twice, we obtain

$$
\begin{aligned}
&(e_g \times e_k)(\lambda,u) \\
&\quad = \int_{\mathcal{O}_\lambda} W_{\rho_\lambda(g)}(v) \int_{\mathcal{O}_\lambda} W_{\rho_\lambda(k)}(w) \\
&\qquad \times \mathrm{Tr}[\Omega_\lambda(u)\Omega_\lambda(v)\Omega_\lambda(w)]\, dm_\lambda(w)\, dm_\lambda(v) \\
&\quad = \mathrm{Tr}[\rho_\lambda(g)\rho_\lambda(k)\Omega_\lambda(u)] = e_{gk}(\lambda,u).
\end{aligned}
$$

From this it follows that $\mathcal{F}(f_1 * f_2) = \mathcal{F}f_1 \times \mathcal{F}f_2$ for $f_1, f_2 \in L^1(G)$, since

$$
\begin{aligned}
&\mathcal{F}(f_1 * f_2)(\lambda,u) \\
&\quad = \int_G \int_G E(g_1 g_2;\lambda,u) f_1(g_1) f_2(g_2)\, dg_1\, dg_2 \\
&\quad = \int_G \int_G (e_{g_1} \times e_{g_2})(\lambda,u) f_1(g_1) f_2(g_2)\, dg_1\, dg_2 \\
&\quad = (\mathcal{F}f_1 \times \mathcal{F}f_2)(\lambda,u).
\end{aligned}
$$

Let χ_λ denote the character of ρ_λ.

Theorem 4: The Fourier kernel (18) has the following properties:

(i) $E^*(g;\lambda,u) = E(g^{-1};\lambda,u)$;

(ii) $\int_{\mathcal{O}_\lambda} E(g;\lambda,u)\, dm_\lambda(u) = \chi_\lambda(g)$;

(iii) $E(hgh^{-1};\lambda,u) = E(g;\lambda,h^{-1}\cdot u)$;

(iv) $$\int_G E(g;\lambda,u) E^*(g;\lambda,v)\, dg = (1/d_\lambda) I_\lambda(u,v). \quad (19)$$

Proof: One has $\Omega_\lambda(u)^\dagger = \Omega_\lambda(u)$; and so

$$
\begin{aligned}
E^*(g;\lambda,u) &= \mathrm{Tr}[\Omega_\lambda(u)^\dagger \rho_\lambda(g)^\dagger] \\
&= \mathrm{Tr}[\rho_\lambda(g^{-1})\Omega_\lambda(u)] = E(g^{-1};\lambda,u).
\end{aligned}
$$

From (12) we get $\int_{\mathcal{O}_\lambda} \Omega_\lambda(u)\, dm_\lambda(u) = I$, which gives (ii). The covariance property (iii) is immediate from (5).

To derive (iv), notice that

$$
\begin{aligned}
&\int_G \langle u|\rho_\lambda(g)|u\rangle \langle v|\rho_\lambda(g^{-1})|v\rangle\, dg \\
&\quad = \int_G \langle u|\rho_\lambda(g)|u\rangle \langle v|\rho_\lambda(g)|v\rangle^*\, dg = \frac{1}{d_\lambda} |\langle u|v\rangle|^2,
\end{aligned}
$$

and thus

$$
\begin{aligned}
&\int_G E(g,\lambda,u) E^*(g,\lambda,v)\, dg \\
&\quad = \sum_{\substack{\gamma,r,k \\ \gamma',r',k'}} \frac{1}{d_\lambda} \left\langle \begin{matrix} \lambda & \bar{\gamma},r \\ \lambda & r \end{matrix} \middle| \begin{matrix} \lambda \\ \lambda \end{matrix} \right\rangle^{-1} \left\langle \begin{matrix} \lambda & \bar{\gamma}',r' \\ \lambda & r' \end{matrix} \middle| \begin{matrix} \lambda \\ \lambda \end{matrix} \right\rangle^{-1} \int_{\mathcal{O}_\lambda} \int_{\mathcal{O}_\lambda} \\
&\qquad \times |\langle u'|v'\rangle|^2 Y^{\lambda *}_{\gamma r k}(u') Y^{\lambda}_{\gamma' r' k'}(v') \\
&\qquad \times dm_\lambda(u')\, dm_\lambda(v') Y^{\lambda}_{\gamma r k}(u) Y^{\lambda *}_{\gamma' r' k'}(v) \\
&\quad = \sum_{\gamma,r,k} \frac{1}{d_\lambda} \left\langle \begin{matrix} \lambda & \bar{\gamma},r \\ \lambda & r \end{matrix} \middle| \begin{matrix} \lambda \\ \lambda \end{matrix} \right\rangle^{-2} \\
&\qquad \times \frac{d_\lambda}{d_\gamma} \left\langle \begin{matrix} \lambda & \bar{\lambda} \\ \lambda & \bar{\lambda} \end{matrix} \middle| \begin{matrix} \gamma,r \\ r \end{matrix} \right\rangle^2 \\
&\qquad \times Y^{\lambda}_{\gamma r k}(u) Y^{\lambda *}_{\gamma r k}(v) = \frac{1}{d_\lambda} I_\lambda(u,v). \qquad \blacksquare
\end{aligned}
$$

As a consequence of Theorem 4, the Fourier transform of a central function is constant on each orbit; note also that (19) gives a representation of the quantizer as an operator integral over the group:

$$\Omega_\lambda(u) = d_\lambda \int_G E^*(g;\lambda,u) \rho_\lambda(g)\, dg.$$

Finally, we get the *Fourier inversion theorem* for compact semisimple groups.

Theorem 5:

$$f(g) = \sum_\lambda d_\lambda \int_{\mathcal{O}_\lambda} E^*(g;\lambda,u) \mathcal{F}f(\lambda,u)\, dm_\lambda(u).$$

Proof: If $g = e$, the identity for G, we find that

$$
\begin{aligned}
&\sum_\lambda d_\lambda \int_{\mathcal{O}_\lambda} \mathcal{F}f(\lambda,u)\, dm_\lambda(u) \\
&\quad = \sum_\lambda d_\lambda \int_G \int_{\mathcal{O}_\lambda} f(g) E(g;\lambda,u)\, dm_\lambda(u)\, dg \\
&\quad = \sum_\lambda d_\lambda \int_G f(g) \chi_\lambda(g)\, dg = f(e),
\end{aligned}
$$

by the Plancherel formula for compact semisimple groups.[24]

The Fourier transformation intertwines the group translation $l_g f(g') := f(g^{-1}g')$ with twisted multiplication by e_g, since

$$
\begin{aligned}
\mathcal{F}(l_g f)(\lambda,u) &= \int_G E(gk;\lambda,u) f(k)\, dk \\
&= \int_G (e_g \times e_k)(\lambda,u) f(k)\, dk \\
&= (e_g \times \mathcal{F}f)(\lambda,u).
\end{aligned}
$$

Thus

$$
\begin{aligned}
&\sum_\lambda d_\lambda \int_{\mathcal{O}_\lambda} E^*(g;\lambda,u) \mathcal{F}f(\lambda,u)\, dm_\lambda(u) \\
&\quad = \sum_\lambda d_\lambda \int_{\mathcal{O}_\lambda} e_{g^{-1}}(\lambda,u) \mathcal{F}f(\lambda,u)\, dm_\lambda(u) \\
&\quad = \sum_\lambda d_\lambda \int_{\mathcal{O}_\lambda} (e_{g^{-1}} \times \mathcal{F}f)(\lambda,u)\, dm_\lambda(u) \\
&\quad = \sum_\lambda d_\lambda \int_{\mathcal{O}_\lambda} \mathcal{F}(l_{g^{-1}} f)(\lambda,u)\, dm_\lambda(u) \\
&\quad = l_{g^{-1}} f(e) = f(g)
\end{aligned}
$$

on using the tracial identity (17). ∎

If f is of the form $h^{\#} * h$, where $h \in L^2(G)$ and $h^{\#}(g) = h^*(g^{-1})$, then $\mathcal{F}f = (\mathcal{F}h)^* \times \mathcal{F}h$ and so another application of the tracial identity (17) yields the *Parseval formula*:

$$\int_G |f(g)|^2 dg = \sum_\lambda d_\lambda \int_{\mathcal{O}_\lambda} |\mathcal{F}f(\lambda,u)|^2 dm_\lambda(u).$$

The Fourier kernel (18) is the perfect analog of an ordinary Fourier exponential; at this point the reader can see that it contains all the necessary information for a functional Fourier theory for compact groups, equivalent to the ordinary operatorial Fourier theory but in a much more concrete form. It is evident now how the theory presented here can be

coupled to ordinary Fourier series theory, yielding the functional Fourier analysis for general compact connected Lie groups.

Remark: The idea of employing the orbit method to obtain an autonomous scalar Fourier theory for compact groups is due to Wildberger.[3] However, he did not have at his disposal the "correct" symbols, so he was stopped short of theorems of the Plancherel–Parseval type. In our formulation, the need for singular integrals[2] and in general the embarrassing convergence difficulties pointed out by Helgason[1] have vanished.

For the case $G = \mathrm{SU}(2)$, the Fourier kernel is calculated explicitly as follows. We use the angle-axis parametrization of SU(2): let

$$g(\psi,\mathbf{m}) := \exp\left(-\tfrac{1}{2}i\psi(m_1\sigma_1 + m_2\sigma_2 + m_3\sigma_3)\right),$$

where $\sigma_1, \sigma_2, \sigma_3$ are the Pauli matrices. Since the diagonal matrix elements (16) of the quantizer are cylindrically symmetric, we write

$$Z^j_{kk}(\cos\theta) = \sum_{l=0}^{2j} \frac{2l+1}{2j+1} \left\langle \begin{matrix} j & l \\ k & 0 \end{matrix} \middle| \begin{matrix} j \\ k \end{matrix} \right\rangle P_l(\cos\theta),$$

where the P_l are the Legendre polynomials. With these conventions, we obtain:

$$E(g(\psi,\mathbf{m});j,\mathbf{n}) = \sum_{k=-j}^{j} e^{-ik\psi} Z^j_{kk}(\mathbf{m}\cdot\mathbf{n}). \tag{20}$$

As an example of the usefulness of the scalar Fourier theory, we note that this formula leads to an interesting expression for the matrix elements of SU(2). According to the definition (18), we have

$$E(g;j,\mathbf{n}) = \sum_{r,s=-j}^{j} \mathscr{D}^j_{rs}(g) Z^j_{rs}(\mathbf{n}).$$

Thus

$$(2j+1)\mathscr{F}(\mathscr{D}^{j*}_{sr})(j,\mathbf{n}) = Z^j_{sr}(\mathbf{n}).$$

Now the Fourier inversion formula may be applied, employing the explicit expression (20) for the Fourier kernel. Details of this derivation are given in Ref. 23. The result is

$$\mathscr{D}^j_{rs}(g(\psi,\mathbf{m})) = \sum_{l=0}^{2j} \frac{\sqrt{2l+1}}{2j+1} \left\langle \begin{matrix} j & l \\ r & s-r \end{matrix} \middle| \begin{matrix} j \\ s \end{matrix} \right\rangle \times \sum_{k=-j}^{j} \left\langle \begin{matrix} j & l \\ k & 0 \end{matrix} \middle| \begin{matrix} j \\ k \end{matrix} \right\rangle e^{-ik\psi}\sqrt{4\pi}\, Y_{l,s-r}(\mathbf{m}).$$

This new form of the representative functions $\mathscr{D}^j_{rs}$ uses what is perhaps the most natural parametrization of the group SU(2). It is clear that analogous formulas may be derived for other compact connected groups, parametrized by $T \times G/T$, where T is a maximal torus.

ACKNOWLEDGMENTS

HF and JCV wish to thank the Ministerio de Educación y Ciencia of Spain for support while visiting the Universities of Zaragoza and Valladolid, during the writing of this paper. Support from the Vicerrectoría de Investigación of the Universidad de Costa Rica is acknowledged.

[1] S. Helgason, "The Fourier Transform on Symmetric Spaces," in *Élie Cartan et les Mathématiques d'aujourd'hui*, Astérisque, hors série, 151 (1985).
[2] T. Sherman, Trans. Am. Math. Soc. **209**, 1 (1975); Bull. Am. Math. Soc. **83**, 378 (1977).
[3] N. Wildberger, "On the Fourier Transform of a Compact Semisimple Lie Group," preprint, Yale University, 1986.
[4] J. E. Moyal, Proc. Cambridge Philos. Soc. **45**, 99 (1949).
[5] E. P. Wigner, Phys. Rev. **40**, 749 (1932).
[6] H. Weyl, *Gruppentheorie und Quantenmechanik* (Hirzel, Leipzig, 1928).
[7] F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer, Ann. Phys. (NY) **111**, 61, 111 (1978).
[8] M. Rieffel, Commun. Math. Phys. **122**, 531 (1989).
[9] J. C. Várilly and J. M. Gracia-Bondía, Ann. Phys. (NY) **190**, 107 (1989).
[10] J. M. Gracia-Bondía and J. C. Várilly, J. Phys. A **21**, L879 (1988).
[11] J. F. Cariñena, J. M. Gracia-Bondía, and J. C. Várilly, "Relativistic Quantum Kinematics in the Moyal Representation," J. Phys. A **23**, 901 (1990).
[12] F. Berezin, Commun. Math. Phys. **40**, 153 (1975).
[13] A. A. Kirillov, *Elements of the Theory of Representations* (Springer, Berlin, 1976).
[14] T. Broecker and T. tom Dieck, *Representations of Compact Lie Groups* (Springer, Berlin, 1985).
[15] B. Kostant, in *Lectures in Modern Analysis and Applications III*, Lecture Notes in Mathematics, Vol. 170 (Springer, Berlin, 1970).
[16] J.-P. Serre, "Représentations Linéaires et Espaces Homogénes Kählériens des Groupes de Lie Compacts," Séminaire Bourbaki, Exposé 100 (1954).
[17] N. R. Wallach, *Harmonic Analysis on Homogeneous Spaces* (Dekker, New York, 1973).
[18] A. A. Kirillov, Funct. Anal. Appl. **2**, 133 (1968).
[19] A. M. Perelomov, *Generalized Coherent States and Applications* (Springer, Berlin, 1986).
[20] R. L. Stratonovich, Sov. Phys. JETP **4**, 891 (1957).
[21] C. Fronsdal, Rep. Math. Phys. **15**, 111 (1978).
[22] L. C. Biedenharn and J. D. Louck, *Angular Momentum in Quantum Physics: Theory and Applications* (Addison-Wesley, Reading, MA, 1981).
[23] J. C. Várilly, J. M. Gracia-Bondía, and W. Schempp, "The Moyal Representation of Quantum Mechanics and Special Function Theory," Acta Appl. Math. (1990), in press.
[24] G. Warner, *Harmonic Analysis of Semisimple Lie Groups* (Springer, Berlin, 1972), 2 vols.

Star quantization on the cylinder

Carlos Alcalde[a)]
Physique–Mathématique, Université de Bourgogne, B.P. 138, F-21004 Dijon Cedex, France

(Received 12 October 1989; accepted for publication 7 January 1990)

Star-product quantization as an autonomous quantum theory is reviewed with emphasis on its relation to path integral and Hilbert space formulations. Quantization of the usual Heisenberg group is used as an example. When phase space is the cylinder, this group is replaced by $E(2)$, the Euclidean group in two dimensions. This case is worked out in detail. In doing so, polarizations on the cylinder are discussed and a phase space path integral is constructed.

I. INTRODUCTION

In most situations, a quantum theory comes about by first considering an analog classical theory. That is to say we start with a classical description of a physical system, a Hamiltonian formulation in phase space, for instance. Physicists commonly use one of two procedures to quantize such a system. On the one hand, one can use Hilbert space methods. In the most elementary prescription, one assigns to functions in phase space operators acting on a Hilbert space and one sets an eigenvalue problem for these operators to find the spectrum of physical observables. The most mathematical formulation of this quantization procedure is perhaps given in the geometric quantization program.[1] However, this program encounters difficulties when quantizing nonlinear phase functions.[2] Another formulation of quantum mechanics is in terms of path integrals first developed by Feynman.[3] In its original form one considers an exponential of the classical action and constructs the quantum mechanical matrix elements directly by summing this quantity "over all paths."[4] Many of the nuances of this approach come from defining a measure for such a sum. Although at first regarded as a clumsy reformulation, Feynman path integrals are now commonly used in quantum field theory.[5]

The moral here is that one should always welcome new methods. In this spirit, this article uses yet another formulation of quantum theory. Quantization via star-products was introduced in Ref. 6 as an autonomous approach to quantum mechanics. Instead of dealing with operators, star-quantization deals directly with phase space functions and an h-deformed associative product: the star-product. It has been shown that it leads to an independent calculation of physical spectra of elementary systems (harmonic oscillator, angular momentum, hydrogen atom).[6] Here we would like to work out the whole program when phase space is the cylinder. Quantization on the cylinder is an interesting laboratory to relearn quantum mechanics. Formal transcriptions from conventional quantization of the plane fail if one does not take into account the periodicity of the configuration space coordinate. This of course will imply the discretization of the spectrum of the momentum observable, which in turn will come back to haunt us if we try to build a phase space path integral by the conventional method. Following the star-product philosophy,[6,7] one would like to see how these known phenomena appear in terms of phase space functions only. It is also interesting to see how star-quantization relates to more conventional formulations of the problem. The basic ingredient in our approach is the utilization of $e(2)$ as the fundamental quantum algebra. It has been noted[7,8] that, on the cylinder, $e(2)$ is the appropriate algebra of preferred observables. Phase space can then be regarded as an orbit of the group $E(2)$ on the dual of its Lie algebra $e(2)^*$.

The paper is organized as follows. Section II reviews some of the relevant mathematial objects used in star-quantization. It can be used as a reference for the following sections. Section III gives the quantization of the Heisenberg group in the star-product language. The idea here is to make clear by the example of this well-known case some of the methods of star-quantization. Rather than stress the autonomous approach of star-quantization, we seek to emphasize its relation with the Hilbert space and path integral methods. Section IV constitutes the bulk of this work. In Sec. IV A the group parametrization to be used is fixed and some formulae are gathered in these coordinates for later use. Next in IV B the so-called star-exponential of the Lie algebra generators (see Sec. II) is constructed. The eigenprojectors and spectrum for the case of the Hamiltonian of the free particle are given as an example of an autonomous quantum calculation in this approach. An adapted Fourier transform is introduced in IV C inspired by the concept of a star-representation [Eq. (2.7)]. It gives the (quantum) phase space functions to be considered. The star-product of two such functions is given in IV D as a prelude to IV E where a phase space path integral is constructed in the cylinder. The connection with Hilbert space methods is given in IV F via the concept of star-polarization (see Sec. II). Both position and momentum wave-functions can be found in this approach. A relation between operators and phase space functions (Weyl–Wigner correspondence) can then be sought. This relation can then shed light on the independent construction of the path integral.[9] These ideas are confirmed in IV G. Finally, the interesting case of the universal cover of $E(2)$ is considered in IV H. Here,the so-called theta sectors and "winding numbers" appear in a natural fashion.

II. GENERAL REMARKS ABOUT STAR-PRODUCTS

It has been shown in Ref. 6 that quantization can be understood as a deformation of the algebra of observables in

[a)] Address after 1 October 1989: Department of Physics, University of California at Los Angeles, Los Angeles, California 90024-1547.

 0022-2488/90/112672-11$03.00

phase space. This deformation is given by defining a star-product, that is to say a new product law (associative but in general noncommutative) between functions in phase space. It is useful to introduce the notion of a $\mathcal{G}$-invariant star-product. Given a phase space W, consider a Lie algebra $\mathcal{G}$ of functions in W with the Poisson bracket. A $\mathcal{G}$-invariant star product satisfies

$$a*f-f*a=i\hbar\{a,f\}, \tag{2.1}$$

for all $a\in\mathcal{G}$, $f\in C^\infty(W)$, where $\{\ ,\ \}$ denotes the Poisson bracket in the symplectic manifold W. The physical meaning of this invariance is the selection of certain observables whose geometric meaning is preserved by quantization. In other words, their equations of motion will be quasiclassical, nonanomalous by construction.

A basic tool of the theory is given by the star-exponential of a given Hamiltonian H:

$$\exp(tH)=\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{t}{i\hbar}\right)^n(H^*)^n. \tag{2.2}$$

When the series converge to a distribution $\exp(tH)$ on W and has a Fourier expansion of the form

$$\exp(tH)=\sum_{\lambda\in I}\pi_\lambda e^{-i/\hbar t\lambda}, \tag{2.3}$$

we can call I (a sequence in $\mathbb{C}$) the spectrum of H, λ an eigenvalue of H, and π_λ the projector associated with λ. In general, the spectrum will be the support of the Fourier transform (in t) of $\exp(tH)$. Concrete examples of this have been given.[6,7] With the aid of exp, the time development of any function of phase space can be given:

$$f(t)=\exp(tH)*f*\exp(-tH). \tag{2.4}$$

Here H denotes the Hamiltonian function. It is easily seen that $f(t)$ obeys

$$\frac{df}{dt}=f*H-H*f, \tag{2.5}$$

which is the full quantum equation of motion in this formalism.

The star-exponential has interesting properties with respect to functions belonging to the Lie algebra $\mathcal{G}$ introduced above. Let G denote the Lie group associated with $\mathcal{G}$ by the exponential map. If three elements $a,b,c\in\mathcal{G}$ are related by the Baker–Campbell–Hausdorff formula $e^a\cdot e^b=e^c$ we have

$$\exp(a)*\exp(b)=\exp(c). \tag{2.6}$$

By defining $E(e^a)=\exp(a)$ we can regard $\exp(a)$ as a function on the group G. This leads to the following.

Definition: A star-representation of G on W is a distribution $\mathcal{E}$ on G with values in $C^\infty(W)$ such that:

(i) The domain D and the kernel of $\mathcal{E}$ are closed under the convolution in the test function space.

(ii) $\mathcal{E}$ is ad_g invariant in the sense that, for every $a\in\mathcal{G}$ $\mathcal{E}\cdot\mathrm{ad}_g(a)=\mathrm{ad}^*_g(a)\mathcal{E}$.

If we have a function $E(g)$ we can try to define a star-representation $\mathcal{E}$ by

$$\mathcal{E}(f)=\int_G f(g)E(g^{-1})dg. \tag{2.7}$$

If conditions (i) and (ii) hold, we arrive at an explicit expression for the g-invariant star-product; namely,

$$\mathcal{E}(f_1\circ f_2)=\mathcal{E}(f_1)*\mathcal{E}(f_2), \tag{2.8}$$

where on the left-hand side we have the convolution of f_1 and f_2 in D. The reason that this gives us a star-product in phase space is that when the Lie algebra g is chosen appropriately, we take for phase space an orbit of G on the dual of its Lie algebra $\mathcal{G}^*$. Since the elements of $\mathcal{G}$ can be considered as the restrictions of linear functions on $\mathcal{G}^*$, we can consider $\exp(a)$ as a function on W.

It is seen, therefore, that the construction of the function exp is a way to define the star-product. This is of practical importance as the star-exponential of the Lie algebra elements can be found directly by solving differential equations. This can be seen as follows. Let $Q=g_{AB}L^AL^B$ be an element of the center of the enveloping algebra of $\mathcal{G}$. Because W can be thought of as an orbit of G on $\mathcal{G}^*$, Q_W (Q restriction to the orbit) is a constant function on W. Let $l(a)$ and $r(a)$ denote the vector fields associated with left and right translations on G. One finds by differentiation of (2.6) that

$$a*\exp=-i\hbar\, l(a)\exp, \tag{2.9}$$

$$\exp*a=i\hbar\, r(a)\exp. \tag{2.10}$$

It follows[7] that

$$g_{AB}l(L^A)\,l(L^B)\exp=-Q_W\exp. \tag{2.11}$$

The eigendistributions of the group satisfy[10] the same equations. Here, however, we look for solutions that are ad_g invariant in the sense of (ii).

One can establish the correspondence of this formalism with conventional Hilbert space quantum mechanics via star-polarization.[11,7] As the name suggests, this method is a generalization of the analogous polarization[1] used in geometric quantization (Ref. 12 provides an introduction for physicists). Here it works as follows.

Let $\mathcal{F}$ be a topological subspace of $C^\infty(W)$. Note that $\mathcal{F}$ is said to be a star-polarization if it consists of $f\in C^\infty(W)$ such that

$$f*a=\lambda(a)f,\quad a\in g_0, \tag{2.12}$$

where g_0 is a subalgebra of $\mathcal{G}$ and $a\to\lambda(a)$ is a character of g_0. Associativity of the star-product guarantees the stability of $\mathcal{F}$ under the transformation $f\to b*f$, $b\in C^\infty(W)$. This stability seems to provide a larger scope of applicability than polarizations in the sense of geometric quantization. It should be stressed, however, that star-quantization is already a complete quantum theory unlike the stage of prequantization in the geometric quantization program. By circumventing Hilbert space methods, we can in fact make a direct connection with quantization via phase space path integrals.[13]

III. QUANTIZATION FOR THE HEISENBERG GROUP

One would like to see how star-quantization works in the simplest example: quantization in the sense of Dirac. Consider therefore,the one-dimensional Heisenberg algebra $\mathcal{H}$ and seek to construct an $\mathcal{H}$-invariant star-product. Let a

basis for $\mathcal{h}$ be (e,q,p). Consider an element of the Heisenberg group H defined via the exponential map

$$u(x_0,x_1,x_2) = e^{x_0 e + x_1 q + x_2 p}.$$

One can utilize the BCH formula to obtain the well-known group law

$$u(x_0,x_1,x_2)\cdot u(y_0,y_1,y_2) = u\left(x_0 + y_0 + \tfrac{1}{2}(x_1y_2 - x_2y_1), x_1 + y_1, x_2 + y_2\right). \quad (3.1)$$

In this parametrization, the vector field $l(y)$ and $r(y)$ corresponding to left and right translations in the group manifold are calculated to be

$$-l(y) = y^A\partial_A + \tfrac{1}{2}(y\wedge x)\partial_0, \quad (3.2)$$

$$r(y) = y^A\partial_A - \tfrac{1}{2}(y\wedge x)\partial_0, \quad (3.3)$$

where

$$\partial_A = \frac{\partial}{\partial x_A}, \quad A = 0,1,2, \quad \text{and} \quad a\wedge b = a_1b_2 - a_2b_1.$$

The orbits of H on the dual of its Lie algebra $\mathcal{h}^*$ are determined by fixing the value of e. We obtain therefore conventional phase space with coordinates $q\cdot p$ as an orbit.[14] The differential equation for exp is therefore given by [see Eq. (2.11)]

$$l(e)\exp = \frac{\partial}{\partial x_0}\exp = -\frac{i}{\hbar}\exp, \quad (3.4)$$

and its most general $\mathcal{h}$-invariant solution given by

$$\begin{aligned} \exp &= f(x_1,x_2)e^{-i/\hbar(x_0e + x_1q + x_2p)}, \\ \exp_\xi(\alpha) &= f(x_1,x_2)e^{-i/\hbar(x\cdot\alpha)}. \end{aligned} \quad (3.5)$$

The arbitrariness of the function $f(x_1,x_2)$ reflects the possible choice of orderings that can be considered. Here we will only consider the case $f = 1$ that corresponds to Weyl (symmetric) ordering.

With the aid of the function exp consider the following adapted Fourier transform [see Eq. (2.7)] given formally as

$$f(\xi) = \int \tilde{f}(\alpha)\exp_\xi(\alpha)d\alpha, \quad (3.6)$$

and explicitly as

$$f(e,q,p) = \int_{R^3} dx_0\,dx_1\,dx_2\,\tilde{f}(x_0,x_1,x_2) \times e^{-i/\hbar(x_0e + x_1q + x_2p)}. \quad (3.7)$$

On the orbit (phase space) we will be able to write simply

$$f(q,p) = \int_{\mathbf{R}^2} dx_1\,dx_2\,\tilde{f}(x_1,x_2)e^{-i/\hbar(x_1q + x_2p)}. \quad (3.8)$$

We are now ready to apply relations (2.9) and (2.10) to the full adapted Fourier transform (3.7) and induce the star-product of q and p with phase space functions of the form (3.8). We have

$$\begin{aligned} q*f(q,p) &= \left(q + \frac{i\hbar}{2}\frac{\partial}{\partial p}\right)f(q,p), \\ f(q,p)*q &= \left(q - \frac{i\hbar}{2}\frac{\partial}{\partial p}\right)f(q,p), \\ p*f(q,p) &= \left(p - \frac{i\hbar}{2}\frac{\partial}{\partial q}\right)f(q,p), \\ f(q,p)*p &= \left(p + \frac{i\hbar}{2}\frac{\partial}{\partial q}\right)f(q,p). \end{aligned} \quad (3.9)$$

The results correspond in the operator language to the multiplication of $\hat{Q}$ and $\hat{P}$ with an arbitrary operator $\hat{F}$ as given in Ref. 9.

Our next task is to define a star product of two arbitrary phase space functions. Formally the result is given by [see Eqs. (2.6) and (3.8)]

$$(f*g)(\xi) = \int d\alpha \int d\beta\,\tilde{f}(\alpha)\tilde{g}(\beta)\exp_\xi(\gamma), \quad (3.10)$$

where γ are parameters given by the BCH formula $e^\alpha\cdot e^\beta = e^\gamma$. Here we have [Eq. (3.1)]

$$\begin{aligned} \gamma_0 &= \alpha_0 + \beta_0 + \tfrac{1}{2}(\alpha_1\beta_2 - \beta_1\alpha_2), \\ \gamma_1 &= \alpha_1 + \beta_1, \\ \gamma_2 &= \alpha_2 + \beta_2. \end{aligned}$$

For this formula to be of practical use however, we need to find the inverse of (3.7) for input into (3.10). In this case it is easy because the adapted Fourier transform is just the regular Fourier transform,

$$\tilde{f}(\alpha) = \left(\frac{1}{2\pi\hbar}\right)\int_{\mathbf{R}^3} d^3\xi\,f(\xi)e^{i/\hbar(\xi\cdot\alpha)}. \quad (3.11)$$

Once we have only phase space functions (via $e = 1$) in Eq. (3.10) it is easy to calculate the formula

$$(f*g)(q,p) = \int_{\mathbf{R}^4} \frac{d^2\xi'\,d^2\xi''}{(\pi\hbar)^2} f(\xi')g(\xi'')e^{2i/\hbar\Gamma(\xi,\xi',\xi'')}, \quad (3.12)$$

where $\xi = (q,p)$ and

$$\Gamma(\xi,\xi',\xi'') = \begin{vmatrix} 1 & 1 & 1 \\ q' & q'' & q \\ p' & p'' & p \end{vmatrix}.$$

One can also make an $\hbar$-expansion of this integral and verify that

$$(f*g)(q,p) = fg + \frac{i\hbar}{2}\{f,g\} + \sum_{n=2}^{\infty}\left(\frac{i\hbar}{2}\right)^n P^n(f,g), \quad (3.13)$$

where P^n denotes the nth-order bidifferential Poisson operator.[6] Because $P^2(f,g) = P^2(g,f)$ this star-product has an added symmetry: it is not only h-invariant but it is also invariant for the algebra for all polynomials of order not greater than two. It will give the midpoint prescription to the path integral that we turn to next.

Following Ref. 13, we define for each $f(q,p)$, a "kernel" $k_f(x,y)$ by

$$k_f(x,y) = \int \frac{dp}{2\pi\hbar} f\left(p, \frac{x+y}{2}\right)e^{ip/\hbar(x-y)}. \quad (3.14)$$

It follows that

$$\int dz\,k_f(x,z)k_g(z,y) = k_{f*g}(x,y). \quad (3.15)$$

In our language this semigroup property is anticipated from the definition of the star-product by group convolution [Eq. (2.8)]. Computation of the path integral is convention-

al. One takes a partition of the time interval of interest: $0<t_1<\cdots<t_{n-1}<t$ and a family of functions $f_k(q,p)=e^{i/\hbar(t_k-t_{k-1})H}$. One can then apply relations (3.12) and (3.15) N times. In the limit as $N\to\infty$ one finds:[13]

$$\int\prod_{\tau}\frac{dq(\tau)\,dp(\tau)}{2\pi\hbar}\exp\left\{\frac{i}{\hbar}\int_0^t\left(p\cdot\frac{dq}{dt}-H\right)dt\right\}$$
$$=\int_{\mathbf{R}}\frac{dp}{2\pi\hbar}\exp\left\{tH\left(\frac{x+y}{2},p\right)\right\}e^{i/\hbar p(x-y)},\tag{3.16}$$

where the left-hand side represents the Feynman phase space path integral and the right-hand side is just a modified Fourier transform of the star-exponential of the Hamiltonian

$$\exp(tH)=\sum_{n=0}^{\infty}\left(\frac{t}{i\hbar}\right)^n\frac{1}{n!}(H^*)^n.$$

A relation between the phase space path integral and the exp function is not surprising as both are related to the time evolution of a quantum system. One can make this relation more apparent by considering the correspondence of star quantization with conventional wave mechanics. Consider the following star-polarization (other polarizations are also possible:[11]

$$\phi\in C^\infty(\mathbf{R}^2)\quad\text{such that } \phi^*q=0.$$

The function ϕ has the form

$$\phi(p,q)=e^{-2i/\hbar pq}\psi(q).\tag{3.17}$$

One can then define operators $\hat{Q}$ and $\hat{P}$ by

$$\begin{aligned}q^*\phi&=e^{-2i/\hbar pq}(\hat{Q}\Psi)(q),\\ p^*\phi&=e^{-2i/\hbar pq}(\hat{P}\Psi)(q)\end{aligned},\tag{3.18}$$

where after normalizing the action of $\hat{Q}$ and $\hat{P}$ is

$$\begin{aligned}(\hat{Q}\Psi)(x)&=x\Psi(x)=\langle x|\hat{Q}\,|\Psi\rangle,\\ (\hat{P}\Psi)(x)&=i\hbar\frac{\partial}{\partial x}\Psi(x)=\langle x|\hat{P}\,|\Psi\rangle.\end{aligned}\tag{3.19}$$

By relating operators $\hat{F}$ to their kernels $\langle x|\hat{F}\,|y\rangle$ on the one hand and by relating the same operator to a phase space function $f(q,p)$ on the other, one can attempt to find a relation between $f(q,p)$ and $\langle x|\hat{F}\,|y\rangle$. This method was developed by Berezin in Ref. 9 where he uses Eqs. (3.9) and (3.19) as his starting point. The result is

$$\langle x|\hat{F}\,|y\rangle=\int\frac{dp}{2\pi\hbar}f\left(p,\frac{x+y}{2}\right)e^{ip/\hbar(x-y)},\tag{3.20}$$

where we find that $\langle x|\hat{F}\,|y\rangle$ is nothing but our $k_f(x,y)$ of Eq. (3.13). In particular

$$\langle x|e^{i\hat{H}t}\,|y\rangle=\int\frac{dp}{2\pi\hbar}\exp\left\{tH\left(\frac{x+y}{2},p\right)\right\}e^{i/\hbar p(x-y)}.$$

Since the left-hand side can be written as a phase space path integral, we arrive again at Eq. (3.15).

IV. QUANTIZATION FOR THE $E(2)$ GROUP

A. Preliminaries

One would like to repeat the previous quantization program for the $E(2)$ group, the group of inhomogeneous rotations of the plane. As discussed previously,[7,8] this group is the appropriate one to consider in the case of a particle on S^1. Note that $E(2)$ is a connected Lie group, the semidirect product of SO(2) and $\mathbf{R}^2$. In terms of 3×3 matrices $E(2)$ can be realized in the form

$$\begin{bmatrix}\Lambda(t) & \mathbf{a}\\ 0 & 1\end{bmatrix}\in E(2),\tag{4.1}$$

where

$$\Lambda(t)=\begin{Bmatrix}\cos t & \sin t\\ -\sin t & \cos t\end{Bmatrix}\in\mathrm{SO}(2)\quad\text{and}\quad\mathbf{a}\in\mathbf{R}^2.\tag{4.2}$$

A natural basis for the $e(2)$ Lie algebra is given by the matrices L_0,P_1,P_2:

$$L_0=\begin{bmatrix}0 & 1 & 0\\ -1 & 0 & 0\\ 0 & 0 & 0\end{bmatrix},$$
$$P_1=\begin{bmatrix}0 & 0 & 1\\ 0 & 0 & 0\\ 0 & 0 & 0\end{bmatrix},\quad P_2=\begin{bmatrix}0 & 0 & 0\\ 0 & 0 & 1\\ 0 & 0 & 0\end{bmatrix},\tag{4.3}$$

with commutation relations

$$[L_0,P_1]=P_2,\quad[L_0,P_2]=-P_1,\quad[P_1,P_2]=0.\tag{4.4}$$

We want to consider group elements in the image of the exponential map:

$$e^{z_0L_0+\mathbf{z}\cdot\mathbf{P}}=\sum_{n=0}^{\infty}\frac{1}{n!}(z_0L_0+\mathbf{z}\cdot\mathbf{P})^n\in E(2).\tag{4.5}$$

By changing coordinates, we can write one such group element as

$$U(\mathbf{x},t)=\begin{bmatrix}\Lambda(t) & \Lambda(t/2)\mathbf{x}\\ 0 & 1\end{bmatrix},\tag{4.6}$$

where $t=z_0$,

$$\mathbf{x}=2\sin t/2\,\mathbf{z}.\tag{4.7}$$

In this parametrization we have

$$U^{-1}(\mathbf{x},t)=U(-\mathbf{x},-t),\tag{4.8}$$

while the group law (BCH formula) is

$$U(\mathbf{x},\alpha)\cdot U(\mathbf{y},\beta)=U(\Lambda(\alpha/2)\mathbf{y}+\Lambda(-\beta/2)\mathbf{x},\alpha+\beta),\tag{4.9}$$

and the vector fields associated with left ($h\to g^{-1}h$) and right ($h\to hg$) translations on $E(2)$ are

$$\begin{aligned}-l(L_0)&=\frac{\partial}{\partial t}-\frac{1}{2}x\wedge\partial,\\ r(L_0)&=\frac{\partial}{\partial t}+\frac{1}{2}x\wedge\partial,\\ -l(P_1)&=\cos\frac{t}{2}\frac{\partial}{\partial x_1}+\sin\frac{t}{2}\frac{\partial}{\partial x_2},\\ r(P_1)&=\cos\frac{t}{2}\frac{\partial}{\partial x_1}-\sin\frac{t}{2}\frac{\partial}{\partial x_2},\\ -l(P_2)&=\cos\frac{t}{2}\frac{\partial}{\partial x_2}-\sin\frac{t}{2}\frac{\partial}{\partial x_1},\\ r(P_2)&=\cos\frac{t}{2}\frac{\partial}{\partial x_2}+\sin\frac{t}{2}\frac{\partial}{\partial x_1}\end{aligned}\tag{4.10}$$

It is not hard to show that an $E(2)$ group element can be decomposed into a pure translation times a pure rotation.

Let $(L_0,\mathbf{P})$ denote the most general element of $e(2)^*$, the dual of the $e(2)$ Lie algebra.[14] The coadjoint action of $E(2)$ on $e(2)^*$ can be described as follows. A pure translation by $\mathbf{x}$ affects only L_0:

$$\begin{aligned} &L_0 \to L_0 + \mathbf{x}\wedge\mathbf{P}, \\ &\mathbf{P}\to\mathbf{P}, \end{aligned} \tag{4.11}$$

while a pure rotation by t just rotates $\mathbf{P}$:

$$\begin{aligned} &L_0 \to L_0, \\ &\mathbf{P}\to\Lambda(t)P. \end{aligned} \tag{4.12}$$

One then concludes that the orbits of $E(2)$ on the dual of its Lie algebra $e(2)^*$ are generated by fixing $\mathbf{P}\cdot\mathbf{P}$, that is, the orbits are isomorphic to cylinders. This is of course the phase space that we are interested in: the cotangent bundle to S^1.

B. The function exp for $E(2)$[7]

The function exp should satisfy the following differential equation [see Eq. (2.11)]:

$$l(\mathbf{P})\cdot l(\mathbf{P})\exp = -(R^2/\hbar^2)\exp. \tag{4.13}$$

In the parametrization we have been using, this equation is $(R=1)$

$$\left(\frac{\partial^2}{\partial x_1^{\,2}}+\frac{\partial^2}{\partial x_2^{\,2}}\right)\exp = -\frac{1}{\hbar^2}\exp. \tag{4.14}$$

We are interested in solutions that are $e(2)$-invariant. This requirement is met only if exp is a function of the combination $(2L_0\sin(t/2)+\mathbf{x}\cdot\mathbf{P})$. The solution that corresponds to symmetric ordering is [see Eq. (2.5)]

$$\begin{aligned} \exp_\xi(\alpha) = &\cos^2(t/4)e^{-i/\hbar(2L_0\sin t/2+\mathbf{x}\cdot\mathbf{P})} \\ &+\sin^2(t/4)e^{i/\hbar(2L_0\sin t/2+\mathbf{x}\cdot\mathbf{P})}. \end{aligned} \tag{4.15}$$

Here one should remark that exp is a function of the group with coordinates defined in Eq. (4.6). In particular it has the symmetry $t\to t+2\pi$, $\mathbf{x}\to-\mathbf{x}$. This is a departure from previous work[7] where such a property was not satisfied. It becomes important when we try to implement the program described in Sec. II.

Consider the star-exponential of the compact generator $L_0(\hbar=1)$:

$$\begin{aligned} \exp(tL_0) = &\cos^2(t/4)e^{-i2L_0\sin t/2} \\ &+\sin^2(t/4)e^{i2L_0\sin t/2}. \end{aligned} \tag{4.16}$$

Recall [Eq. (2.2)] we can also write

$$\exp(tL_0)=\sum_{n=0}^{\infty}\frac{(-it)^n}{n!}(L_0{}^*)^n. \tag{4.17}$$

Therefore

$$(L_0{}^*)^n = i^n\frac{d^n}{dt^n}(\exp tL_0)\,|_{t=0}. \tag{4.18}$$

Notice $\exp(tL_0)$ is a function of period 2π, so it is natural to consider the Fourier expansion

$$\exp(tL_0)=\sum_n \Pi_n(L_0)e^{\mathrm{int}}. \tag{4.19}$$

One finds

$$\Pi_n(L_0)=(1+(n/L_0))J_{2n}(2L_0), \tag{4.20}$$

where J_n denotes the Bessel function of integer order. If we interpret (4.17) as an example of Eq. (2.3) we find the spectrum of L_0 to be the integers as expected. From (4.17) we find

$$L_0{}^*\exp(tL_0)=i\frac{\partial}{\partial t}\exp(tL_0), \tag{4.21}$$

and the following relations for the projectors Π_n hold:[6]

$$\begin{aligned} &L_0{}^*\Pi_n = n\Pi_n, \\ &L_0=\sum_n n\Pi_n, \\ &\Pi_n{}^*\Pi_m=\delta_{nm}\Pi_m. \end{aligned} \tag{4.22}$$

It is interesting to consider the star-exponential of $H=\frac{1}{2}L_0{}^2$, the Hamiltonian of the free particle. Since $L_0{}^*L_0=L_0{}^2$, we have

$$\exp(tH)=\sum_{n=0}^{\infty}\frac{1}{n!}\left(\frac{-it}{2}\right)^n(L_0{}^*)^{2n}. \tag{4.23}$$

We can use (4.18) and (4.19) to find

$$(L_0{}^*)^{2n}=\sum_m m^{2n}\Pi_m(L_0), \tag{4.24}$$

and (4.17) becomes

$$\begin{aligned} \exp(tH) &= \sum_{n=0}^{\infty}\sum_m\frac{1}{n!}\left(\frac{-it}{2}\right)^n m^{2n}\Pi_m(L_0) \\ &= \sum_m \Pi_m(L_0)e^{-im^2/2t}. \end{aligned} \tag{4.25}$$

Notice that we have found the spectrum of H without any Hilbert space methods. Skeptical readers can look at another derivation of $\exp(tH)$ given in Appendix B.

The last equation can be written in an integral form if we realize that the $\Pi_n(L_0)$ are the Fourier coefficients of $\exp(\theta L_0)$:

$$\Pi_n(L_0)=\frac{1}{2\pi}\int_0^{2\pi}d\theta\exp(\theta L_0)e^{in\theta}. \tag{4.26}$$

Equation (4.25) becomes

$$\exp(tH)=\int_0^{2\pi}\frac{d\theta}{d\pi}\Theta(\theta,t)\exp(\theta L_0), \tag{4.27}$$

where

$$\Theta(\theta,t)=\sum_n e^{in\theta}e^{-in^2/2t} \tag{4.28}$$

is the elliptic Θ-function. Equation (4.27) motivates the next section.

C. The adapted Fourier transform of $E(2)$

We want to consider a transform with respect to the function $\exp_\xi(\alpha)$ [Eqs. (2.7) and (3.6)]:

$$f(\xi)=\int\tilde{f}(\alpha)\exp_\xi(-\alpha)d\alpha, \tag{4.29}$$

with $\exp_\xi(\alpha)$ given by Eq. (4.13). Note that $\tilde{f}(\alpha)$ is a function of the $E(2)$ group but more precisely of the image of the exponential map. That is,it has the symmetry (our group coordinates are $\mathbf{x}$ and t)

$$\tilde{f}(t+2\pi,-\mathbf{x})=\tilde{f}(t,\mathbf{x}).$$

For such functions we can rewrite (4.29) as ($\hbar=1$)

$$f(L_0,\mathbf{P})=\int_0^{4\pi}dt\int_{\mathbf{R}^2}d^2x\,\tilde{f}(t,\mathbf{x})\cos^2\frac{t}{4}e^{i(2L_0\sin t/2+\mathbf{x}\cdot\mathbf{P})},\quad(4.30)$$

where $f(L_0,\mathbf{P})$ is a function on $e(2)^*$ and by restriction to the orbit $\mathbf{P}\cdot\mathbf{P}=1$ can be considered a function on T^*S^1. Let $\hat{F}$ denote the Fourier transform of $\tilde{f}$ on the x-variables:

$$\hat{F}(t,\mathbf{P})=\int_{\mathbf{R}^2}d^2x\,\tilde{f}(t,\mathbf{x})e^{i\mathbf{x}\cdot\mathbf{P}}.\quad(4.31)$$

Define functions on the orbit by

$$f(L_0,\theta)=f(L_0,P_1=\cos\theta/2,\quad P_2=\sin\theta/2).\quad(4.32)$$

With these conventions we shall consider phase space functions of the form

$$f(L_0,\theta)=\int_0^{4\pi}dt\,\hat{F}(t,\theta)\cos^2\frac{t}{4}e^{i2L_0\sin t/2}.\quad(4.33)$$

We can now use Eqs. (4.10) to find the star-product of the Lie algebra generators with phase space functions of the form (4.30) and (4.33). It is more natural here to express the star product as an integral operator. That is, by specifying the transformation of function $\hat{F}(t,\theta)$. We have the correspondence

$$\begin{aligned}
&L_0*f(L_0,\theta)\leftrightarrow i\hbar\left(\frac{\partial}{\partial\theta}-\frac{\partial}{\partial t}\right)\hat{F}(t,\theta),\\
&f(L_0,\theta)*L_0\leftrightarrow -i\hbar\left(\frac{\partial}{\partial\theta}+\frac{\partial}{\partial t}\right)\hat{F}(t,\theta),\\
&P_2*f(L_0,\theta)\leftrightarrow\sin\left(\frac{\theta-t}{2}\right)\hat{F}(t,\theta),\\
&f(L_0,\theta)*P_2\leftrightarrow\sin((\theta+t)/2)\hat{F}(t,\theta),\\
&P_1*f(L_0,\theta)\leftrightarrow\cos((\theta-t)/2)\hat{F}(t,\theta),\\
&f(L_0,\theta)*P_1\leftrightarrow\cos((\theta+t)/2)\hat{F}(t,\theta).
\end{aligned}\quad(4.34)$$

These relations will be useful in finding a correspondence between functions and operators. That is, the analog of the Weyl–Wigner correspondence.

D. A star-product of two arbitrary functions

One begins by going back to Eqs. (4.29) or (4.30) to write the formal result

$$(f*g)(\xi)=\int d\alpha\int d\beta\,\tilde{f}(\alpha)\tilde{g}(\beta)\exp_\xi(-\gamma).$$

The parameters γ can be read out of Eq. (3.9):

$$\begin{aligned}
&-\gamma_0=\alpha+\beta,\\
&-\gamma=\Lambda(-\alpha/2)\mathbf{y}+\Lambda(\beta/2)\mathbf{x}.
\end{aligned}\quad(4.35)$$

The idea here is to write $\tilde{f}$ and $\tilde{g}$ in terms of $\hat{F}$ and $\hat{G}$. For example,

$$\tilde{f}(\alpha,\mathbf{x})=\int_{\mathbf{R}^2}\frac{d^2Q}{(2\pi)^2}\hat{F}(\alpha,\mathbf{Q})e^{-i\mathbf{x}\cdot\mathbf{Q}}.\quad(4.36)$$

We can then perform the integrations in $\mathbf{x}$ and $\mathbf{y}$ space to obtain delta functions that we can again integrate. By restricting to the orbit as before, we can write our final result as

$$\begin{aligned}
(f*g)(L_0,\theta)&=\int_0^{4\pi}d\alpha\int_0^{4\pi}d\beta\,\hat{F}(\alpha,\theta+\beta)\hat{G}(\beta,\theta-\alpha)\\
&\times\cos^2\frac{\alpha+\beta}{4}e^{i2L_0\sin(\alpha+\beta)/2}.
\end{aligned}\quad(4.37)$$

Using the invariance properties of the group measure, we can shift our integration variables and write

$$(f*g)(L_0,\theta)=\int_0^{4\pi}d\alpha(\hat{F}\circ\hat{G})\alpha,\theta)\cos^2\frac{\alpha}{4}e^{i2L_0\sin\alpha/2},\quad(4.38)$$

where $\hat{F}\circ\hat{G}$ is related to the convolution of the rotation group [see Eqs. (2.7) and (2.8)]:

$$(\hat{F}\circ\hat{G})(\alpha,\theta)=\int_0^{4\pi}d\alpha\,\hat{F}(\alpha-\beta,\tau+\beta)\,\hat{G}(\beta,\theta+\beta-\alpha).\quad(4.39)$$

We would like to point out some properties of $\hat{F}$. Because $\tilde{f}$ has the symmetry $\tilde{f}(\alpha+2\pi,-\mathbf{x})=\tilde{f}(\alpha,\mathbf{x})$; $\hat{F}$ inherits

$$\hat{F}(\alpha+2\pi,\theta+2\pi)=\hat{F}(\alpha,\theta).\quad(4.40)$$

This leads us to define k_f, the kernel of f, as

$$k_f(x,y)=\hat{F}(x-y,x+y).\quad(4.41)$$

Notice that we now have periodicity in x,y independently

$$\begin{aligned}
&k_f(x+2\pi,y)=k_f(x,y),\\
&k_f(x,y+2\pi)=k_f(x,y)
\end{aligned}\quad(4.42)$$

Equations (4.38) and (4.41) imply

$$k_{f*g}((\theta+\alpha)/2,(\theta-\alpha)/2)=(\hat{F}\circ\hat{G})(\alpha,\theta)$$

so that we can establish transitivity for the kernel (after rescaling):

$$k_{f*g}(x,y)=\int_0^{2\pi}dz\,k_f(x,z)k_g(z,y).\quad(4.43)$$

Finally, one can write the star product in terms of phase space functions only. To accomplish this, one needs to find the inverse of (4.33) for input into (4.38). The inverse can be written as (see Appendix A):

$$\begin{aligned}
\hat{F}(t,\theta)=\int_{\mathbf{R}}\frac{dp}{2\pi}\,&\mathrm{sgn}\left(\cos\frac{t}{2}\right)\\
&\times\left\{f(p,\theta)\cos^2\frac{t}{4}e^{-i2p\sin t/2}\right.\\
&\left.-f(p,\theta+2\pi)\sin^2\frac{t}{4}e^{i2p\sin t/2}\right\}.
\end{aligned}\quad(4.44)$$

The final result is not too illuminating. We shall consider a special case in the following.

E. A path integral in the cylinder

Calculations simplify considerably if one takes functions even in the Lie algebra generators. That is, functions on T^*S^1 even in L_0 and of period 2π in θ. For such functions $\hat{F}(t,\theta)$ has in addition to the properties given above the following:

$$\hat{F}(t,\theta)=\hat{F}(t+2\pi,\theta)=\hat{F}(-t,\theta). \tag{4.45}$$

Equation (4.33) simplifies to

$$f(L_0,\theta)=\frac{1}{2}\int_0^{4\pi} dt\,\hat{F}(t,\theta)e^{i2L_0\sin t/2}$$
$$=\int_{-\pi}^{\pi} dt\,\hat{F}(t,\theta)e^{i2L_0\sin t/2}, \tag{4.46}$$

while its inverse [Eq. (4.44)] can be written as

$$F(t,\theta)=\left|\cos\frac{t}{2}\right|\int_{\mathbf{R}^2}\frac{dp}{2\pi}f(p,\theta)e^{-i2p\sin t/2}$$
$$=\int_{\mathbf{R}}\frac{dp}{2\pi\hbar}\left(1+\frac{\hbar^2}{4}\frac{\partial^2}{\partial p^2}\right)^{1/2}f(p,\theta)e^{-i2p/\hbar\sin t/2}, \tag{4.47}$$

where in the last equation, $\hbar$ is made explicit in order to better appreciate the quantum corrections ($\hbar$-deformations) inherent in star products. The star-product of two arbitrary functions in phase space with these properties is

$$(f*g)(L_0,\theta)=\int_{-\pi}^{\pi}d\alpha\int_{-\pi}^{\pi}d\beta\int_{\mathbf{R}}\frac{dp}{2\pi}\int\frac{dQ}{2\pi}$$
$$\times\left|\cos\frac{\alpha}{2}\cos\frac{\beta}{2}\right|f(P,\theta+\beta)g(Q,\theta-\alpha)$$
$$\times\exp\left\{i2\left[L_0\sin\frac{\alpha+\beta}{2}-P\sin\frac{\alpha}{2}\right.\right.$$
$$\left.\left.-Q\sin\frac{\beta}{2}\right]\right\}. \tag{4.48}$$

We now repeat the calculation leading to Eq. (3.16), this time utilizing equation (4.43) with

$$k_f(x,y)=\hat{F}(x-y,x+y)$$
$$=\int_{\mathbf{R}}\frac{dp}{2\pi\hbar}\left|\cos\frac{x+y}{2}\right|f(p,x+z)e^{-i2p/\hbar\sin(x-y/2)}. \tag{4.49}$$

Partition the time interval as before and consider the same family of functions: $f_k(L_0,\theta)=e^{i/\hbar(t_k-t_{k-1})H}$. After N repetitions of (4.43), we have

$$\int d\tilde{x}_1\cdots d\tilde{x}_N\int dp_1\cdots dp_{N+1}$$
$$\times\exp\left\{\frac{i}{\hbar}\sum_k\left(L_0\sin\left(\frac{x_{k+1}-x_k}{2}\right)-(t_{k+1}-t_k)H\right\}\right.$$
$$=\int\frac{dp}{2\pi\hbar}\left|\cos\frac{x-y}{2}\right|(f_N^**\cdots*f_1)e^{-i2p/\hbar\sin(x-y/2)}, \tag{4.50}$$

where $x=x(0)$, $y=x(t)$, and

$$d\tilde{x}_k=\left|\cos\left(\frac{x_{k+1}-x_k}{2}\right)\right|\frac{dx_k}{2\pi\hbar}.$$

In the limit $N\to\infty$ we can write suggestively with the precise meaning given above:

$$\int\prod_\tau dx(\tau)\,dp(\tau)\exp\left\{\frac{i}{\hbar}\int_0^t(P\dot{x}-H)d\tau\right\}$$
$$=\int_{\mathbf{R}}\frac{dp}{2\pi\hbar}\left(1+\frac{\hbar^2}{4}\frac{\partial}{\partial P^2}\right)^{1/2}$$
$$\times\exp tH(P,x+y)e^{-i2p/\hbar\sin(x-y/2)}, \tag{4.51}$$

thereby finding a phase space path integral on the cylinder intimately related to the star-exponential of the Hamiltonian albeit formally.

F. Polarizations in the cylinder

Once again we can apply the method of star-polarization to make contact with the more conventional form of quantum mechanics. Consider the following star-polarization:

$$\phi(L_0,\theta)\in C^\infty(T^*S^1)\quad\text{such that } \phi*P_2=0.$$

Using Eq. (4.34), we see ϕ is of the form

$$\phi(L_0,\theta)=\int_0^{4\pi}dt\,\delta(\theta+t)\cos^2\frac{t}{4}e^{i2L_0\sin t/2}\Psi(\theta)$$
$$=\cos^2\frac{\theta}{4}e^{-2L_0\sin\theta/2}\Psi(\theta). \tag{4.52}$$

One then defines operators $\hat{L}_0$, $\hat{S}$, $\hat{C}$ by

$$L_0*\phi(L_0,\theta)=\cos^2\frac{\theta}{4}e^{-i2L_0\sin\theta/2}(\hat{L}_0\Psi)(\theta),$$
$$P_1*\phi(L_0,\theta)=\cos^2\frac{\theta}{4}e^{-i2L_0\sin\theta/2}(\hat{C}\Psi)(\theta), \tag{4.53}$$
$$P_2*\phi(L_0,\theta)=\cos^2\frac{\theta}{4}e^{-i2L_0\sin\theta/2}(\hat{S}\Psi)(\theta).$$

Using Eq. (4.34) again we find

$$(\hat{L}_0\Psi)(\theta)=i\frac{\partial}{\partial\theta}\Psi(\theta),$$
$$(\hat{C}\Psi)(\theta)=\cos\theta\ \Psi(\theta), \tag{4.54}$$
$$(\hat{S}\Psi)(\theta)=\sin\theta\ \Psi(\theta)$$

which is the usual representation of $e(2)$ on $L^2(S^1)$. We could also take another star-polarization. Take $\phi(L_0,\theta)\in C^\infty(T^*S^1)$ such that $\phi*L_0=0$. This polarization is absent in the geometric quantization approach. Here ϕ takes the form

$$\phi(L_0,\theta)=\int_0^{4\pi}dt\,\Psi(t)\cos^2\frac{t+\theta}{4}e^{+i2L_0\sin(\theta+t)/2}$$
$$=\sum_n a_n\pi_n(L_0)e^{in\theta}, \tag{4.55}$$

where the $\pi_n(L_0)$ are the eigenprojectors introduced in Eq. (4.30) and a_n is the wave function in momentum representation:

$$a_n=\langle n|\Psi\rangle=\int_0^{2\pi}dt\,\Psi(t)e^{\text{int}}. \tag{4.56}$$

The action of operators $\hat{L}_0$, $\hat{C}$, $\hat{S}$ is induced as before, this time for action on $\Psi(t)$ or equivalently on a_n. In particular,

$$L_0 * \phi(L_0,\theta) = \sum_n \langle n|\hat{L}_0\Psi\rangle \pi_n(L_0) e^{in\theta}, \tag{4.57}$$

with

$$\langle n|\hat{L}_0|\Psi\rangle = n\langle n|\Psi\rangle. \tag{4.58}$$

It is seen therefore that via star-polarization we can "half" the phase space variables to obtain a Hilbert space formulation. We remark again that this procedure is not necessary (although it might be practical) to find eigenprojectors and spectrum.

G. Weyl–Wigner correspondence on the cylinder

In Ref. 9 Berezin calculated the Wigner–Weyl correspondence in $\mathbf{R}^2$. His starting point is the knowledge of the one-to-one relation between Heisenberg algebra elements acting on operators and their corresponding phase space function. His method in this case is as follows. If we associate $f \in C^\infty(T^*S^1)$ to the kernel of an operator $\langle x|\hat{F}|y\rangle = k(x,y)$, star-quantization tells us that we should also relate

$$\begin{aligned}
&L_0 * f \leftrightarrow i\hbar \frac{\partial}{\partial x} k(x,y),\\
&f * L_0 \leftrightarrow i\hbar \frac{\partial}{\partial y} k(x,y),\\
&p_2 * f \leftrightarrow \sin x\, k(x,y),\\
&f * p_2 \leftrightarrow \sin y\, k(x,y),\\
&p_1 * f \leftrightarrow \cos x\, k(x,y),\\
&f * p_1 \leftrightarrow \cos y\, k(x,y).
\end{aligned} \tag{4.59}$$

Comparison with Eqs. (4.34) allows us to guess

$$x = (\theta - t)/2, \quad y = (\theta + t)/2,$$

so that

$$k((\theta - t)/2, (\theta + t)/2) = \hat{F}(t,\theta),$$

and Eq. (4.33) becomes

$$f(L_0,\theta) = \int_0^{4\pi} d\alpha \left\langle \frac{\theta+\alpha}{2}|\hat{F}|\frac{\theta-\alpha}{2}\right\rangle \cos^2\frac{\alpha}{4} e^{-i2L_0 \sin\alpha/2}, \tag{4.60}$$

which is a Wigner application (operators to functions). Its inverse [given in the general case by (4.44)] is called a Weyl transformation. As a check consider the case $\hat{F} = e^{i\hat{H}t}$, where $\hat{H} = \frac{1}{2}\hat{L}_0^2$. We have

$$\left\langle \frac{\theta+\alpha}{2}|e^{i\hat{H}t}|\frac{\theta-\alpha}{2}\right\rangle = \sum_n e^{in\alpha} e^{-tn^2/2}. \tag{4.61}$$

Substituting in (4.60) and using the 2π periodicity we arrive at Eq. (4.27). That is, the phase space function corresponding to the operator $e^{i\hat{H}t}$ is precisely the star-exponential $\exp(tH)$.

One can also verify that our path integral reproduces the standard result in this case. Recall that our construction is based on the conventional time slicing plus repeated use of the superposition principle. One would like to know, therefore, the short-time behavior of our matrix elements. For example consider

$$G(x,\Delta t) = \langle x|e^{i\hat{H}\Delta t}|0\rangle.$$

By (4.49) we set this equal to

$$\int_{\mathbf{R}} \frac{dp}{2\pi\hbar}\left(1 + \frac{\hbar^2}{4}\frac{\partial^2}{\partial p^2}\right)^{1/2} \exp(\Delta t H) e^{-i2p/\hbar \sin x/2}. \tag{4.62}$$

In our formalism it easy to extract a short-term expansion. Up to terms of order $(\Delta t)^2$ we have

$$\begin{aligned}
&\left(1 + \frac{\hbar^2}{4}\frac{\partial^2}{\partial p^2}\right)^{1/2} \exp(\Delta t H)\\
&\quad = \left(1 + \frac{\hbar^2}{4}\frac{\partial^2}{\partial p^2}\right)^{1/2}\left(1 + \frac{i\Delta t}{\hbar}p^2\right)\\
&\quad = 1 + \frac{i\Delta t}{\hbar}\left(\frac{1}{2}p^2 + \frac{\hbar^2}{8}\right)\\
&\quad = \exp\left\{\frac{i\Delta t}{\hbar}\left(\frac{1}{2}p^2 + \frac{\hbar^2}{8}\right)\right\},
\end{aligned}$$

so that (4.62) becomes a Gaussian integral and we can write the final result as

$$G(x,\Delta t) = \left(\frac{1}{2\pi i\hbar\Delta t}\right)^{1/2} \exp\left\{\frac{-i}{\hbar\Delta t}(1 - \cos x) - i\Delta t\frac{\hbar^2}{8}\right\}, \tag{4.63}$$

which is the correct short-time kernel for the free particle calculated directly in Ref. 15.

H. The case of the universal cover

One would like to see the modifications in our procedure when one considers $\tilde{E}(2)$, the universal cover of $E(2)$. Note that $\tilde{E}(2)$ can be considered as the semidirect product of SO(2) (the real line) and $\mathbf{R}^2$. One can take the same group parametrization as before and seek to find an appropriate star-exponential. Here exp must still satisfy the differential equation (4.14), which as we have seen, fixes the value of the Casimir. The general $e(2)$-invariant solution (any covering) can be written as

$$\begin{aligned}
\exp_\xi(\alpha) = {} & A(t) e^{-i\hbar(2L_0 \sin t/2 + \mathbf{x}\cdot\mathbf{P})}\\
& + B(t) e^{i/\hbar(2L_0 \sin t/2 + \mathbf{x}\cdot\mathbf{P})}.
\end{aligned} \tag{4.64}$$

As in Refs. 6 and 7 we impose the following conditions on A and B:

(i) Normalization:

$$A(0) + B(0) = 1,$$
$$A'(0) = B'(0) = 0$$

(ii) Symmetry:

$$A(-t) = \overline{A(t)},$$
$$B(-t) = \overline{B(t)}.$$

For $E(2)$ itself we also impose $A(t) = B(t + 2\pi)$ and $A(t + 2\pi) = B(t)$. For $E(2)$ we can relax this condition as we no longer demand periodicity since $t \in \mathbf{R}$. We can introduce arbitrary boundary conditions via a parameter $\gamma \in (0,1)$ and modify our previous star-exponential:

$$\begin{aligned}
\exp_\xi(\alpha,\gamma) = e^{i\gamma t}\Big\{ & \cos^2\frac{t}{4} e^{-i/\hbar(2L_0 \sin t/2 + \mathbf{x}\cdot\mathbf{P})}\\
& + \sin^2\frac{t}{4} e^{i/\hbar(2L_0 \sin t/2 + \mathbf{x}\cdot\mathbf{P})}\Big\}.
\end{aligned} \tag{4.65}$$

The star-exponential of L_0 is now

$$\exp(tL_0,\gamma) = e^{i\gamma t}\left\{\cos^2\frac{t}{4}e^{-i2L_0\sin t/2} + \sin^2\frac{t}{4}e^{i2L_0\sin t/2}\right\}. \quad (4.66)$$

If we look for a Fourier expansion of the form of Eq. (2.3) we find

$$\exp(tL_0,\gamma) = \sum_n \Pi_n(L_0)e^{i(n+\gamma)t}, \quad (4.67)$$

where Π_n was given in Eq. (4.20). Remark that the spectrum is now shifted by γ. Next, consider the adapted Fourier transform of $E(2)$. That is, we consider an equation like (4.29) where the range of t is now $\mathbf{R}$:

$$f(L_0,\mathbf{P}) = \int_{\mathbf{R}} dt \int_{\mathbf{R}} d^2x \tilde{f}(t,\mathbf{x})\exp_\gamma(t,\mathbf{x}). \quad (4.68)$$

This is once again an example of Eq. (2.7). Here $\exp_\gamma(t,\mathbf{x})$ is given by Eq. (4.65). Since the star-exponential now satisfies

$$\exp_\gamma(t+2\pi,-\mathbf{x}) = e^{i2\pi\gamma}\exp_\gamma(t,\mathbf{x}), \quad (4.69)$$

we can decompose the integral t into a sum of integrals of length 2π. We have

$$f(L_0,\mathbf{P}) = \int_0^{2\pi} dt\, A(t,\mathbf{P})\exp_\gamma(t,\mathbf{x}).$$

And on the orbit [compare with Eq. (4.33)]

$$F(L_0,\theta) = \int_0^{4\pi} dt\, A(t,\theta)e^{i\gamma t}\cos^2\frac{t}{4}e^{i2L_0\sin t/2}. \quad (4.70)$$

Here

$$A(t,\theta) = \sum_n e^{i2\pi n\gamma}\hat{\tilde{F}}(t+2\pi n,\theta+2\pi n). \quad (4.71)$$

Note that $\hat{\tilde{F}}$ denotes the Fourier transform of $\tilde{f}$ on the x-variables as given in Eq. (4.41). To arrive at (4.70) use has been made of the relation

$$A(t+2\pi,\theta+2\pi) = e^{-2\pi\gamma}A(t,\theta).$$

Calculations now can be repeated as before. Polarizations now give wave-functions with twisted boundary conditions. For instance,

$$\Psi(\theta+2\pi) = e^{i2\pi\gamma}\Psi(\theta).$$

By the Weyl–Wigner correspondence one finds

$$A(t,\theta) = \left\langle\frac{\theta-t}{2}\middle|\hat{\tilde{F}}\middle|\frac{\theta+t}{2}\right\rangle,$$

so that $A(t,\theta)$ gives the matrix elements in the cylinder. It is not hard to see that $\hat{\tilde{F}}(t,\theta)$ in Eq. (4.71) is the matrix element in the plane. Equation (4.71) therefore justifies and generalizes the prescription of "sum over winding numbers."[16] Equation (4.71) was also found in Ref. 17. Here it is a consequence of Eq. (4.68) and the group manifold formulation of the adapted Fourier transform.

V. DISCUSSION

We have shown how star-products can be used to give a consistent quantum theory on the cylinder. As we have seen, in this approach it is not canonical variables but an algebra of preferred observables that plays a prominent role. Quantization is then carried solely in terms of phase space functions and the star-product. Since the star-product is associative, but in general noncommutative, the ring of functions with this product does not have an associated underlying manifold. This is consistent with the uncertainty principle. Quantization without polarizations might be useful in theories where halving the number of variables might break a symmetry such as gauge invariance. On the other hand, when polarizations are desired, they can be found by the method of star-polarization. Due to the stability of star-polarizations (see the end of Sec. II) polarizations might be found in this sense where they do not exist in the sense of geometric quantization.

The connection of star-products with path integrals while still at the formal level is a significant result. As in the case of the plane, a path integral in the cylinder is defined as an integral transform of the star-exponential of the Hamiltonian of interest. This mere fact does not make computations any easier since a direct calculation of the star-exponential of a given Hamiltonian is very involved.[6] However, knowledge of the first terms of a star-exponential might help define the path integral measure.

The $e(2)$ Lie algebra has been shown to be relevant in the quantization of some field theory models.[8] Applications of star-products to quantum field theories are therefore important. Work in this regard is in progress.

ACKNOWLEDGMENTS

The author would like to express thanks for the hospitality of the Université de Bourgogne where parts of this work were completed. He has benefited from many helpful discussions with F. Bayen, J. C. Cortet, M. Flato, C. Fronsdal, J. Jezierski, C. Martin, and D. Sternheimer.

This work is supported by the National Science Foundation under Grant No. PHY 86-13201.

APPENDIX A: THE INVERSE OF THE ADAPTED FOURIER TRANSFORM

We want to show how the inverse of (4.33) is constructed. We start by showing how (4.46) and (4.47) are inverse relations. Clearly the dependence on θ will not pose a problem. It is enough to prove the following.

Proposition: Let $u(p)\in L^2(\mathbf{R})$ such that

$$(\mathcal{F}u)(x) = \int_{\mathbf{R}} u(p)e^{-ipx}dp = 0 \quad \text{for } |x|>2. \quad (A1)$$

Then the inverse of the transform

$$u(p) = \int_{-\pi}^{\pi} dt\,\tilde{u}(t)e^{i2p\sin t/2}$$

is

$$\tilde{u}(t) = \int_{\mathbf{R}}\frac{dp}{2\pi}\left(1+\frac{1}{4}\frac{\partial^2}{\partial p^2}\right)^{1/2}u(p)e^{-i2p\sin t/2}.$$

Proof: By direct verification. On one side,

$$\int_{-\pi}^{\pi} dt\, \tilde{u}(t) e^{i2p \sin t/2}$$
$$= \int_{-\pi}^{\pi} dt \left\{ \int_{\mathbf{R}} \frac{dq}{2\pi} \left(1 + \frac{1}{4}\frac{\partial^2}{\partial q^2}\right)^{1/2} u(q) e^{-i2q \sin t/2} \right\} \times e^{i2p \sin t/2}$$
$$= \int_{-\pi}^{\pi} \frac{dt}{2\pi} \cos\frac{t}{2} \int_{\mathbf{R}} dq\, u(q) e^{-i2q \sin t/2} e^{i2p \sin t/2}.$$

Letting $x = 2 \sin t/2$,

$$\int_{-2}^{2} \frac{dx}{2\pi} \int_{\mathbf{R}} dq\, u(q) e^{iqx, ipx}$$
$$= \int_{-2}^{2} \frac{dx}{2\pi} (\mathcal{F}u)(x) e^{ipx}$$
$$= \int_{\mathbf{R}} \frac{dx}{2\pi} (\mathcal{F}u)(x) e^{ipx} = u(p).$$

On the other side,

$$\int_{\mathbf{R}} dp \left(1 + \frac{1}{4}\frac{\partial^2}{\partial p^2}\right)^{1/2} u(p) e^{-i2p \sin t/2}$$
$$= \int_{\mathbf{R}} dp \left(1 + \frac{1}{4}\frac{\partial^2}{\partial p^2}\right)^{1/2} \left\{ \int_{-\pi}^{\pi} dx\, \tilde{u}(x) e^{-i2p \sin x/2} \right\} \times e^{-i2p \sin t/2}$$
$$= \int_{\mathbf{R}} dp \int_{-\pi}^{\pi} dx\, \tilde{u}(x) \left|\cos\frac{x}{2}\right| e^{-i2p(\sin s/2 - \sin t/2)}$$
$$= \int_{-\pi}^{\pi} dx\, \tilde{u}(x) \left|\cos\frac{x}{2}\right| \delta\left(2\sin\frac{x}{2} - 2\sin\frac{t}{2}\right).$$

In this interval, the delta function picks up only one point:

$$\int_{-\pi}^{\pi} dx\, \tilde{u}(x) \delta(x - t) = \tilde{u}(t).$$

This finishes the proof. In the general case

$$f(L_0, \theta) = \int_0^{4\pi} dt\, \hat{F}(t,\theta) \cos^2\frac{t}{4} e^{i2L_0 \sin t/2}$$

and it cannot be put in the form (4.46). We can decompose f however, into

$$g(L_0,\theta) = \tfrac{1}{2}\{f(L_0,\theta) + f(-L_0, \theta + 2\pi)\},$$
$$h(L_0,\theta) = \tfrac{1}{2}\{f(L_0,\theta) - f(-L_0 + 2\pi)\}$$

And we can put $g(L_0,\theta)$ and $h(L_0,\theta)$ into the form (4.46) so we can find their inverse $\hat{G}(t,\theta)$ and $\hat{H}(t,\theta)$. It is easy to show that

$$\hat{F}(t,\theta) = \hat{G}(t,\theta) + \hat{H}(t,\theta), \tag{A2}$$

since

$$\hat{G}(t,\theta) = \int_{\mathbf{R}} dp \left|\cos\frac{t}{2}\right| g(p,\theta) e^{-i2p \sin t/2}$$

and

$$\hat{H}(t,\theta) = \int_{\mathbf{R}} dp\, \mathrm{sgn}\left(\cos\frac{t}{2}\right) h(p,\theta) e^{-i2p \sin t/2}.$$

Equation (A2) readily gives Eq. (4.44). It is interesting to consider the case of functions of L_0 only. From Eq. (4.33) we have that such functions are expressed as

$$f(L_0) = \int_0^{4\pi} dt\, \tilde{f}(t) \cos^2\frac{t}{4} e^{i2L_0 \sin t/2},$$

with $\tilde{f}(t + 2\pi) = \tilde{f}(t)$. It is not hard to show that this implies

$$f(L_0) = \int_0^{2\pi} dt\, \tilde{f}(t) \exp(tL_0),$$

with $\exp(tL_0)$ given by Eq. (4.16). Expansion of $\exp(tL_0)$ into eigenprojectors [Eq. (4.19)] leads to a Neumann series[18]

$$f(L_0) = \sum_n a_n \Pi_n(L_0),$$

where

$$a_n = \int_0^{2\pi} dt\, \tilde{f}(t) e^{int}.$$

Recall that

$$\Pi_n = (1 + (n/p)) J_{2n}(2L_0),$$

so that we will have expansion of the type

$$f(L_0) = \sum_{n=0}^{\infty} b_n J_{2n}(2L_0), \quad f \text{ even},$$
$$f(L_0) = \sum_{n=0}^{\infty} c_{2n+1} J_{2n+1}(2L_0), \quad f \text{ odd}$$

In Ref. 19 we find (A1) as a necessary and sufficient condition for convergence of such a series.

APPENDIX B: STAR-EXPONENTIAL OF $H = \frac{1}{2} L_0^2$

One can give another derivation for the star-exponential of the Hamiltonian of the free particle. We want to calculate the sum

$$\phi(L_0,t) = \sum_{n=0}^{\infty} \left(\frac{t}{i\hbar}\right)^n \frac{1}{n!} (H^*)^n \tag{B1}$$

for $H = \frac{1}{2}L_0^2$. Note that ϕ must satisfy

$$i\hbar \frac{\partial}{\partial t} \phi(L_0,t) = H^* \phi(L_0,t). \tag{B2}$$

Because ϕ is an even function of L_0, we can use (4.46) to express it as

$$\phi(L_0,t) = \frac{1}{2} \int_0^{4\pi} d\alpha\, \tilde{F}(\alpha) e^{i2L_0 \sin \alpha/2}. \tag{B3}$$

Using (4.34) we have

$$H^* \phi(L_0,t) = \frac{1}{2} \int_0^{4\pi} d\alpha - \frac{\hbar^2}{2} \frac{\partial^2}{\partial \alpha^2} \times \tilde{F}(\alpha) e^{i2L_0 \sin \alpha/2}. \tag{B4}$$

Integrating by parts and redefining ϕ so that it depends on $x = 2L_0/\hbar$, we can induce the action of H as a differential operator:

$$H^* \phi = \frac{1}{2} \frac{\hbar^2}{4} \left[x^2 \left(1 + \frac{\partial^2}{\partial x^2}\right) + x \frac{\partial}{\partial x} \right] \phi. \tag{B5}$$

Once we recognize this operator as Bessel's differential operator, separation of variables permits us to write the solution to (B2) as a linear superposition of independent functions:

$$\phi(L_0,t) = \sum_{n>0} a_n J_{2n}\left(\frac{2L_0}{\hbar}\right) e^{-in^2/2\hbar t}.$$

The condition $\exp Ht|_{t=0} = 1$ tells us that a_n should be equal to the Neumann symbol ϵ_n:

$$\epsilon_n = \begin{cases} 1 & n = 0, \\ 2 & n \neq 0. \end{cases}$$

Therefore

$$\exp(tH) = \sum_{n>0} \epsilon_n J_{2n}\left(\frac{2L_0}{\hbar}\right) e^{-in^2/2\hbar t},$$

which can be easily be shown to be equivalent to the form (4.25).

[1] B. Kostant, in *Lectures in Modern Analysis and Applications III*, Lecture Notes in Mathematics, Vol. 170 (Springer, Berlin, 1970).
[2] N. Woodhouse, *Geometric Quantization* (Oxford U. P., New York, 1980).
[3] R. Feynman, Rev. Mod. Phys. **20**, 367 (1948).
[4] R. Feynman and A. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).
[5] L. Faddeev and A. Slavnov, *Gauge Fields Introduction to Quantum Theory* (Benjamin/Cummings, New York, 1980).
[6] F. Bayen, M. Flato, C. Fronsdal, A. Lichnerowicz, and D. Sternheimer, Ann. Phys. **111**, 61, 111 (1978).
[7] C. Fronsdal, Rep. Math. Phys. **15**, 111 (1978).
[8] C. Alcalde and D. Sternheimer, Lett. Math. Phys. **17**, 117 (1989).
[9] F. Berezin, Theo. Math. Phys. **6**, 141 (1971).
[10] A. A. Kirillov, *Elements of the Theory of Representations* (Springer, Berlin, 1976).
[11] T. Huynh, Lett. Math. Phys. **4**, 201 (1980).
[12] A. Niemi, Ann. Phys. **187**, 469 (1988).
[13] P. Sharon, Phys. Rev. D **20**, 414 (1979).
[14] One identifies a Lie algebra $\mathcal{g}$ with the space $\mathcal{g}^{**}$ of linear forms over its dual $\mathcal{g}^*$. This identification allows us to consider a basis L^A of $\mathcal{g}$ as a set of coordinates for $\mathcal{g}^*$.
[15] H. Fukutaka and T. Kashiwa, Ann. Phys. **176**, 301 (1987).
[16] L. Schulman, *Techniques and Applications of Path Integration* (Wiley, New York, 1981).
[17] D. Arnal and J. C. Cortet, "Star Representations of $E(2)$," to be published.
[18] G. Watson, *Theory of Bessel Functions*, (Cambridge U. P., London, 1966).
[19] E. Titchmarsh, *Introduction to the Theory of Fourier Integrals* (Oxford–Clarendon, New York, 1959), 2nd ed.

The gravitational field of plane symmetric thick domain walls

Guenter Goetz
Astronomy and Astrophysics Center, University of Chicago, Chicago, Illinois 60637 and NASA/Fermilab Astrophysics Center, Fermi National Accelerator Laboratory, Batavia, Illinois 60510

(Received 18 August 1989; accepted for publication 27 June 1990)

Exact solutions of Einstein's equations for a scalar field with a potential $V(\Phi) = V_0 \cos^{2(1-n)}(\Phi/f(n))$ $(0<n<1)$ are presented describing the gravitational field of thick, plane symmetric domain walls. The scalar field has a time-independent kinklike distribution, whereas the metric depends on a time coordinate. The metric is conformally flat and the hypersurfaces parallel to the wall ($z = \text{const}$) are three-dimensional de-Sitter spaces. A particle horizon exists on which the metric becomes Minkowski space. It is shown that the gravitational field experienced by a test particle is repulsive.

I. INTRODUCTION

The renewed interest in the cosmological significance of domain walls is mainly due to a proposal for a new scenario of galaxy formation by Hill, Schramm, and Fry.[1] In the Hill, Schramm, and Fry scenario, the seeds for galaxies are the topological defects (domain walls) produced during a phase transition after the time of recombination of matter and radiation. The phase transition is triggered by the breaking of a descrete symmetry of a weakly coupled scalar field of pseudo-Goldstone bosons. The mass of the scalar field m_Φ is assumed to be in a range so that the spatial scale of the topological defects $1/m_\Phi$ lies in the range of Mpc. Such light, thick domain walls are assumed to provide the gravitational field necessary for the clustering of dark matter and baryons after recombination. Thus the scenario alleviates the problem of too short time scales in hot dark matter and pure baryonic scenarios that do not allow enough clustering to take place between recombination time and the present. This scenario also avoids the severe restrictions on the amplitude of the density inhomogeneities imposed on other scenarios of galaxy formation by the high isotropy of the microwave background because clustering of matter takes place entirely after recombination. Since the domain walls associated with such a late-time phase transition are light compared to wall-like topological defects associated with a GUT phase transition, they do not dominate the energy density of the universe, as is the case for GUT domain walls.[2]

Since the scenario for galaxy formation based on a late-time phase transition has been proposed very recently, almost no details have been worked out up to the present. In Ref. 3 an attempt has been made to determine approximately the gravitational field of thick, plane symmetric domain walls and in Ref. 4 the stability of infinitely thin, spherical walls has been investigated. The numerical simulations in Refs. 5 and 6 deal with the dynamics and the interaction of thick domain walls on a background metric that is not affected by the gravitational field of the walls. A rough estimate of how matter accretes onto topological defects is contained in Ref. 7 and in Ref. 8 the cosmological consequences of light domain walls are discussed. The gravitational field of infinitely thin walls has been calculated by Vilenkin[9-11] (see also Ref. 12).

The purpose of this paper is to study the gravitational field of plane symmetric thick domain walls in the framework of general relativity. Since the thickness of the wall $1/m_\Phi$ is not small compared to a typical curvature radius, the thin wall approximation of Refs. 9 and 12 does not apply.

In Sec. II a metric for a plane symmetric domain wall is derived. We find a solution for a potential $V(\Phi) = V_0 \cos^{2(1-n)}(\Phi/f(n))$ $(V_0, n = \text{const}, 0<n<1)$ with the usual kinklike distribution of Φ whose metric is conformally flat and where the hypersurfaces $z = \text{const}$ (z is a coordiante running perpendicular to the wall) are three-dimensional de-Sitter spaces. Due to the conformal flatness and a vanishing energy momentum tensor at $|z| \to \infty$, the space-time becomes Minkowski space for $|z| \to \infty$. A particle horizon exists in all directions: In the planes parallel to the wall we have the usual de-Sitter horizon and in the z direction there is a horizon whose distance depends on the parameters n and V_0. In Sec. III the geodesic equations for a test particle moving in the gravitational field of these domain walls are investigated. It is shown that the gravitational field is repulsive.

A thick domain wall can be viewed as a solitonlike solution of the scalar field equation coupled to gravity. In order to determine the gravitational field one has to solve Einstein's equation

$$G_{\mu\nu} = 8\pi G T_{\mu\nu} \tag{1.1}$$

with an energy momentum tensor $T_{\mu\nu}$ describing a scalar field Φ with self-interactions contained in a potential $V(\Phi)$:

$$T_{\mu\nu} = \partial_\mu \Phi \, \partial_\nu \Phi - g_{\mu\nu}(\tfrac{1}{2}\partial_\sigma \Phi \, \partial^\sigma \Phi - V(\Phi)). \tag{1.2}$$

In the following we will derive solutions of (1.1) and (1.2) for a potential $V(\Phi)$:

$$V = V_0 \cos^{2(1-n)}(\Phi/f(n)) \tag{1.3}$$

$(n, V_0 = \text{const}, 0<n<1)$ which is similar to that proposed in Ref. 1 as the potential for weakly interacting pseudo-Goldstone bosons. Since a domain wall is essentially a one-dimensional planar object we seek solutions to (1.1), (1.2), and (1.3) with a three-dimensional symmetry group: Two commuting Killing vectors describing translational invariance in the planes parallel to the wall and a third Killing vector related to a rotational symmetry about the axis perpendicular to the wall. In addition, we assume that the scalar field is

 0022-2488/90/112683-05$03.00

time independent in some coordinate system, whereas the metric is allowed to depend on time.

II. PLANE SYMMETRIC DOMAIN WALLS

The general metric for a plane symmetric space-time can be parametrized as follows:[13]

$$ds^2 = e^{\nu(z,t)}dt^2 - e^{\lambda(z,t)}dz^2 - e^{\psi(z,t)}(dx^2 + dy^2). \tag{2.1}$$

This metric admits three Killing vectors:

$$\partial_x, \quad \partial_y, \quad x\partial_y - y\partial_x. \tag{2.2}$$

Now we assume that the scalar field does not depend on the time coordinate t, i.e., we have $\Phi = \Phi(z)$. For such a time independent scalar field the components of the energy momentum tensor (1.2) are

$$T_t^t = T_x^x = T_y^y = \tfrac{1}{2}e^{-\lambda}\Phi'^2 + V(\Phi) \equiv \rho,$$
$$T_z^z = -\tfrac{1}{2}e^{-\lambda}\Phi'^2 + V(\Phi) \equiv -p. \tag{2.3}$$

But we do not assume that the metric is independent of t, which clearly imposes some restrictions on the metric variables ν, λ, ψ. We proceed similar to Ref. 3 and consider first the scalar field equation $\Phi^{;\mu}{}_{;\mu} + dV/d\Phi = 0$:

$$\Phi'' + \Phi'\left(\psi' + \frac{1}{2}(\nu' - \lambda')\right) = e^{\lambda}\frac{dV(\Phi)}{d\Phi}, \tag{2.4}$$

"prime" and "dot" denote derivatives with respect to z and t, respectively. Since $\dot{\Phi} = 0$, the coefficient of the Φ' term in (2.4) and the coefficient of the potential term must be functions of z only:

$$\lambda = \lambda(z), \quad \partial_t(\psi' + \tfrac{1}{2}(\nu' - \lambda')) = 0. \tag{2.5}$$

Moreover, $\dot{\Phi} = 0$ implies

$$G_t^z = 0 \Rightarrow 2\dot{\psi}' - \nu'\dot{\psi} + \dot{\psi}\psi' = 0, \tag{2.6}$$

which together with (2.5) puts the line element (2.1) in the following form:

$$ds^2 = A(z)[dt^2 - dz^2 - b(t)(dx^2 + dy^2)]. \tag{2.7}$$

The freedom to transform the coordinates $t \to \tilde{t} = g(t)$ and $z \to \tilde{z} = h(z)$ has been used to eliminate the time dependence of the g_{tt} component of the metric tensor and to set $g_{zz} = -g_{tt} = -A$. The three functions $A(z)$, $b(t)$, and $\Phi(z)$ are determined by the remaining three Einstein equations:

$$G_t^t - G_x^x = 0 \Rightarrow \ddot{b}b - \dot{b}^2 = 0, \tag{2.8}$$

$$G_x^x - G_z^z = -\frac{A''}{A^2} + \frac{3}{2}\frac{A'^2}{A^3} - \frac{c^2}{2A} = 8\pi G\frac{1}{A}\Phi'^2, \tag{2.9}$$

$$G_t^t + G_z^z = -\frac{A''}{A^2} + \frac{c^2}{A} = 16\pi G V(\Phi), \tag{2.10}$$

where c is the integration constant in the nontrivial solution of (2.8):

$$b = e^{ct}, \tag{2.11}$$

where (2.9) and (2.10) are the two equations that determine $A(z)$ and $\Phi(z)$ for a given $V(\Phi)$. Starting with the ansatz that $A(z)$ be proportional to some power of $\cosh(z)$, we find a solution to Eqs. (2.9) and (2.10) with a scalar field potential given by (1.3) whose energy momentum tensor vanishes for $|z| \to \infty$. Note that all the coordinates can be rescaled so that all trivial integration constants, and also the constant c in (2.11), are removed from the solution. Then the line element and the scalar field are given by

$$ds^2 = \frac{1}{L^2}\frac{1}{\cosh^{2n}(z)}[dt^2 - n^2dz^2 - e^{2t}(dx^2 + dy^2)], \tag{2.12}$$

$$\Phi = f\arcsin[\tanh(z)], \tag{2.13}$$

$$V(\Phi) = V_0\cos^{2(1-n)}(\Phi/f), \tag{2.14}$$

$$f \equiv \left[\frac{n(1-n)}{4\pi G}\right]^{1/2}, \quad V_0 \equiv \frac{2n+1}{(8\pi G)n}L^2, \quad 0 < n < 1, \tag{2.15}$$

where n and L are two nontrivial constants in the solution. Here, n is dimensionless and determines the energy scale f of the scalar field and the power of the cosine in $V(\Phi)$; L has the dimension of a mass and determines the amplitude V_0 of the scalar field potential. The energy density ρ, the pressure p perpendicular to the wall [see (2.3)] and V as a function of z are given by

$$\rho = V_0\frac{n+2}{2n+1}\frac{1}{\cosh^{2(1-n)}(z)}, \tag{2.16}$$

$$p = -V_0\frac{3n}{2n+1}\frac{1}{\cosh^{2(1-n)}(z)}, \tag{2.17}$$

$$V = V_0\frac{1}{\cosh^{2(1-n)}(z)}. \tag{2.18}$$

The scalar field Φ has a kinklike distribution that becomes $\Phi \to (\pi f)/2$ for $z \to +\infty$ and $\Phi \to -(\pi f)/2$ for $z \to -\infty$. The energy density ρ and the potential energy of the field V both have a maximum at $z = 0$ and vanish for $|z| \to \infty$. Such a scalar field distribution and energy momentum tensor characterize a domain wall. The pressure in the (x,y) planes is clearly isotropic and equal to $-\rho$. The pressure p perpendicular to the wall is negative for any choice of the parameters and has a minimum at $z = 0$. Note also that the potential probes only a half-period of the cosine in (2.14).

The hypersurfaces $z = \text{const}$ are obviously three-dimensional de-Sitter spaces. It is easy to see that the metric (2.12) is conformally flat. A coordinate transformation

$$\hat{t} = -e^{-t}\cosh(nz), \quad \hat{z} = e^{-t}\sinh(nz), \tag{2.19}$$

casts the line element (2.12) into the form

$$ds^2 = \frac{2^n}{L^2}\frac{1}{[(\hat{t}-\hat{z})^{1/n} + (\hat{t}+\hat{z})^{1/n}]^{2n}} \times [d\hat{t}^2 - d\hat{z}^2 - dx^2 - dy^2]. \tag{2.20}$$

The conformal flatness of this thick wall metric is a property shared by the metric of an infinitely thin wall.[9] The energy momentum tensor vanishes for $|z| \to \infty$. Since the only conformally flat vacuum space-time is Minkowski space the metric becomes flat space far away from the wall $|z| \to \infty$. For $|z| \to \infty$ the line element in $(\hat{t},\hat{z})$ coordinates becomes

$$ds^2 \to \frac{1}{L^2}\frac{1}{(\hat{t}-\hat{z})^2}[d\hat{t}^2 - d\hat{z}^2 - dx^2 - dy^2], \tag{2.21}$$

whose Riemann tensor vanishes.

An interesting property of this metric for a thick domain

wall is the existence of a particle horizon in every direction. Since the hypersurfaces $z = \text{const}$ are de-Sitter spaces we have the usual de-Sitter horizon in the x and y direction. This will be demonstrated in more detail in the context of a discussion of the null geodesics in Sec. III, see (3.10) and (3.11). There is also a horizon in the z direction since the proper distance s_H between $z = 0$, the center of the wall, and the Minkowski vacuum at $z \to \infty$, measured along a spacelike curve running perpendicular to the wall ($t,x,y = \text{const}$), is finite:

$$s_H = \frac{n}{L}\int_0^\infty \frac{1}{\cosh^n(z)}\,dz = \frac{1}{L}\,2^n\frac{[\Gamma(1+n/2)]^2}{\Gamma(1+n)}, \tag{2.22}$$

$\Gamma(x)$ is the standard Gamma function. This means that a particle moving in the gravitational field of the wall can at most travel a distance s_H in the z direction from the center of the wall. Here, s_H is always of the order $1/L$ and attains the following values for the two extremal values of n:

$$s_H = 1/L, \quad \text{for } n \to 0,$$
$$s_H = (\pi/2)(1/L), \quad \text{for } n \to 1. \tag{2.23}$$

Finally, we determine the energy density per surface element σ of the domain wall. We define σ as the integral of ρ over the element of proper length along the z axis:

$$\sigma \equiv \int_{-s_H}^{s_H} \rho\,ds = \frac{n}{L}\int_{-\infty}^{+\infty} \rho\,\frac{1}{\cosh^n(z)}\,dz$$
$$= \frac{L(n+2)}{8\pi G}\,2^{1-n}\,\frac{[\Gamma(1-n/2)]^2}{\Gamma(2-n)}. \tag{2.24}$$

This surface energy density is always of the order of L/G because the dependence on n is very weak:

$$\sigma = (1/2\pi)(L/G), \quad \text{for } n \to 0,$$
$$\sigma = (3/8)(L/G), \quad \text{for } n \to 1. \tag{2.25}$$

III. GEODESICS

In this section we discuss the properties of the gravitational field of the domain wall configuration set forth in the last section by studying the geodesic equations for test particles.

For the metric (2.12) the first integrals of the geodesic equations for a particle on a trajectory $x^\mu = (t(\tau),z(\tau),x(\tau),y(\tau))$ are (henceforth, "dot" denotes a derivative with respect to the affine parameter τ):

$$\dot{x} = ue^{-2t}(1/A), \quad \dot{y} = we^{-2t}(1/A), \tag{3.1}$$
$$\dot{t}^2 = (1/A^2)[E^2 + (u^2+w^2)e^{-2t}], \tag{3.2}$$
$$\dot{z}^2 = (1/n^2A)[E^2/A - \mu^2], \tag{3.3}$$
$$A(z) \equiv (1/L^2)[1/\cosh^{2n}(z)], \tag{3.4}$$

where E, u, w are integration constants and $\mu^2 = 0,1$ for light rays and for particles with nonzero mass, respectively. Due to the rotational symmetry about the z axis [see (2.2)] $x(\tau)$ and $y(\tau)$ have the same functional dependence on t,z. The acceleration in the z direction is given by

$$\ddot{z} = \frac{2}{n}\tanh(z)\,\frac{1}{A}\left[\frac{E^2}{A} - \frac{1}{2}\mu^2\right]. \tag{3.5}$$

Since $E^2/A - \mu^2$ must be positive, the term in brackets in (3.5) is also positive. Thus, the acceleration is positive for $z > 0$ and negative for $z < 0$, meaning that the gravitational field (2.12) is repulsive in the z direction. The acceleration increases with increasing distance from the wall and becomes infinite on the horizon $|z| \to \infty$. Repulsive gravitational fields that are related to the negative stress components in the energy momentum tensor are also known to exist for infinitely thin domain walls.[12]

It is possible to integrate (3.1) once more and express x,y as a function of t instead of τ:

$$\frac{1}{u}(x - x_0) = \frac{1}{w}(y - y_0)$$
$$= -\frac{1}{u^2+w^2}\sqrt{E^2 + (u^2+w^2)e^{-2t}},$$
$$\text{for } \mu^2 = 0,1, \tag{3.6}$$

$x_0,y_0 = \text{const}$ are some initial positions of the particle. In (3.6) the positive root of $\dot{t}^2$ in (3.2) was used. In general, it is not possible to determine the relation between t and τ since (3.2) cannot be integrated analytically, both for $\mu^2 = 1$ and $\mu^2 = 0$. For light rays ($\mu^2 = 0$) one can integrate $dz/dt = \dot{z}/\dot{t}$ and determine $z(t)$:

$$\coth(n(z - z_0)) = \pm\,[1 + (1/E^2)(u^2+w^2)e^{-2t}]^{1/2},$$
$$\text{for } \mu^2 = 0;\ u^2 + w^2 \neq 0, \tag{3.7}$$

$z_0 = \text{const}$ is again an initial value. For a photon with zero momentum in the x and y directions ($u = w = 0$) the integration yields

$$z - z_0 = \pm\,(1/n)t, \quad \text{for } \mu^2 = 0;\ u = w = 0. \tag{3.8}$$

The two signs in (3.7) and (3.8) are due to the two roots of $\dot{z}^2$ in (3.3). Thus all null geodesics of the metric (2.12) are given by (3.6), (3.7), and (3.8). This result could also have been inferred from the fact that the metric is conformally flat and the coordinate transformation (2.19) which transforms (2.12) into (2.20).

Next, we want to discuss the geometric shape of the particle trajectories in a space like hypersurface for $\mu^2 = 0$. Since, by means of a rotation of the coordinates in the (x,y) plane, one can always orientate the coordinate system such that the y component of the position vector of the particle coincides with $y = 0$, we henceforth assume $w = 0 \Leftrightarrow \dot{y} = 0$, without loss of generality. Therefore, the motion of the particle is entirely confined to the (x,z) plane. By eliminating t from (3.6) and (3.7) one has

$$(x - x_0)(u/E) \pm \coth(n(z - z_0)) = 0, \quad \text{for } \mu^2 = 0;\ u \neq 0. \tag{3.9}$$

As $z \to z_0$, x becomes infinite and for $|z| \to \infty$ the x coordinate approaches a constant value $|x - x_0| \to |E/u|$. Thus a photon starting at the horizon $|z| \to \infty$ with nonvanishing momentum in the x direction is always bent away from the wall and reaches the plane $z - z_0 = 0$ at $|x| \to \infty$. Vice versa, photons emerging from the center of the wall are forced onto trajectories parallel to the z axis as they approach the hori-

zon. The location where their trajectories intersect the horizon orthogonally is determined by the ratio E/u. The interesting feature is that photons with nonzero momentum in the x direction are either confined to the region $z-z_0>0$ or to the region $z-z_0<0$, corresponding to the two branches of the coth function in (3.7) and (3.9). Since the particles are drifting toward $|x|\to\infty$, as they approach the plane $z-z_0=0$, it is not possible for them to cross $z-z_0=0$ continuously. Only photons moving strictly perpendicular to the wall ($u=w=0$) can traverse the entire horizon between $z\to+\infty$ and $z\to-\infty$. But these trajectories orthogonal to the wall are unstable in the sense that the slightest initial momentum in the x direction increases and therefore prevents the particle from crossing the plane $z-z_0=0$.

Next, we want to show that the proper distance that corresponds to an infinite value of the x coordinate is finite. The proper distance measured along the x axis at a constant $z=z_0$ is

$$s=(1/L)[1/\cosh^n(z_0)]e^t(x-x_0). \tag{3.10}$$

According to (3.6), $|x|$ becomes infinite for $t\to-\infty$. Therefore, as the particle moves toward $z\to z_0$, $|x|\to\infty$, the proper distance in the x direction between x_0 and $x\to\infty$ becomes

$$s\to(1/L)[1/\cosh^n(z_0)]. \tag{3.11}$$

Note that this holds both for photons and massive particles since (3.6) is valid for $\mu^2=0$ and $\mu^2=1$. (3.11) is nothing else than the usual de-Sitter horizon in the $z=$ const hypersurfaces. Since we know that the maximum proper distance a particle can travel in the z direction is also finite (2.22), there exists a horizon in every direction which is of the order of $1/L$.

Although it is not possible to integrate the geodesic equations for massive particles ($\mu^2=1$) completely one can determine the qualitative behavior from (3.2), (3.3), and (3.6). Assuming, without loss of generality, $w=0$, the path of a massive particle in the (x,z) plane is given by

$$\int_{z_0}^{z}\frac{dz'}{[1-(1/E^2)A(z')]^{1/2}}\pm\frac{1}{n}\operatorname{arcoth}\left[\frac{u}{E}(x-x_0)\right]=0, \quad \text{for } \mu^2=1,\quad u\neq 0, \tag{3.12}$$

where $A(z)$ is given by (3.4). For $|z|\to\infty$, $A(z)$ goes to zero and the integral in (3.12) is $\propto z$. Therefore, far away from the wall and close to the horizon the trajectories of massive particles are similar to the paths of photons (3.9). The integrand in (3.12) is everywhere finite only if $(EL)^2>1$. For $(EL)^2\ll 1$ the integrand is singular at $|z|=z_T$ where $\dot z^2$ becomes zero (3.3) and therefore the lower limit in the integral must be $|z_0|\geq z_T$. In the case $(EL)^2>1$ for z close to zero the integrand in (3.12) becomes almost a constant and the integral is again approximately a linear function of z implying that close to the wall the shape of the trajectories resembles (3.9). Paths of particles with zero momentum in the x direction, $u=0$, are determined by

$$\int_{z_0}^{z}\frac{dz'}{[1-(1/E^2)A(z')]^{1/2}}\pm\frac{1}{n}t=0, \quad \text{for } \mu^2=1,\quad u=0. \tag{3.13}$$

IV. CONCLUDING REMARKS

To summarize, we have found solutions of the coupled Einstein-scalar field equations describing a thick domain wall with kinklike scalar field distribution and vanishing energy momentum tensor far away from the center of the wall. The metric is conformally flat and the hypersurfaces $z=$ const are three-dimensional de-Sitter spaces. For all combinations of the parameters a particle horizon exists on which the space-time becomes Minkowski space. The gravitational field of the domain wall has been shown to be repulsive. Massive and massless test particles coming from the center of the wall are moving on trajectories which become orthogonal to the wall as they approach the horizon. Particles moving toward the wall and with nonzero momentum components parallel to the wall approach paths parallel to the wall as they come close to the center and are therefore prevented from crossing the wall and forced toward the de-Sitter horizon. Only particles moving strictly orthogonal to the wall can cross the wall and traverse the entire horizon. The horizons in all directions are of the order of $1/L$.

An important question we have not addressed in this paper is whether these scalar field configurations are stable. The 1-D, planar kink solution in Minkowski space is known to be stable with respect to small perturbations due to Derrick's theorem.[14] Thus, there is at least a chance that the domain walls discussed in this paper are stable. However, for the application of these scalar fields in the context of a cosmological scenario of structure formation the stability is not of crucial importance because the domain walls are supposed to provide only the seeds for baryon inhomogeneities. Such a scenario could work if the domain walls survive a certain period of time during which enough baryonic matter accretes in the gravitational field of the walls to account for the observed masses of galaxies and clusters. Therefore, the scenario is only viable if the time scale on which the scalar fields decay is equal or larger than the typical time scale of accretion. But it is not necessary that the domain walls are stable.

To emulate the parameters in the scalar field potential (1.3) of the scenario proposed in Ref. 1, one has to choose $n\approx 10^{-8}$ for an energy scale f of the scalar field $f\approx 10^{15}$ GeV [see (2.15)]. In principle, there are two values of n for a given f. But the second one, which is close to 1 for $f\ll m_{pl}\equiv 1/\sqrt{G}$, would yield a potential $V(\Phi)$ and an energy momentum tensor that are almost constant [see (2.16)–(2.18)]. For n close to zero the potential is approximately $V\propto\cos^2(\Phi/f)$. The other free parameter, V_0, corresponds to a nonzero neutrino mass m_ν in Ref. 1: $V_0\approx m_\nu^4$, with $m_\nu\approx 10^{-2}$ eV. This gives a value for the mass scale $L\approx\sqrt{n}\,m_\nu^2/m_{pl}\approx 10^{-45}$ GeV. Note that for $n\ll 1$ the density ρ and V are of the same order of magnitude [see (2.16)–(2.18)] whereas the pressure $|p|$ perpendicular to the wall is by a factor of n smaller than ρ.

ACKNOWLEDGMENTS

This work was financially supported by the Alexander von Humboldt foundation (Feodor–Lynen program) and

partially supported by the NSF and NASA at the University of Chicago and by NASA and DOE at Fermilab.

[1] C. T. Hill, D. N. Schramm, and J. N. Fry, Comm. Nucl. Part. Phys. **19**, 25 (1989).
[2] Y. B. Zeldovich, I. Y. Kobzarev, and L. B. Okun, Sov. Phys. JETP **40**, 1 (1975).
[3] L. M. Widrow, Phys. Rev. D **39**, 3571 (1989).
[4] L. M. Widrow, Phys. Rev. D **39**, 3576 (1989).
[5] L. M. Widrow, Phys. Rev. D **40**, 1002 (1989).
[6] W. H. Press, B. S. Ryden, and D. N. Spergel, preprint NSF-ITP-89-51, Princeton University, 1989.
[7] I. Wasserman, Phys. Rev. Lett. **57**, 2234 (1986).
[8] A. Stebbins and M. S. Turner, Ap. J. **339**, L13 (1989).
[9] A. Vilenkin, Phys. Lett. B **133**, 177 (1983).
[10] A. Vilenkin, Phys. Rev. D **23**, 852 (1981).
[11] A. Vilenkin, Phys. Rep. **121**, 263 (1985).
[12] J. Ipser and P. Sikivie, Phys. Rev. D. **30**, 712 (1984).
[13] D. Kramer *et al., Exact Solutions of Einstein's Field Equations* (Cambridge U.P., Cambridge, 1980).
[14] G. H. Derrick, J. Math. Phys. **5**, 1252 (1964).

Anomalous foliations of Einstein space-times

Roger Ove[a)]
National Center for Supercomputing Applications, University of Illinois at Urbana–Champaign, 5600 Beckman Institute, 405 North Mathews Street, Urbana, Illinois 61801

(Received 13 September 1989; accepted for publication 20 June 1990)

$T^3 \times R$ space-times with a single global spatial Killing vector can be foliated such that the trace of the two-dimensional metric momentum tensor is a constant over the hypersurfaces. This is analogous to constant mean curvature slicing, but is somewhat more algebraically natural in the case with symmetry. Here, the question is addressed of whether Einstein solutions foliated in such a manner exist where the trace of the extrinsic curvature is not of consistent sign over the surface. The relationship between such behavior and the possibility of singularities local in the quotient space of the symmetry is considered.

I. INTRODUCTION

IT is well known that cosmological space-times with topology $T^3 \times R$ may be foliated with constant mean curvature (CMC) hypersurfaces. Such surfaces are defined by the requirement that the four-dimensional divergence of the unit normal vector, or equivalently the trace of the extrinsic curvature K, is a constant. The observed singularity avoidance behavior of families of these surfaces makes them suitable as a time coordinate.

The time rate of change of the volume of the hypersurfaces is related to suitably measured spatial integral of K. In an expanding space-time, for example, this average of K is positive.

In the case of such space-times with a single spatial Killing vector, so called $2+1+1$ space-times, a natural alternative presents itself. The action takes the form

$$S = \int dt\, d^2x \{\pi^{ab} g_{ab,t} + e^a \beta_{a,t} + p\phi_{,t} - \tilde{N}\tilde{H} - \tilde{N}^a \tilde{H}_a - \beta_0 e^a_{,a}\} \tag{1}$$

and the slicing condition

$$g_{ab}\pi^{ab}/\sqrt{g} = \tau(t) \tag{2}$$

is found to be effective in both simplifying the dynamics and evolving the field equations (specifying a consistent global Cauchy problem). In these expressions the roman indices range over the quotient space of the symmetry, a two-dimensional manifold. We can also calculate K for the congruence defined by this prescription, and it appears that they are of consistent sign in an average sense. It does not seem possible for expansion in the direction of symmetry to dominate expansion of the two-dimensional surfaces in such a way that a space-time with expanding areas could be decreasing in volume. There does not seem to be an obvious problem if such behavior is possible, but it has not been observed in numerical examples and we conjecture that the expansions are always globally consistent.

However, it may be possible for K to be locally of different sign than its average in such foliated space-times. The situation could exist in which an expanding space-time contains a local region that is collapsing. The resolution of this question is dependent in a fairly interesting way on the nonlinearity of the Hamiltonian constraint. We address the question by numerical means, which appears to be the only means by which it can be done, and establish that such behavior is indeed possible. Both the local question and the global conjecture are dependent on the interplay between the Einstein equations, the foliation conditions, and symmetry.

We then provide some evidence to indicate that the behavior is not generic, that if a hypersurface is initialized with such a local collapsing patch it leaves this state and the evolution is not characterized thereafter by a magnification of the local inconsistency in the signs of τ and K. The question is of interest in that such local patches of inconsistency are likely candidates to form local (in the quotient space) singularities in expanding space-times. It is not known if the Einstein equations allow such singularities.

It should be emphasized that in this paper we use K to refer to the divergence of the normal to the surfaces defined by the $\tau = \tau(t)$ condition, not a CMC foliation. There may be differences in sign convention as well. Here, K is defined by

$$K \equiv -n^\alpha_{;\alpha} \tag{3}$$

where n^α is the unit normal.

II. FORMALISM

Our description of the formalism will be necessarily brief and largely incomplete. The construction is due to Moncrief,[1] and is related to the formalism underlying the solution generating methods of Geroch.[2,3] Additional details specific to the torus can also be found elsewhere.[4] Allowing the Killing vector to play the role of a Kaluza–Klein "extra" dimension, one can cast the four-dimensional Einstein equations in a form equivalent to three-dimensional gravitation coupled to a scalar and vector field. In the five-dimensioal theory this vector field would obey the Maxwell equations and have two degrees of freedom, but in the lower-dimensional case this is reduced to one. Gravitation in three dimensions is not a dynamical theory (it has no degrees of freedom), and the full two degrees of freedom of the four-dimensional Einstein equations are represented by the dy-

[a)] Present address: Departments of Mathematics and Astronomy, Univ. of Illinois at Urbana–Champaign, 5600 Beckman Inst., 405 N. Mathews St. Urbana, IL 61801.

 0022-2488/90/112688-06$03.00

namics of the scalar and vector fields.

The metric takes the form

$$ds^2 = e^{-2\phi}\, d\sigma^2 + e^{2\phi}\{dx^3 + \beta_a\, dx^a + \beta_0\, dt\}^2, \tag{4}$$

where

$$d\sigma^2 = -\tilde{N}^2\, dt^2 + g_{ab}(dx^a + \tilde{N}^a\, dt)(dx^b + \tilde{N}^b\, dt). \tag{5}$$

All variables are functions of t, x^1, and x^2 only, as x^3 has been singled out as the direction of symmetry. Terms denoted by a tilde are two-dimensional quantities, and the letters $(a,b,c,\ldots)$ will be used for their indices. Four-dimensional quantities will have their indices denoted by lower case Greek letters, while $(i,j,k,\ldots)$ will be used for three-dimensional quantities.

Parametrized in this way, the Einstein action is

$$S = \int dt\, d^2x\{\pi^{ab} g_{ab,t} + e^a \beta_{a,t} + p\phi_{,t} - \tilde{N}\tilde{H} - \tilde{N}^a \tilde{H}_a - \beta_0 e^a_{,a}\}, \tag{6}$$

where

$$\tilde{H} = \frac{1}{\sqrt{g}}\left\{\pi^{ab}\pi_{ab} - (\pi^a_a)^2 + \frac{1}{8}p^2 + \frac{1}{2}e^{-4\phi} g_{ab} e^a e^b\right\} + \sqrt{g}\{-{}^{(2)}R + 2g^{ab}\phi_{,a}\phi_{,b} + \tfrac{1}{4}e^{4\phi} g^{ac} g^{bd}(\beta_{a,b} - \beta_{b,a})(\beta_{c,d} - \beta_{d,c})\}, \tag{7}$$

$$\tilde{H}_a = -2\pi^b_{a;b} + p\phi_{,a} + e^b(\beta_{b,a} - \beta_{a,b}).$$

The dynamics of the equations are represented by the scalar field ϕ and the vector field β, and these and their momenta are the only fields that are explicitly evolved. Additional quantities are determined by imposing coordinate conditions and solving the Hamiltonian and momentum constraints, all of which can be cast as elliptic equations (and except for the Hamiltonian constraint, linear elliptic equations).

The metric g_{ab} is a metric on the two-torus. Since all such metrics are conformal to a flat metric, the metric can be put in this form by making use of the coordinate freedom to impose

$$g_{ab} = e^{2\lambda} f_{ab}(t), \quad \det(f) = 1. \tag{8}$$

This condition leads to elliptic equations determining the shift functions.

One of the constraints involving the "electric" field,

$$e^a_{,a} = 0, \tag{9}$$

can be formalized away by

$$e^a = \epsilon^{ab}\omega_{,b}, \tag{10}$$

which defines a twist potential ω. The momentum conjugate to this field is $\epsilon^{ab}\beta_{c,d}$, which we denote by r. The evolution equations for the dynamical fields (ϕ,p) and (ω,r) will not be reproduced here, nor will those of the flat metric and its "momentum."

In addition to prescribing how the coordinates thread themselves through the successive hypersurfaces (done by preserving the flatness of the metric f_{ab}), a coordinate condition must be imposed to specify the local proper distance between successive slices. We choose a time slicing condition that is a two-dimensional variant of the familiar CMC condition. As mentioned earlier, this is done by imposing the requirement

$$g_{ab}\pi^{ab}/\sqrt{g} = \tau(t). \tag{11}$$

The expansion of the normal to the two-dimensional surfaces (the symmetry direction has been factored out) is a constant over the surface. Here, $\tau(t)$ is an arbitrary monotonic function, and the requirement that the form of this equation be invariant in time determines the lapse.

Of central importance to the current discussion is the Hamiltonian constraint, which determines the conformal factor factor λ. It is equivalent to the nonlinear elliptic equation

$$\Delta\lambda + \alpha + Pe^{-2\lambda} - Qe^{2\lambda} = 0, \tag{12}$$

where

$$\alpha \equiv f^{ab}\phi_{,a}\phi_{,a} + \tfrac{1}{4}e^{-4\phi} f^{ab}\omega_{,a}\omega_{,b},$$

$$P \equiv \frac{1}{2}r^b_a r^a_b + \frac{p^2}{16} + \frac{r^2}{4}e^{4\phi},$$

$$Q \equiv \tau^2/4, \tag{13}$$

and r^b_a is related to the two-dimensional momentum tensor. This quantity must be determined beforehand by solving the momentum constraints. However, both constraints can be solved without knowledge of the lapse or shift functions.[4] Application of the minimum principle allows one to define a bounding function

$$\Lambda = \tfrac{1}{2}\ln[(\alpha + \sqrt{\alpha^2 + 4PQ})/2Q], \tag{14}$$

$$\Lambda|_{\min} \leqslant \lambda \leqslant \Lambda|_{\max}, \tag{15}$$

where the minimum and maximum are taken over the two-torus.

The time derivative of the slicing condition results in a linear elliptic equation for the field $\tilde{N}$, which is related to the true lapse $\tilde{N}e^{-\phi}$. In terms of the definitions for the Hamiltonian constraint above, this elliptic equation takes the form

$$\Delta\tilde{N} - 2\tilde{N}(Pe^{-2\lambda} + Qe^{2\lambda}) = -e^{2\lambda}\frac{d\tau}{dt}. \tag{16}$$

We will refer to this as the lapse equation, and make use of it later. It is clear from this equation that $\tilde{N}$ is of the same sign everywhere and does not vanish anywhere on the hypersurface. The sign of $\tilde{N}$ and $d\tau/dt$ are the same.

III. AN AREA "THEOREM"

A consequence of the Hamiltonian constraint and the lapse equation is that an upper bound exists on the pseudo-lapse

$$|\tilde{N}| \leqslant 2\left|\frac{d\tau}{dt}\right|(\tau^2)^{-1}. \tag{17}$$

No positive lower bound appears to exist. If a lower bound existed we could make the rather strong statement that in polarized space-times ($\omega = r = 0$) either (1) the two-dimensioal slicing condition does not generally avoid singularities; or (2) singularities are excluded by the symmetry. The true lapse is $\tilde{N}e^{-\phi}$, which could become arbitrarily small

because of the effect of ϕ. However, the polarized field equations possess an exact symmetry under reversal of the sign of ϕ. As a consequence, if a solution existed where the lapse tended to zero in the vicinity of a local singularity, another solution could be found where the lapse was correspondingly large.

Numerical experiments indicate that it is difficult to force the pseudolapse to become small (relative to the upper bound), and that solutions tend to "bounce" away from such states. Therefore, we have pursued the matter further. Defining

$$L \equiv \frac{d\tau}{dt}(\tau^2)^{-1}, \tag{18}$$

we can demonstrate the following result.

If the pseudolapse is less than a certain value ν in a region of the hypersurface, the fractional area (with proper measure) of that region is less than or equal to the ratio ν/L.

The result follows from integral relationships generated from the Hamiltonian constraint and the lapse equation. Taking a linear combination of these gives

$$\frac{\Delta\tilde{N}}{\tilde{N}} + \Delta\lambda + \alpha - 2Qe^{2\lambda} = -\frac{e^{2\lambda}}{2\tilde{N}}\frac{d\tau}{dt}. \tag{19}$$

We then integrate this equation over the two-dimensional spatial surfaces. Making use of the fact that

$$\int \frac{\Delta\tilde{N}}{\tilde{N}}\, d^2x = \int \frac{f^{ab}\tilde{N}_{,a}\tilde{N}_b}{\tilde{N}^2}\, d^2x \geqslant 0, \tag{20}$$

we arrive at

$$\int \left(\tau^2 - \frac{d\tau}{dt}(\tilde{N})^{-1}\right) e^{2\lambda}\, d^2x \geqslant 0. \tag{21}$$

The equality holds for Kasner space-times. Using the upper bound we can put this in the form

$$\frac{1}{2} \leqslant \left[\int \left(\frac{L}{\tilde{N}}\right) e^{2\lambda}\, d^2x\right]\left(\int e^{2\lambda}\, d^2x\right)^{-1} \leqslant 1. \tag{22}$$

Of course the upper bound itself, a local condition, is much stronger than the left side of this expression. From the right-hand inequality the area result immediately follows.

This implies that the picture of black hole formation accompanied by an ever increasing region of vanishing lapse does not apply to these cosmological space-times with a global Killing vector. As the lapse tends to zero, the region of space over which this happens tends to zero to proportion. The result is supportive of evidence provided elsewhere[5] that local singularities (in the quotient space of the symmetry) are excluded.

IV. THE EXPANSION

The trace of the extrinsic curvature of the foliation can be calculated in a straightforward manner, with the result

$$K = \tau e^{\phi}(1 + pe^{-2\lambda}/4\tau). \tag{23}$$

The quantity $e^{2\phi}$ is the distance around universe in the direction of symmetry. From the evolution equation for ϕ,

$$(\phi_{,t} - \tilde{N}^a\phi_{,a})/\tilde{N} = e^{-2\lambda}p/4, \tag{24}$$

it is clear that the expansion is composed of contributions from the rate of change in this distance and the two-dimensional expansion τ.

From the evolution equation for the conformal factor λ, the time derivative of the area of the quotient space can be found to be

$$\frac{d}{dt}A(t) = -\int \tilde{N}\tau e^{2\lambda}\, d^2x. \tag{25}$$

There is no ambiguity of sign because the slicing is defined with respect to these surfaces.

The situation is different for the full three-dimensional surfaces. Making use of the evolution equations for ϕ and the conformal factor λ, the volume rate of change of the spatial hypersurfaces can also be easily calculated:

$$\begin{aligned}\frac{d}{dt}V(t) &= -\int d^3x\, \tilde{N}e^{-\phi}\left(\tau e^{2\lambda} + \frac{p}{4}\right) \\ &= -\int (K\tilde{N}e^{-\phi})e^{2\lambda-\phi}\, d^3x \\ &= -\int KT_{,t}\sqrt{{}^{(3)}g}\, d^3x,\end{aligned} \tag{26}$$

where T is a local measure of the proper time, $dT = \tilde{N}e^{-\phi}\, dt$. This is consistent with a well-known result, the "first variation formula."[6,7]

We conjecture that the sign of $(d/dt)A(t)$ is equal to that of $(d/dt)V(t)$, but are unable to demonstrate this (perhaps due to blindness). Solutions exist in which locally the integrands do not obey this requirement, as we will demonstrate, so clearly a proof of the conjecture will involve the lapse condition as well as the Hamiltonian constraint. For example, showing that the integral of K is of definite sign would be insufficient. The measure for the volume result involves the pseudolapse, which, in principle, could be negligible except where K is of the wrong sign.

The question involves the nonlinearity of the Hamiltonian constraint, and from this constraint one can construct an integral expression analogous to the familiar exponent sum rules of Kasner space-times. The quantity p is the canonical momentum of the scalar ϕ, and as such is free to be chosen arbitrarily negative as a specification of a Cauchy problem. The constraint can counterbalance the effect by leading to a solution for λ such that $pe^{-2\lambda}$ is bounded in an average sense. This is indeed the behavior of the system. Multiplying the Hamiltonian constraint through by $e^{-2\lambda}$ and integrating over the surface gives

$$\begin{aligned}&\int \left\{2f^{ab}\lambda_{,a}\lambda_{,b} + \alpha + \left(\frac{1}{2}r_a^b r_b^a + \frac{r^2}{4}e^{4\phi}\right)e^{-2\lambda}\right\} e^{-2\lambda}\, d^2x \\ &\quad + \int \left\{\frac{p^2e^{-4\lambda}}{16}\right\} d^2x - \int \left\{\frac{\tau^2}{4}\right\} d^2x = 0,\end{aligned} \tag{27}$$

which leads to the result

$$\langle |pe^{-2\lambda}/2\tau| \rangle \leqslant 1. \tag{28}$$

The angle brackets denote the average, with the measure d^2x. A similar expression with the measure $e^{2\lambda}\, d^2x$ results when the constraint is integrated directly. From this it is apparent that the influence of the constraint is consistent

with the conjecture. Numerical results are also supportive, and it seems probable that an analytic proof could be constructed.

V. THE LOCAL QUESTION

The above arguments indicate that in an average sense the Einstein equations tend to prevent the expansion in the direction of symmetry from dominating the evolution. We now consider the local behavior.

Can the sign of K be different from τ locally?

The answer is yes. We demonstrate this by numerical solution of the initial value problem. Due to the nonlinearity of field equations, it does not appear that this problem can be dealt with by any other means. We then address questions of the implications of such local inconsistent patches by evolving the field equations numerically. A discussion of the computational considerations of numerical solution of this system can be found elsewhere.[8]

The initial data configurations used here all involve a Gaussian in the field p of varying widths and amplitudes, with the width chosen sufficiently small that discontinuities at the "boundary" are not significant. Other independent initial data fields are chosen to be zero. The flat metric is initially defined by $f_{ab} = \mathrm{diag}(1,1)$, and, in general, changes as a consequence of nonzero initial momentum (c^{ab}). The time slicing function $\tau(t)$ is chosen to be t^{-1}; the space-time expands as the time increases. Units of measurement do not enter, as all quantities represent relative distances and are dimensionless.

Figure 1 shows the minimum value of K on the initial hypersurface, as a function of the amplitude of the Gaussian. In this case we have

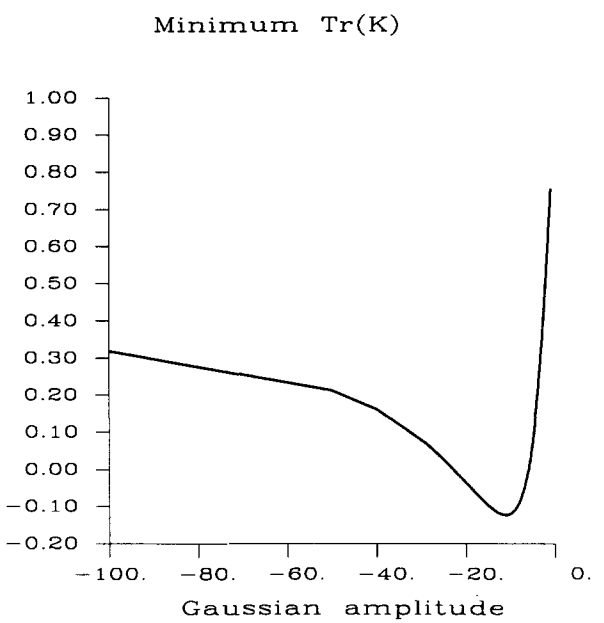


FIG. 1. The minimum of K over the initial hypersurface as a function of the amplitude of the Gaussian in the field p. The width is $\pi/5$ and $c_1^1 = -c_2^2 = \frac{1}{2}$. The accuracy of the numbers is approximately 1 part per thousand, less than the thickness of the line.

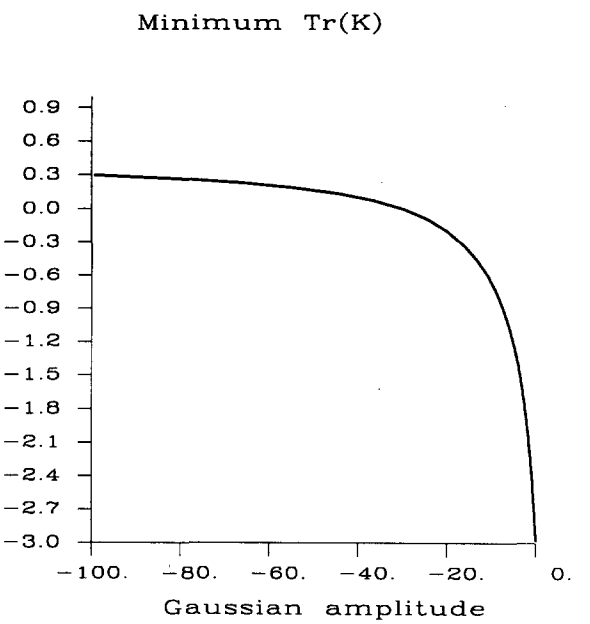


FIG. 2. The minimum of K, with everything as was defined for the Fig. 1 example, except that there is no global stretching $c_b^a = 0$.

$$c_b^a = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & -\frac{1}{2} \end{pmatrix} \tag{29}$$

and the width of the Gaussian is $\pi/5$. The plot indicates that there exist values for the amplitude that lead to negative K. Figure 2 is equivalent except that in this case there is no Kasner style global anisotropy, that is $c_b^a = 0$. Here there are also values of the amplitude that lead to an inconsistent hypersurfaces, in this case including values arbitrarily close to zero (zero itself is ill defined, since the resulting space-time would be static). As the value of c_1^1 is increased further to 1, regions of negative K do not seem to arise, as shown in Fig. 3.

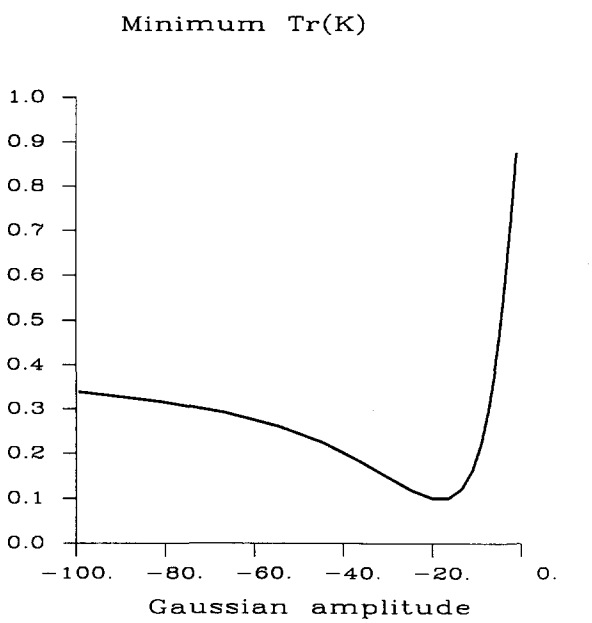


FIG. 3. The minimum of K, similar to Figs. 1 and 2, with the global stretching increased to $c_1^1 = -c_2^2 = 1$.

The behavior is also sensitive to the width of the initial distribution. In the example depicted by Fig. 1, the region of negative K would disappear if the width were increased slightly to $2\pi/15$. It is not clear whether the inconsistent region exists for all values of the width for the case $c_b^a = 0$. Numerical experiments indicate that the qualitative behavior holds until the width is increased beyond reasonable approximation, where the discontinuity at the "boundary" is pronounced and cannot be neglected.

Given that Cauchy surfaces exist with such inconsistent regions, it is of interest to see if there is any qualitative behavior that can be associated with resulting evolution. For example, expanding space-times with locally collapsing patches are intuitively suggestive candidates to form localized singularities. Some evidence has been generated in support of the view that local singularities in such expanding space-times are excluded by the presence of the global Killing vector symmetry,[5] and a stronger analytic result has been generated for the two Killing vector case.[9] In the absence of symmetry singularities would be free to form, but the issue has not been settled for the single Killing vector cosmological space-times under consideration here. The importance of studying the implications of symmetry rests on the fact that much of the lore of general relativity is based upon examples with symmetry.

We continue the evolution of the field equations for the initial data represented by the minimum of Fig. 1, which corresponds to an amplitude of -10. Figure 4 shows the minimum of K as a function of time. Initially there is a slight tendency for the configuration to focus and the minimum decreases, but rather than continue in this manner the situation is reversed and the minimum eventually becomes a maximum. The figure also indicates that the inconsistent patches recur periodically during the evolution, but do not seem to lead to any catastrophic behavior.

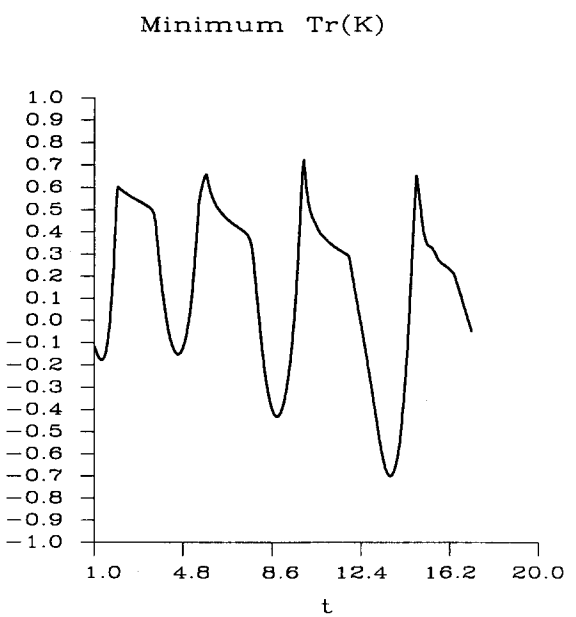


FIG. 4. The evolution of the minimum of K for the most extreme case depicted in Fig. 1, where the amplitude was approximately -10.

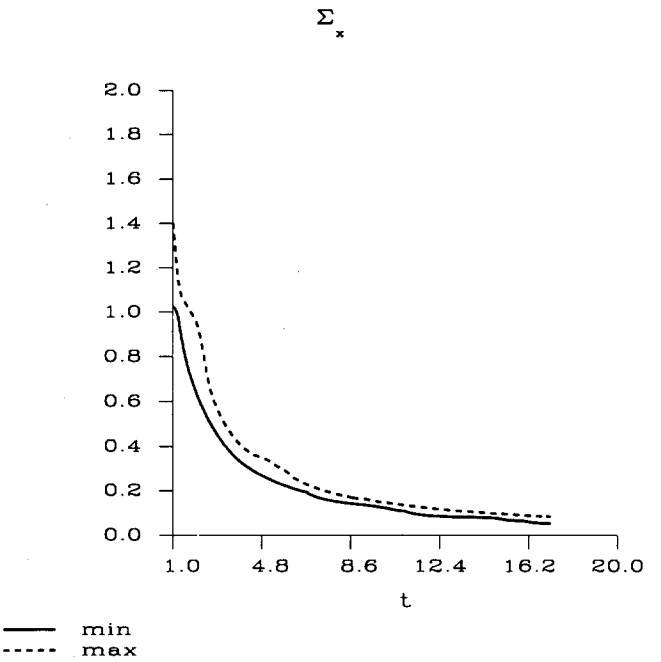


FIG. 5. The range of values covered by the expansion quantity Σ defined with respect to the x coordinate curves, for the evolution depicted in Fig. 4.

In earlier work we used null geodesics to probe numerically generated space-times, and null expansion as an indicator of trapped surfaces.[5] If $\xi_\alpha^{\pm}$ are the in and outgoing null forms associated with a closed spatial two-surface (for convenience chosen here to lie along coordinate trajectories), the expansion sum Σ defined by

$$\sigma^{\pm} \equiv g^{\mu\nu} \xi^{\pm}_{\alpha;\beta} P^{\alpha}_{\mu} P^{\beta}_{\nu}, \qquad \Sigma \equiv \sigma^{+} + \sigma^{-}, \tag{30}$$

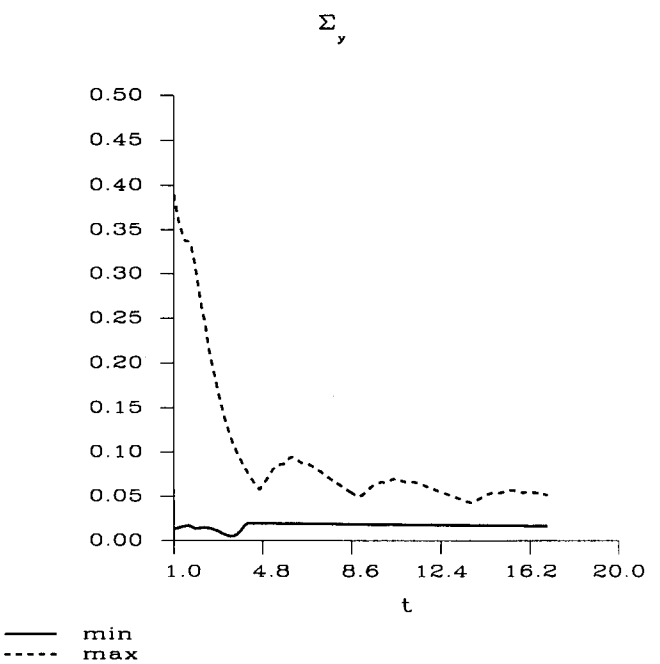


FIG. 6. The range of values covered by the expansion quantity Σ defined with respect to the y coordinate curves, for the evolution depicted in Fig. 4.

serves to identify hypersurfaces likely to contain trapped surfaces. The physical picture is that of photons leaving a surface containing a black hole, with the proper distance between the photons everywhere decreasing.[10]

Figures 5 and 6 show the range of Σ as the space-time evolves, for both sets of coordinate families. Negative values here would provide the first indication that a trapped surface may exist. The fact that these quantities remain positive is supportive of the above result, that the inconsistent regions do not seem to be associated in a causal manner with the formation of local singularities.

VI. CONCLUSIONS

The object of this work is to understand the effects of symmetry on solutions to the Einstein equations, in particular solutions on $T^3 \times R$ with one Killing vector. We have shown that space-times foliated according to a constant curvature prescription defined in the quotient space of the symmetry allow local patches that are inconsistent with the global behavior. Evidence is provided which indicates that such local anomalies do not generically lead to a magnification of the inconsistency and the formation of a local singularity. We also establish a result that supports the view that in an average sense the expansions in the two-dimensional and three-dimensional formulations agree, and that the action of the Einstein equations is such that expansion in one direction cannot dominate the evolution.

Based on these efforts and past numerical work, it appears that the effect of the Killing vector symmetry is to restrict the freedom and range of qualitative features present in solutions. For example, the lack of local singularities in such space-times is not reflective of the natue of truly generic solutions. This is what one might expect, as even the elegant formulation employed here, where the Einstein theory is represented in terms of the dynamics of interacting scalar fields, is an artifact of symmetry. Eliminating the dependence of symmetry and its coloration of scientific intuition is the motivation behind the development of uncompromising solvers of the Einstein equations.

ACKNOWLEDGMENTS

The author would like to thank D. Eardley, V. Moncrief, and L. Smarr for helpful discussions.

This work was made possible by the NSF Supercomputing program, in particular, the National Center for Supercomputing Applications in Champaign, Illinois.

[1] V. Moncrief, Ann. Phys. **167**, 118 (1986).
[2] R. Geroch, J. Math. Phys. **12**, 918 (1971).
[3] R. Geroch, J. Math. Phys. **13**, 394 (1972).
[4] R. Ove, *Dynamical Spacetimes and Numerical Relativity,* edited by J. M. Centrella (Cambridge U. P., Cambridge, England, 1986).
[5] R. Ove, Phys. Rev. D **15**, **39**, 3587 (1989).
[6] J. E. Marsden and F. J. Tipler, Phys. Rep. **66**(3), 109 (1980).
[7] Y. Choquet-Bruhat, A. Fischer, and J. Marsden, in *Isolated Gravitating Systems in General Relativity,* edited by J. Ehlers (Italian Physical Society, Amsterdam, 1979), pp. 396–456.
[8] R. Ove, *Frontiers in Numerical Relativity,* edited by D. W. Hobill, C. R. Evans, and L. S. Finn (Cambridge U. P., Cambridge, England, 1988).
[9] J. Isenberg and V. Moncrief, Commun. Math. Phys. **86**, 485 (1982).
[10] S. Hawking and G. Ellis, *The Large-Scale Structure of Spacetime* (Cambridge U. P., Cambridge, England, 1973).

Large family of colliding waves in the Einstein–Maxwell theory

Mustafa Halilsoy[a)]
Nuclear Engineering Department, King Abdulaziz University, P. O. Box 9027, Jeddah-21413, Saudi Arabia

(Received 28 April 1989; accepted for publication 2 May 1990)

A general class of similarity integrals of the Ernst equation with the method of Chandrasekhar and Xanthopoulos are combined to obtain a family of solutions for colliding waves in the Einstein–Maxwell theory. The family is parametrized by an arbitrary harmonic function that satisfies the Euler–Darboux equation. The first nontrivial member of the family is shown to be the cross-polarized version of the Bell–Szekeres solution.

I. INTRODUCTION

Collision of electromagnetic (em) waves with linear polarization in the Einstein–Maxwell (EM) theory was formulated and solved first by Bell and Szekeres[1] (BS) in 1974. Since then, a considerable amount of work has been done both in the realm of colliding gravitational waves (CGW's) and colliding waves in the EM theory. Colliding pure em waves form an interesting subset of the EM theory that corresponds to the classical limit of the photon–photon scattering. It was shown by BS that linearly polarized waves interact and give rise to delta function curvature on the null boundaries, leaving the interaction region conformally flat. We have shown,[2] more recently, that when the incident waves are endowed with relative polarization, some of the features of colliding linearly polarized em waves modify. The interaction region becomes curved and the Ψ_2 component of the Weyl curvature, which is related to the mass aspect, emerges nonzero. Further, the space-time singularity that develops as a result of mutual focusing in pure gravitational waves does not arise in the em case. Instead, a coordinate singularity presents itself,[3] whose exact physical implication has not been well understood. It is believed, however, that similar to the weakening of the singularity by the addition of the second polarization, coupling of em waves to the gravitational waves serves to weaken the singularity further. This particular aspect shows that there is a significant difference between the collisional behaviors between the em and gravitational waves. For this reason we reiterate our view that the EM solutions that admit independent em and gravitational limits are more important than the other possible types. The most interesting case within the context of colliding EM waves consists of the colliding superposed waves. Recently,[4] we have formulated this as an initial value problem without being able to obtain an exact solution.

In this paper we present a large family of solutions with the second polarization for colliding waves in the EM theory. Our general solution is parametrized in terms of an arbitrary harmonic function X and for particular choices of X it reduces to the solutions obtained previously. The method of obtaining this family is to combine the similarity solutions of the Ernst equation[5] with the general treatment of Chandrasekhar and Xanthopoulos[6] (CX). We state our result as a theorem, according to which, given a vacuum solution for CGW's, we can construct an electrovacuum solution that describes colliding EM waves. We show that the simplest nontrivial member of the family leads to the generalization of the BS solution that was known previously. Some properties of this particular solution are given in the Appendix, where we show, in particular, that it is a type-D metric and is regular everywhere within the range of the coordinates used.

In Sec. II we present the formalism and give the solution for the general family. Particular solutions come next in Sec. III, which is followed by concluding remarks in Sec. IV. A number of properties about the generalized BS solution are proved in Appendix A.

II. METHOD FOR SOLVING EM EQUATIONS

The symmetrical pair of Ernst[7] equations for the EM theory are given by

$$(\xi\bar{\xi} + \eta\bar{\eta} - 1)\nabla^2\xi = 2\nabla\xi(\bar{\xi}\nabla\xi + \bar{\eta}\nabla\eta),$$
$$(\xi\bar{\xi} + \eta\bar{\eta} - 1)\nabla^2\eta = 2\nabla\eta(\bar{\xi}\nabla\xi + \bar{\eta}\nabla\eta), \tag{1}$$

where ξ and η represent the gravitational and em complex potentials, respectively. By introducing the new parametrization

$$Z = (1+\xi)/(1-\xi), \quad H = \eta/(1-\xi), \tag{2}$$

the Ernst equations can be expressed equivalently by

$$(\mathrm{Re}\, Z - |H|^2)\nabla^2 Z = (\nabla Z)^2 - 2\bar{H}\nabla Z\cdot\nabla H,$$
$$(\mathrm{Re}\, Z - |H|^2)\nabla^2 H = \nabla H\cdot\nabla Z - 2\bar{H}(\nabla H)^2. \tag{3}$$

The geometry on which the operators ∇ and ∇^2 act in these equations is given by

$$ds_0^2 = \frac{d\tau^2}{\Delta} - \frac{d\sigma^2}{\delta} + \Delta\delta\, d\phi^2, \tag{4}$$

where $\Delta = 1-\tau^2$, $\delta = 1-\sigma^2$, and ϕ is considered to be a Killing coordinate. The coordinates (τ,σ) are given in terms of the null coordinates (u,v), unless otherwise modified, by

$$\tau = u\sqrt{1-v^2} + v\sqrt{1-u^2}, \quad \sigma = u\sqrt{1-v^2} - v\sqrt{1-u^2}, \tag{5}$$

suitable for the description of CGW's. We adopt the space-time line element introduced by CX that reads

$$ds^2 = e^{\nu+\mu_3}\sqrt{\Delta}\left(\frac{d\tau^2}{\Delta} - \frac{d\sigma^2}{\delta}\right) - \sqrt{\Delta\delta}\left[\chi\, dy^2 + \frac{1}{\chi}(dx - q_2\, dy)^2\right], \tag{6}$$

[a)] Present address: Physics Department, Eastern Mediterranean University, G. Magosa, Mersin 10, Turkey.

 0022-2488/90/112694-05$03.00

where the metric functions depend on (τ,σ) alone. Following CX, we introduce two auxiliary real potentials Ψ and Φ in accordance with the expression

$$Z=\Psi+|H|^2-i\Phi. \tag{7}$$

Once a set (Z,H) of solutions to the Ernst equations is known, the metric function χ is given by

$$\chi=\sqrt{\Delta\delta}/\Psi, \tag{8}$$

whereas q_2 and $\nu+\mu_3$ are integrated from the following coupled equations:

$$\begin{aligned}
&q_{2,\tau}=(\delta/\Psi^2)(\Phi_\sigma-2\,\mathrm{Im}\,H\bar{H}_\sigma),\\
&q_{2,\sigma}=(\Delta/\Psi^2)(\Phi_\tau-2\,\mathrm{Im}\,H\bar{H}_\tau),\\
&-(\sigma/\delta)(\nu+\mu_3)_\tau-(\tau/\Delta)(\nu+\mu_3)_\sigma\\
&\quad=(1/\chi^2)(\chi_\tau\chi_\sigma+q_{2,\tau}q_{2,\sigma})\\
&\qquad+(2\chi/\sqrt{\Delta\delta})(H_\tau\bar{H}_\sigma+\bar{H}_\tau H_\sigma),\\
&2\tau(\nu+\mu_3)_\tau+2\sigma(\nu+\mu_3)_\sigma\\
&\quad=\frac{3}{\Delta}+\frac{1}{\delta}-\frac{4\chi}{\sqrt{\Delta\delta}}(\Delta H_\tau\bar{H}_\tau+\delta H_\sigma\bar{H}_\sigma)\\
&\qquad-\frac{1}{\chi^2}[\Delta(\chi_\tau^2+q_{2,\tau}^2)+\delta(\chi_\sigma^2+q_{2,\sigma}^2)].
\end{aligned} \tag{9,10}$$

The solution that we shall seek in this article is expressed in terms of the Ernst potentials (ξ,η) by

$$\xi=a\xi_0, \tag{11}$$
$$\eta=\sqrt{1-a^2}\,\xi_0\ (a=\text{real constant},\ 0\leqslant|a|\leqslant1),$$

where ξ_0 satisfies the vacuum Ernst equation

$$(|\xi_0|^2-1)\nabla^2\xi_0=2\bar{\xi}_0(\nabla\xi_0)^2. \tag{12}$$

This particular choice for (ξ,η) has the advantage that any future solution is constrained to possess independent em and gravitational limits. To obtain a solution for the vacuum Ernst equation (12) we parametrize ξ_0 in accordance with

$$\xi_0=Y(X)e^{i\beta(X)}, \tag{13}$$

where Y and β are both functions of the function X, which satisfies the Euler–Darboux equation

$$(\Delta X_\tau)_\tau-(\delta X_\sigma)_\sigma=0. \tag{14}$$

For this reason, the function X will be referred to as a harmonic function in the rest of the paper. The solution for ξ_0 is[5]

$$\begin{aligned}
\xi_0=&\left(\frac{\sqrt{1+\sin^2\alpha}\cosh 2X-1}{\sqrt{1+\sin^2\alpha}\cosh 2X+1}\right)^{1/2}\\
&\times\exp[-i\tan^{-1}(\sin\alpha\coth 2X)],
\end{aligned} \tag{15}$$

where α is a constant of integration that, as we shall justify below, can be interpreted as the angle of the second polarization. In terms of Ψ and Φ this solution reads as

$$Z=\Psi-i\Phi=\frac{1-i\sin\alpha\cosh 2X}{\sqrt{1+\sin^2\alpha}\cosh 2X-\sinh 2X}. \tag{16}$$

In order to obtain the corresponding EM solution we shall summarize the procedure of CX as the following theorem.

Theorem: Let $(Z,\Psi,\Phi,\chi,\nu+\mu_3,q_2)$ denote Ernst potentials and the metric functions for a vacuum solution of the Einstein equations. Then, in order to obtain an electrovacuum solution $[Z_e,\Psi_e,\Phi_e,\chi_e,(\nu+\mu_3)_e,q_{2e},H]$, it suffices to make the following replacements:

$$\begin{aligned}
&Z_e=\Psi_e+|H|^2-i\Phi_e,\quad \chi_e=\sqrt{\Delta\delta}/\Psi_e=(\Omega^2/4)\chi,\\
&\Psi_e=(4/\Omega^2)\Psi,\quad \Phi_e=(4a/\Omega^2)\Phi,\\
&H=\sqrt{1-a^2}\,\frac{(\Psi-i\Phi-1)}{(1-a)(\Psi-i\Phi)+a+1},\\
&(\nu+\mu_3)_e=(\nu+\mu_3)+\ln(\Omega^2/4),
\end{aligned}$$

where

$$\Omega^2=(1-a)^2(\Psi^2+\Phi^2)+2(1-a^2)\Psi+(1+a)^2, \tag{17}$$

and the integrability equations for q_{2e} become

$$\begin{aligned}
q_{2e,\tau}=&\frac{1}{4}(1+a)^2q_{2,\tau}+\frac{\delta}{4}(1-a)^2\\
&\times\left[\frac{1}{\Psi^2}(\Phi^2-\Psi^2)\Phi_\sigma+2\frac{\Phi}{\Psi}\Psi_\sigma\right],\\
q_{2e,\sigma}=&\frac{1}{4}(1+a)^2q_{2,\sigma}+\frac{\Delta}{4}(1-a)^2\\
&\times\left[\frac{1}{\Psi^2}(\Phi^2-\Psi^2)\Phi_\tau+2\frac{\Phi}{\Psi}\Psi_\tau\right].
\end{aligned}$$

For the proof of this theorem, up to some minor changes in the notation, we would like to refer to the detailed analysis of CX in Ref. 6.

It can easily be seen that for $a=1$ the EM solution obtained reduces to the vacuum solution that describes colliding pure gravitational waves. Similarly, for $a=0$, the solution describes colliding pure em waves in which the gravitational curvatures arise only due to the existing em field. We shall proceed now by employing the vacuum solution (16) in the Theorem to construct the electrovacuum solution that describes colliding waves in the EM theory. The remarkable feature of the vacuum solution (16) is that when it is substituted into the expressions (9), (10), and (17) it yields significant reductions and cancellations. In particular, the equations that determine q_{2e} and $\nu+\mu_3$ for the vacuum take the forms

$$\begin{aligned}
&q_{2e,\tau}=(1+a^2)\sin\alpha\,\delta X_\sigma,\\
&q_{2e,\sigma}=(1+a^2)\sin\alpha\,\Delta X_\tau,
\end{aligned} \tag{18}$$

and

$$\begin{aligned}
(\nu+\mu_3+\ln\Psi)_\tau=&\frac{2\tau}{\delta-\Delta}+\frac{\tau}{\Delta}+\frac{2\delta}{\delta-\Delta}[2\sigma\,\Delta X_\tau X_\sigma\\
&-\tau(\Delta X_\tau^2+\delta X_\sigma^2)],\\
(\nu+\mu_3+\ln\Psi)_\sigma=&\frac{2\sigma}{\Delta-\delta}+\frac{2\Delta}{\Delta-\delta}[2\tau\,\delta X_\tau X_\sigma\\
&-\sigma(\Delta X_\tau^2+\delta X_\sigma^2)],
\end{aligned} \tag{19}$$

in which Ψ is given in (16).

It is observed from (18) that the integrability condition for q_{2e} amounts to the same Euler–Darboux equation satisfied by X in (14). This form for q_{2e} also provides for us a

justification for interpreting α as a measure of polarization, since, when $\alpha = 0$, it leads to $q_{2e} = 0$. We complete our expressions by giving Ω^2 and χ,

$$\frac{1}{2}\Omega^2 = \frac{\sqrt{1+\sin^2\alpha}(1+a^2)\cosh 2X - 2a\sinh 2X + 1 - a^2}{\sqrt{1+\sin^2\alpha}\cosh 2X - \sinh 2X}, \tag{20}$$

$$\chi = \sqrt{\Delta\delta}\quad(\sqrt{1+\sin^2\alpha}\cosh 2X - \sinh 2X). \tag{21}$$

Expressions (18)–(21) provide the necessary information to be used in the family of electrovacuum solutions in terms of an arbitrary harmonic function X. Taking $a = 0$ and $a = 1$ the family reduces to the pure em and pure gravitational (vacuum) limits, respectively, as it should. The remaining task now is to make particular choices for X and find the metric functions explicitly.

III. PARTICULAR SOLUTIONS

It is well known that a general class of separable solutions for the Euler–Darboux equation (14) is given by

$$X(\tau,\sigma) = \sum_n [a_n P_n(\tau)P_n(\sigma) + b_n Q_n(\tau)Q_n(\sigma) + c_n P_n(\tau)Q_n(\sigma) + d_n P_n(\sigma)Q_n(\tau)], \tag{22}$$

where P and Q are the Legendre functions of the first and second kinds, respectively, and a_n, b_n, c_n, and d_n are arbitrary constants. Now, we shall show that the simplest nontrivial solutions for X lead to the previously known solutions.

(a) The choice $P_0 = 1$, $Q_0 = \frac{1}{2}\ln[(1+\sigma)/(1-\sigma)]$, $c_0 = 1$, $c_n = 0$ $(n \neq 0)$, $a_n = b_n = d_n = 0$. (This is equivalent to $X = \tanh^{-1}\sigma$.) Also, in order to obtain a pure em solution we make the choice $a = 0$ and the vacuum functions for this case become

$$\Psi - i\Phi = [1 - \sigma^2 - i(1+\sigma^2)\sin\alpha]/[\sqrt{1+\sin^2\alpha}(1+\sigma^2) - 2\sigma],$$
$$\chi = \sqrt{\Delta/\delta}[\sqrt{1+\sin^2\alpha}(1+\sigma^2) - 2\sigma], \tag{23}$$
$$e^{\nu+\mu_3} = [\sqrt{1+\sin^2\alpha}(1+\sigma^2) - 2\sigma]/\sqrt{\Delta},$$
$$q_2 = \tau\sin\alpha,$$

and

$$\tfrac{1}{2}\Omega^2 = [\sqrt{1+\sin^2\alpha}(1+\sigma^2) + (1-\sigma^2)]/[\sqrt{1+\sin^2\alpha}(1+\sigma^2) - 2\sigma].$$

Applying the Theorem to this vacuum metric, we obtain the line element

$$ds^2 = F\left(\frac{d\tau^2}{\Delta} - \frac{d\sigma^2}{\delta}\right) - \Delta F\,dy^2 - \frac{\delta}{F}(dx - \tau\sin\alpha\,dy)^2, \tag{24}$$

where

$$2F = \sqrt{1+\sin^2\alpha}(1+\sigma^2) + 1 - \sigma^2.$$

This solution, with a different parametrization, $\tan\theta = \sin\alpha$, was reported previously.[2] For $\alpha = 0$ it reduces to the BS solution and therefore it describes colliding em shock waves with the second polarization. In Appendix A we show that it belongs to a class of particular type-D space-times. Also, in Appendix B we formulate the problem of colliding pure em waves as a variational principle from a Lagrangian. CX[8] also gave a solution for the EM theory as a generalization of the BS solution with second polarization. Their solution, however, involves two essential parameters and it represents colliding dual rotated em shocks accompanied by gravitational waves. In this sense the solution (24) is the only available solution to date that represents colliding pure em waves with the second polarization.

We obtain a new solution now by relaxing $a = 0$ and considering $0 \leqslant a \leqslant 1$. Applying the Theorem to the same seed vacuum metric that led us to (24), we obtain

$$q_{2e} = (1+a^2)\tau\sin\alpha,$$
$$\chi_e = \tfrac{1}{4}\Omega^2\sqrt{\Delta/\delta}[\sqrt{1+\sin^2\alpha}(1+\sigma^2) - 2\sigma],$$
$$e^{(\nu+\mu_3)_e} = \tfrac{1}{4}(\Omega^2/\sqrt{\Delta})[\sqrt{1+\sin^2\alpha}(1+\sigma^2) - 2\sigma], \tag{25}$$

where

$$\frac{1}{2}\Omega^2 = \frac{\sqrt{1+\sin^2\alpha}(1+a^2)(1+\sigma^2) + (1-a^2)(1-\sigma^2) - 4a\sigma}{\sqrt{1+\sin^2\alpha}(1+\sigma^2) - 2\sigma}.$$

One can easily show that this solution reduces to (24) in the limit $a = 0$. Also, in the linear polarization limit $\alpha = 0$ $(a \neq 0)$ it gives us a metric that describes colliding em shock waves of BS coupled with gravitational waves.

(b) The choices, $c_0 = (k_1 - k_2)/4, d_0 = (k_1 + k_2)/4$ (all other coefficients zero), is equivalent from (22) to

$$e^{2X} = [(1+\tau)/(1-\tau)]^{(k_1+k_2)/4} \times [(1+\sigma)/(1-\sigma)]^{(k_1-k_2)/4}. \tag{26}$$

This leads to the Szekeres[9] family of solutions for colliding gravitational waves with linear polarization.

We have shown recently that by making use of this expression for X we obtain a family of solutions with the second polarization.[10] Here now, we shall obtain from the Theorem an electrovacuum generalization of this family. In order to employ this family we modify the definition of (τ,σ) coordinates in accordance with

$$\tau = u^{n_1/2}\sqrt{1-v^{n_2}} + v^{n_2/2}\sqrt{1-u^{n_1}}, \tag{27}$$
$$\sigma = u^{n_1/2}\sqrt{1-v^{n_2}} - v^{n_2/2}\sqrt{1-u^{n_1}},$$

where the parameters (n_1,n_2) are related to (k_1,k_2) by

$$k_i^2 = 8(1 - 1/n_i) \quad (i = 1,2). \tag{28}$$

The vacuum solution describing CGW's with second polarization is given by the functions[10]

$$\chi = \sqrt{\Delta\delta}/\Psi,$$
$$q_2 = \tfrac{1}{2}\sin\alpha[(k_1 - k_2)\tau + (k_1 + k_2)\sigma], \tag{29}$$
$$e^{\nu+\mu_3} = (1/\Psi)(\tau+\sigma)^{1-k_1^2/4}(\tau-\sigma)^{1-k_2^2/4}$$
$$\times(1-\tau^2)^{-1/2+(k_1+k_2)^2/16}(1-\sigma^2)^{(k_1-k_2)^2/16},$$

where

$$\Psi = 1/(\sqrt{1+\sin^2\alpha}\cosh 2X - \sinh 2X)$$

and X is given by (26). Generalization to colliding EM waves follows immediately from the Theorem,

$$\chi_e = (\Omega^2/4)\chi,$$
$$q_{2e} = \tfrac{1}{2}(1+a^2)q_2, \tag{30}$$
$$e^{(\nu+\mu_3)_e} = (\Omega^2/4)e^{\nu+\mu_3},$$

where Ω is given by (20), in which X is expression (26).

IV. CONCLUDING REMARKS

Since the solutions are parametrized in terms of the solutions of a linear Euler–Darboux equation the family presented in this paper is quite large. Starting from a diagonal seed metric for CGW's our method enables us to obtain a cross-polarized electrovacuum solution for colliding EM waves. The simplest member of the family is shown to generalize the BS solution to the case with second polarization. The same solution was obtained by a method that made use of an analogy with cylindrical waves and also involved tedious integrations of the BS equations. The fact that this turns out to be a type-D metric is not surprising since it has already been pointed out by CX[8] that there is a connection between type-D space-times and the absence of curvature singularities in colliding waves.

Further extension of our method to the case of parametrizing the Ernst potential of CGW's by n harmonic functions seems possible, however, we maintain that the physical significance of solutions is much more important than their mathematical existence.

APPENDIX A: PROPERTIES OF THE CROSS POLARIZED BS METRIC

The BS line element that describes colliding em waves is

$$ds^2 = 2e^{-M}du\,dv - e^{-U}[e^V\cosh W dx^2$$
$$+ e^{-V}\cosh W dy^2 - 2\sinh W dx\,dy], \tag{A1}$$

in which all metric functions depend on the null coordinates u and v. The solution (24) was obtained by a different method in the coordinates

$$\tau = \sin(au+bv), \quad \sigma = \sin(au-bv), \tag{A2}$$

where a and b are constants. The BS metric functions and the em field strengths are given as follows:

$$e^{-M} = \Sigma, \quad e^{-U} = \sqrt{1-\tau^2}\sqrt{1-\sigma^2},$$
$$\sinh W = (\tau/\Sigma)\sqrt{(1-\sigma^2)/(1-\tau^2)}\sin\alpha,$$
$$e^{-V}\tanh W = \tau\tan\alpha,$$
$$\phi_2 = (a/\sqrt{k})(\cos\alpha/\Sigma)^{1/2}e^{i\gamma}, \tag{A3}$$
$$\phi_0 = (b/\sqrt{k})(\cos\alpha/\Sigma)^{1/2}e^{i\epsilon},$$

where

$$\Sigma = \cos^2(\alpha/2) + \sigma^2\sin^2(\alpha/2),$$
$$\sin(\gamma-\epsilon) = \tanh W, \tag{A4}$$
$$\tan[(\gamma+\epsilon)/4] = \sigma\tan(\alpha/2),$$

and we note that we have adopted the notation of Ref. 2 with $(\theta\to\alpha)$. In order to see the difference between this solution and the BS solution, we evaluate the electric (**E**) and the magnetic (**H**) fields in an orthonormal frame $\{\omega^a\}$, where

$$ds^2 = (d\omega^0)^2 - (d\omega^i)^2.$$

The field components are

$$E_x = F_{02} = \mathrm{Re}(\phi_0 - \phi_2),$$
$$H_y = F_{12} = -\mathrm{Re}(\phi_0 + \phi_2), \tag{A5}$$
$$E_y = F_{03} = \mathrm{Im}(\phi_0 + \phi_2),$$
$$H_x = F_{31} = \mathrm{Im}(\phi_0 - \phi_2),$$

which imply, in contrast to the BS case, that both field invariants $\mathbf{E}\cdot\mathbf{H}$ and $\mathbf{E}^2 - \mathbf{H}^2$ are nonzero.

In order to calculate the curvature components we make use of the CX line element

$$ds^2 = \Sigma(d\psi^2 - d\theta^2) - [(\sin\psi\sin\theta/1-\epsilon\bar{\epsilon}]$$
$$\times|(1-\epsilon)dx + i(1+\epsilon)dy|^2. \tag{A6}$$

Here,

$$\epsilon = (Z-1)/(Z+1),$$
$$Z = \sin\theta(\Sigma\sin\psi - i\sin\alpha\sin\theta\cos\psi)^{-1}; \tag{A7}$$

Σ is given in (A4) and we have chosen $a = 1 = b$, such that the new coordinates (ψ,θ) are related to (τ,σ) by

$$\tau = \cos\psi, \quad \sigma = \cos\theta. \tag{A8}$$

After significant reduction the nonvanishing Weyl and Maxwell scalars are found as follows:

$$\Psi_2 = R, \quad \Psi_4 = -3e^{i\lambda}R, \quad \Psi_0 = -3\,\mathrm{Re}^{-i\lambda},$$
$$2\phi_{00} = 2\phi_{22} = \cos\alpha/\Sigma^2, \tag{A9}$$
$$-2\phi_{20} = e^{i\lambda}\cos\alpha/\Sigma^2,$$

where

$$R = \frac{\sin(\alpha/2)}{2\Sigma}\frac{\sin(\alpha/2) - i\cos\theta\cos(\alpha/2)}{[\cos(\alpha/2) + i\cos\theta\sin(\alpha/2)]^2},$$
$$e^{i\lambda} = \frac{\sin\theta + \Sigma\sin\psi + i\sin\alpha\sin\theta\cos\psi}{\sin\theta + \Sigma\sin\psi - i\sin\alpha\sin\theta\cos\psi}. \tag{A10}$$

From these expressions it is readily seen that the relations

$$9\Psi_2^2 = \Psi_0\Psi_4,$$
$$3\phi_{00}\Psi_2 = \phi_{20}\Psi_0, \tag{A11}$$

hold; therefore the solution (A3) belongs to a class of type-D metrics. On this account we make two successive null-tetrad rotations[8] to make the principal null directions of the Weyl and Maxwell tensors coincident. As a result of these transformations we obtain the principal values of the invariants as

$$\Psi_2 = -2R, \quad \Psi_0 = \Psi_1 = \Psi_3 = \Psi_4 = 0,$$
$$2\phi_{11} = \cos\alpha/\Sigma^2, \quad \phi_0 = \phi_2 = 0. \tag{A12}$$

By choices of the polarization angle α these invariants can be shown to be free of singularities.

APPENDIX B: THE VARIATIONAL PRINCIPLE

The problem of colliding pure em waves can be formulated as a variational principle, provided we introduce two potential functions as follows. We define the em potential one-form by

$$\tilde{A} = \tilde{A}_\mu \, dx^\mu = A \, dx + B \, dy, \tag{B1}$$

where A and B are the components in the Killing directions. The em field two-form is given, accordingly, by

$$F = d\tilde{A}, \tag{B2}$$

where d stands for the exterior derivative. The null-tetrad basis one-forms are chosen as

$$l = e^{-M/2} \, du, \quad n = e^{-M/2} \, dv, \tag{B3}$$

$$m = \frac{e^{-U/2}}{\sqrt{2}} \left\{ e^{V/2} \left(i \sinh \frac{W}{2} - \cosh \frac{W}{2} \right) dx + e^{-V/2} \left(\sinh \frac{W}{2} - i \cosh \frac{W}{2} \right) dy \right\},$$

and the field two-form is given by

$$F = \phi_2 \, l \wedge m + \bar{\phi}_2 \, l \wedge \bar{m} - \bar{\phi}_0 \, n \wedge m - \phi_0 \, n \wedge \bar{m}, \tag{B4}$$

in which $\wedge$ is the wedge product. The dual basis two-forms are

$$*(l \wedge m) = -il \wedge m, \quad *(n \wedge m) = in \wedge m,$$
$$*(l \wedge \bar{m}) = il \wedge \bar{m}, \quad *(n \wedge \bar{m}) = -in \wedge \bar{m}, \tag{B5}$$

so that the dual field two-form becomes

$$-i*F = -\phi_2 \, l \wedge m + \bar{\phi}_2 \, l \wedge \bar{m} - \bar{\phi}_0 \, n \wedge m + \phi_0 \, n \wedge \bar{m}. \tag{B6}$$

The sourceless Maxwell equations,

$$dF = 0 = d*F, \tag{B7}$$

are satisfied everywhere, including the null boundaries. The em field strengths ϕ_0 and ϕ_2 are expressed in terms of the metric functions and the potentials by

$$\phi_2 = \frac{e^{U/2}}{\sqrt{2}} \left\{ e^{-V/2} \left(i \sinh \frac{W}{2} - \cosh \frac{W}{2} \right) A_u + e^{V/2} \left(i \cosh \frac{W}{2} - \sinh \frac{W}{2} \right) B_u \right\},$$
$$\phi_0 = \frac{e^{U/2}}{\sqrt{2}} \left\{ e^{-V/2} \left(i \sinh \frac{W}{2} + \cosh \frac{W}{2} \right) A_v + e^{V/2} \left(i \cosh \frac{W}{2} + \sinh \frac{W}{2} \right) B_v \right\}. \tag{B8}$$

As a result of all these expressions it can be checked by direct calculation that a suitable Lagrangian that describes colliding em waves is given by

$$L = e^{-U} (M_u U_v + M_v U_u + U_u U_v - V_u V_v \cosh^2 W), - 2k\{(B_u B_v e^V + A_u A_v e^{-V}) \times \cosh W + (A_u B_v + A_v B_u) \sinh W\}, \tag{B9}$$

where k is a coupling constant. BS equations now follow from the variational principle

$$\delta I = 0, \quad I = \int L \, du \, dv. \tag{B10}$$

[1] P. Bell and P. Szekeres, Gen. Relativity Gravitation **5**, 275 (1974).
[2] M. Halilsoy, Phys. Rev. D **37**, 2121 (1988).
[3] R. A. Matzner and F. Tipler, Phys. Rev. D **29**, 1575 (1984).
[4] M. Halilsoy, Phys. Rev. D **39**, 2172 (1989).
[5] M. Halilsoy, Lett. Nuovo Cimento **37**, 231 (1983).
[6] S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **398**, 209 (1985).
[7] F. J. Ernst, Phys. Rev. **168**, 1415 (1968).
[8] S. Chandrasekhar and B. C. Xanthopoulos, Proc. R. Soc. London Ser. A **410**, 311 (1987).
[9] P. Szekeres, J. Math. Phys. **13**, 286 (1972).
[10] M. Halilsoy, Phys. Rev. D **38**, 2979 (1988).

Effective actions with vector-axial-vector couplings on Riemann–Cartan manifolds

Guido Cognola
Dipartimento di Fisica, Università di Trento, 38050 Povo (TN), Italy and Istituto Nazionale di Fisica Nucleare, Gruppo Collegato di Trento, Trento, Italy

Paola Giacconi and Roberto Soldati
Dipartimento di Fisica, Università di Bologna, Via Irnerio 46, 40126 Bologna, Italy and Istituto Nazionale di Fisica Nucleare, Sezione di Bologna, Bologna, Italy

(Received 14 February 1990; accepted for publication 30 May 1990)

The general forms of quantum effective actions in the presence of vector and axial-vector gauge potentials on Riemann–Cartan manifolds with torsion are discussed. The relations between different forms of anomalies are elucidated.

I. INTRODUCTION

In recent years anomalies have played an increasingly important role in the study of field theories and particles physics. Anomalies appear when some symmetry at the classical level gets lost after the quantization process. A complete understanding of anomalies is crucial to implement those symmetries in physical problems.

The discovery of the anomalous breaking of the chiral and gauge symmetries has been done in the framework of a perturbative analysis based on Feynman diagrams.[1–4]

Later on, Fujikawa has suggested a way to describe chiral anomalies within the context of a path-integral formalism.[5,6] In this approach anomalies are related to the noninvariance of the fermionic functional measure with respect to the corresponding symmetry group of transformations and therefore, eventually, to Jacobian functional determinants.

Although heuristically powerful, that procedure is not completely satisfactory as it involves formal manipulations with ill-defined mathematical quantities; as a matter of fact it has been shown that, in general, it gives rise to controversial or ambiguous results.[7] A very convenient and rigorous method to treat functional determinants is provided by the ζ-function regularization,[8,9] which allows a suitable definition of the effective actions as path integrals. Following this method, the well-known result of Bardeen[3] for the non-Abelian consistent axial anomaly has been recovered.[10] The above technique can be suitably generalized in order to deal with curved spaces as shown in Ref. 11; in this work the consistent and covariant forms of the non-Abelian anomalies have been computed in terms of the coefficients of the heat kernel asymptotic expansion on a Riemann–Cartan manifold with a general nonvanishing torsion.

It is the aim of the present note to clarify the relation between the consistent and covariant forms of the non-Abelian anomalies, within the context of the ζ-function regularization method and in the presence of general vector and axial-vector gauge potentials on Euclidean four-dimensional Riemann–Cartan manifolds. As a matter of fact, it is known[12] that the covariant form of the anomaly for a left-(right) handed chiral coupling in a flat space-time can be obtained from the consistent one by adding to the expectation value of the left (right) current $\langle J^k_{L(R)} \rangle$ a local polynomial functional of the gauge potential. The resulting current $\langle \tilde{J}^k_{L(R)} \rangle$ is determined by the requirement of being covariant under local gauge transformations, in such a way that its covariant divergence is also covariant. This means that it is always possible to set up the covariant form of the anomaly from the knowledge of the consistent one and vice versa (modulo trivial cocycles), within a given regularization scheme. Morever, it is claimed[12] that only the consistent form correspond to a local variation of an effective action functional, as it turns out to satisfy the Wess–Zumino consistency conditions.

In the present work we will show that, on general ground, both forms of the non-Abelian anomaly do indeed correspond to local variations of suitable effective actions. Furthermore we explicitly derive the expression of the local functional relating the consistent form of the nonsinglet anomaly in the vector-axial-vector theory to the covariant one, for the gauge theory of a massless Dirac field with arbitrary internal degrees of freedom, on a Euclidean four-dimensional curved space and in the presence of an external arbitrary torsion field. The expression we obtain reduces to the one of Bardeen and Zumino[12] in a suitable limit.

II. THE CLASSICAL MODEL

Our starting point is the classical Euclidean action, for a Dirac fermion ψ on a four-dimensional curved space with arbitrary torsion (for more details on this model see Ref. 11)

$$S = \int \left[i\psi^\dagger \gamma^k \left(\frac{1}{2} \nabla_k + V_k + \gamma^5 A_k \right) \psi \right] \sqrt{g}\, d^4x, \qquad (1)$$

where

$$\nabla_k \psi = (\partial_k + \tfrac{1}{4} \Gamma^{\mu\nu}_k \sigma_{\mu\nu}) \psi, \qquad (2)$$

$i,j,...$, and $\mu,\nu,...$, being holonomic and anholonomic indices, respectively, g_{jk} a metric with signature $(+,+,+,+)$ and determinant g, $\gamma^\mu = (\gamma^\mu)^\dagger$, $\gamma^5 = (\gamma^5)^\dagger$ and $\sigma^{\mu\nu} = (1/2)[\gamma^\mu,\gamma^\nu]$ the usual Euclidean Dirac matrices and $\Gamma^{\mu\nu}_k$ the Cartan connection. For vector and axial-vector gauge potentials we use the matrix notation $V_k = -iV^a_k \tau_a$, $A_k = -iA^a_k \tau_a$, τ_a being the Hermitean generators of the gauge group, in the given representation ($a = 1,...,N$).

 0022-2488/90/112699-03$03.00

We notice that, even if we choose ψ and $\psi^\dagger$ as Euclidean conjugate spinors belonging to a unitary representation of a compact group, the classical action of Eq. (1) does develop an immaginary part due to presence of the axial-vector coupling. As a consequence it is quite natural to consider a general linear complex group as the gauge group. From Eq. (1), we obtain the field equations

$$i\gamma^k(\nabla_k + V_k + \gamma^5 A_k + \tfrac{1}{2}S_k)\psi = \not{D}\psi = 0, \tag{3}$$

S_k being the trace of the torsion tensor. The classical action is invariant with respect to vector and axial-vector general linear complex transformations. In the infinitesimal form they read ($\epsilon(x) = \epsilon^a(x)\tau_a$):

$$\delta_v\psi = -\epsilon\psi; \quad \delta_v V_k = \partial_k\epsilon + [V_k,\epsilon],$$
$$\delta_v\bar{\psi} = \bar{\psi}\epsilon; \quad \delta_v A_k = [A_k,\epsilon], \tag{4}$$

$$\delta_A\psi = -\gamma^5\epsilon\psi; \quad \delta_A A_k = \partial_k\epsilon + [V_k,\epsilon],$$
$$\delta_A\bar{\psi} = -\bar{\psi}\gamma^5\epsilon; \quad \delta_A V_k = [A_k,\epsilon]. \tag{5}$$

The corresponding classical conservation laws can be written in the form

$$\hat{\nabla}_k J^k + [V_k,J^k] + [A_k,{}^5J^k] = 0, \tag{6}$$

$$\hat{\nabla}_k {}^5J^k + [V_k,{}^5J^k] + [A_k,J^k] = 0, \tag{7}$$

where $J^k = \bar{\psi}\gamma^k\tau_a\psi\tau^a$ and ${}^5J^k = \bar{\psi}\gamma^k\gamma^5\tau_a\psi\tau^a$ are the vector and axial-vector currents, respectively, while $\hat{\nabla}_k$ is the covariant derivative with respect to the Levi-Civita connection. As it is well known, in the quantization process the conservation laws may acquire an anomalous term. This means that the right-hand sides of the quantum version of Eqs. (6) and (7) can be, in general, different from zero, that is equal to $\mathscr{A}_a\tau^a$ and ${}^5\mathscr{A}_a\tau^a$, respectively. The values of these quantities can be obtained once a precise definition for the effective action has been given.

III. THE EFFECTIVE ACTION

Let us now define the generating functional $W[V,A,g]$ for connected Green's functions as

$$W[V,A,g] = \frac{1}{2}\frac{d}{ds}\zeta(H,s)|_{s=0}, \tag{8}$$

where $\zeta(H,s)$ is the ζ function[8,9] related to the second-order differential operator $H = \not{D}^\dagger\not{D}$ we supose to be positive definite (the possible presence of kernels in H can be accustomed with an obvious modification of the ζ-function definition of the corresponding functional determinant). It should be emphasized that, due to the spectral theorem, the present definition corresponds, in the path-integral formulation, to a nonlocal action in the exponential integrand, in contrast with the classical one of Eq. (1). By means of the transformations (4), (5) we get

$$\frac{\delta_V}{\delta\epsilon^a}W[V,A,g] = \mathscr{A}_a = \frac{\delta_V}{2\delta\epsilon^a}\zeta'(H,0), \tag{9}$$

$$\frac{\delta_A}{\delta\epsilon^a}W[V,A,g] = {}^5\mathscr{A}_a = \frac{\delta_A}{2\delta\epsilon^a}\zeta'(H,0). \tag{10}$$

We shall study the variation of the effective action of Eq. (8) under infinitesimal vector and axial-vector transformations. For the Dirac operator we have

$$\delta_V\not{D} = -[\not{D},\hat{\epsilon}], \tag{11}$$

$$\delta_A\not{D} = -\{\not{D},\gamma^5\hat{\epsilon}\}, \tag{12}$$

whence

$$\delta_V H = 2\not{D}^\dagger\hat{\epsilon}\not{D} - \not{D}^\dagger\not{D}\hat{\epsilon} - \hat{\epsilon}\not{D}^\dagger\not{D}, \tag{13}$$

$$\delta_A H = -2\not{D}^\dagger\gamma^5\hat{\epsilon}\not{D} - \not{D}^\dagger\not{D}\gamma^5\hat{\epsilon} - \gamma^5\hat{\epsilon}\not{D}^\dagger\not{D}, \tag{14}$$

with $\hat{\epsilon}\equiv\frac{1}{2}(\epsilon+\epsilon^\dagger)$; we see that only the noncompact parts of the transformations lead to nonvanishing variations. Taking into account the differentiability properties[13] of zeta function one gets:

$$\delta_V\zeta'\equiv\zeta'(0,H+\delta_V H) - \zeta'(0,H)$$
$$= -\frac{d}{ds}[s\,\mathrm{Tr}(H^{-s-1}\delta_V H)]$$
$$= -\frac{d}{ds}[s\,\mathrm{Tr}(H^{-s-1}\not{D}\hat{\epsilon}\not{D}) - \mathrm{Tr}(H^{-s}\hat{\epsilon})]. \tag{15}$$

If we denote with ϕ_n and Φ_n a complete set of orthonormal functions satisfying:

$$H\Phi_n = \not{D}^\dagger\not{D}\Phi_n = \lambda_n\Phi_n, \tag{16}$$

$$\not{D}\not{D}^\dagger\phi_n = \lambda_n\phi_n, \tag{17}$$

where

$$\not{D}\Phi_n = \sqrt{\lambda_n}, \tag{18}$$

$$\not{D}^\dagger\phi_n = \sqrt{\lambda_n}\Phi_n, \tag{19}$$

we can write

$$\delta_v\zeta' = -\frac{d}{ds}s\Big[\sum_n\sum_m\langle\Phi_n,(\not{D}^\dagger\not{D})^{-s-1}\Phi_m\rangle$$
$$\times\langle\Phi_m,\not{D}^\dagger\hat{\epsilon}\not{D}\Phi_m\rangle - \langle\Phi_n,(\not{D}^\dagger\not{D})^{-s-1}\hat{\epsilon}\Phi_n\rangle\Big]$$
$$= \frac{d}{ds}s\,\mathrm{Tr}[(\not{D}^\dagger\not{D})^{-s} - (\not{D}\not{D}^\dagger)^{-s}]\hat{\epsilon}, \tag{20}$$

so that, taking into account the value in zero of the ζ function is related (in four dimensions) to the second coefficient $a_2(H,x)$ of the heat kernel expansion,[9] we obtain

$$\tilde{\mathscr{A}}_a = (1/16\pi^2)\mathrm{tr}\{\tau_a[a_2(\not{D}^\dagger\not{D},x) - a_2(\not{D}\not{D}^\dagger,x)]\}, \tag{21}$$

$${}^5\tilde{\mathscr{A}}_a = (1/16\pi^2)\mathrm{tr}\{\tau_a\gamma^5[a_2(\not{D}^\dagger\not{D},x) + a_2(\not{D}\not{D}^\dagger,x)]\}. \tag{22}$$

These formulas provide the "covariant" expression of the anomalies.

The explicit computation of the heat kernel expansion coefficients and the detailed derivation of the anomalies can be found in Ref. 11. Here, we simply report the final expressions that are necessary to our purposes. Using Eqs. (21) and (22) we have:

$$\tilde{\mathscr{A}}_a = (1/8\pi^2)\mathrm{tr}[F^*G + G^*F]\tau_a, \tag{23}$$

$${}^5\tilde{\mathscr{A}}_a = (1/8\pi^2)\mathrm{tr}[F^*F + G^*G]\tau_a, \tag{24}$$

where

$$F_{ij} = \partial_i V_j - \partial_j V_i + [V_i, V_j] + [A_i, A_j],$$
$$G_{ij} = \partial_i A_j - \partial_j A_i + [A_i, A_j]$$

and the dual quantities are built up by means of the completely antisymmetric Levi-Civita symbol $e_{\mu\nu\alpha\beta}$, that is ${}^*F_{\mu\nu} = (1/2)e_{\mu\nu\alpha\beta}F^{\alpha\beta}$ and so on. These expressions of the anomalies are formally the same that one obtains in the flat space[14] and, when the chiral limit $V/2 = A/2 = V'$ is taken, they give rise to the covariant form of the anomalies.

It is not difficult to verify that the choice $\not{D}\not{D}^\dagger$ in place of $\not{D}^\dagger\not{D}$ to define the effective action gives rise to the same result for the anomaly. It is important to remark that an alternative definition for the effective action is provided through the operator $H = \not{D}^2$. We would like to remark that, in order to implement this definition, it is necessary to perform the replacement $A_k \to iA_k$ in such a way to obtain a self-adjoint operator $\not{D}(V_k, iA_k)$.[14,15] Notice that, even in this case, the group of invariance of the analitically continued classical action is still the general linear complex group of vector and axial-vector local transformations.

Using Eq. (9) and recalling again the properties of the ζ function, we get:[16]

$$\mathscr{A}_a = 0, \tag{25}$$
$${}^5\mathscr{A}_a = (1/8\pi^2)\mathrm{tr}[\tau_a \gamma^5 a_2(\not{D}^2, x)]. \tag{26}$$

These formulas give an expression for the axial anomaly satisfying the Wess–Zumino integrability conditions.[17] Notice in fact that, in the present case, the effective action is related to a path integral where the exponential integrand contains the analitically continued classical action. From Eqs. (25) and (26), following the method of Ref. 11 one obtains:

$$\mathscr{A}_a = 0, \tag{27}$$

$${}^5\mathscr{A}_a = \frac{1}{8\pi^2}\mathrm{tr}\left[F{}^*F + \frac{1}{3}G{}^*G + \frac{16}{3}A^\mu A^\nu A^\alpha A^\beta e_{\mu\nu\alpha\beta}\right] - \frac{8}{3}[{}^*F_{ij}A^i A^j + A^i A^j {}^*F_{ij} + A^i {}^*F_{ij} A^j]\tau_a. \tag{28}$$

We observe that the above expressions for the anomalies are formally the same as obtained by Bardeen[3] in the flat space-time. They satisfy the usual Wess–Zumino integrability conditions,[17] at variance with the corresponding covariant ones as given in Eqs. (23) and (24). The reason for this feature is due to the fact that the covariant form of the anomalies arises, as we have seen, from the infinitesimal noncompact variation of a generating functional corresponding to a nonlocal action [see Eq. (8)]. We would like to remark that, concerning the above last point, there are inaccurate and confusing statements in the recent literature.[18]

In Ref. 12 it has been shown that it is possible to obtain the covariant forms of the current and of the anomaly from the knowledge of the consistent ones in the chiral case. Here, we explicitly set up, in the present general case, the local polynomials χ^k and ${}^5\chi^k$ of the gauge potentials so as to define two new currents $\tilde{J}^k = J^k + \chi^k$ and ${}^5\tilde{J}^k = {}^5J^k + {}^5\chi^k$ satisfying the conservation laws:

$$\hat{\nabla}_k \tilde{J}^k + [V_k, \tilde{J}^k] + [A_k, {}^5\tilde{J}^k] = \tilde{\mathscr{A}}_a \tau^a, \tag{29}$$
$$\hat{\nabla}_k {}^5\tilde{J}^k + [V_k, {}^5\tilde{J}^k] + [A_k, \tilde{J}^k] = {}^5\tilde{\mathscr{A}}_a \tau^a. \tag{30}$$

We recall that J^k and ${}^5J^k$ satisfy similar conservation laws, but with $\mathscr{A}_a$ and ${}^5\mathscr{A}_a$ in place of $\tilde{\mathscr{A}}_a$ and ${}^5\tilde{\mathscr{A}}_a$, respectively. This means that χ and ${}^5\chi$ fulfill the relations

$$\hat{\nabla}_k \chi^k + [V_k, \chi^k] + [A_k, {}^5\chi^k] = (\tilde{\mathscr{A}}_a - \mathscr{A}_a)\tau^a, \tag{31}$$
$$\hat{\nabla}_k {}^5\chi^k + [V_k, {}^5\chi^k] + [A_k, \chi^k] = ({}^5\tilde{\mathscr{A}}_a - {}^5\mathscr{A}_a)\tau^a. \tag{32}$$

Then we easily get, after some algebra

$$\chi^k_a = -(1/8\pi^2)\mathrm{tr}\,\tau_a\left[\tfrac{2}{3}({}^*G^{kj}A_j + A_j {}^*G^{kj})\right], \tag{33}$$

$${}^5\chi^k_a = -(1/8\pi^2)\mathrm{tr}\,\tau_a \times\left[2({}^*F^{kj}A_j + A_j {}^*F^{kj}) - \tfrac{8}{3}e^{kjrs}A_j A_r A_s\right]. \tag{34}$$

The known results of Bardeen and Zumino[12] can be reached by taking the chiral limit $V/2 = A/2 = V'$ and reducing to the flat space case.

IV. CONCLUSION

To sum up we have seen that, following the ζ-function regularization method to set up effective actions, it is possible to derive the covariant and consistent forms of the non-singlet anomalies in the general vector-axial-vector theory on Riemann–Cartan manifolds as infinitesimal variations with respect to general linear groups of gauge transformations. It is important to appreciate that, in a quite general framework, both forms do indeed arise as limits of local functional derivatives of suitable ζ-regularized effective actions. Moreover, it is possible to obtain the explicit expression of the local functional that relates, in the present general case, the two above-mentioned forms of the anomalies; that functional correctly reduces to the Bardeen–Zumino one in the flat space and chiral limits.

[1] S. Adler, Phys. Rev. **177**, 1426 (1969).
[2] J. Bell and R. Jackiw, Nuovo Cimento A **60**, 47 (1969).
[3] W. A. Bardeen, Phys. Rev. **184**, 1848 (1969).
[4] D. J. Gross and R. Jackiw, Phys. Rev. D **6**, 477 (1972).
[5] K. Fujikawa, Phys. Rev. Lett. **42**, 1195 (1975).
[6] K. Fujikawa, Phys. Lett. **42**, 1195 (1979).
[7] R. E. Gamboa Saravi, M. A. Muschietti, F. A. Shaposnik, and J. E. Solomin, Ann. Phys. **157**, 360 (1984).
[8] J. S. Dowker and R. Critchley, Phys. Rev. D **13**, 3224 (1976).
[9] S. W. Hawking, Comm. Math. Phys. **55**, 133 (1977).
[10] A. P. Balachandran, G. Marmo, V. P. Nair, and C. G. Trahern, Phys. Rev. D **25**, 2713 (1982).
[11] G. Cognola and P. Giacconi, Phys. Rev. D **39**, 2987 (1989).
[12] W. A. Bardeen and B. Zumino, Nucl. Phys. B **244**, 421 (1984).
[13] R. E. Gamboa-Saravi, M. A. Muschietti, F. A. Shaposnik, and J. E. Solomin, Comm. Math. Phys. **89**, 363 (1983).
[14] K. Fujikawa, *Proceedings of Kyoto Summer Institute on Quantum Gravity and Cosmology*, edited by H. Sato and T. Inami (World Scientific, Singapore, 1985).
[15] A. Andrianov and L. Bonora, Nucl. Phys. B **233**, 232 (1984).
[16] G. Cognola and S. Zerbini, Mod. Phys. Lett. A **3**, 599 (1988).
[17] J. Wess and B. Zumino, Phys. Lett. B **37**, 95 (1971).
[18] M. Reuter, Phys. Rev. D **31**, 1374 (1985).

Separation of variables and exact solution to the Dirac equation in curvilinear orthogonal coordinates with cylindrical symmetry

Víctor M. Villalba[a)]
Centro de Física, Instituto Venezolano de Investigaciones Científicas (IVIC), Apartado 21827, Caracas 1020-A, Venezuela

(Received 22 September 1989; accepted for publication 13 June 1990)

In the present article, a complete separation of variables in the Dirac equation for a free particle is achieved in parabolic cylinder and elliptical coordinates. The resulting system of ordinary differential equations is solved and the asymptotic behavior of the spinor solution is discussed.

I. INTRODUCTION

One of the most interesting problems in mathematical physics is the study of systems of partial differential equations and the search for their exact solutions which, in general, represent a different task, due to the lack of a general method[1] to study such systems.

The method of separation of variables is, perhaps one of the most successful techniques available to find exact solutions of partial differential equations because it allows us to reduce a particular problem to a system of ordinary differential equations which are more widely studied in the literature. Because of the complexity of the Dirac equation, that is, a system of four coupled partial differential equations, only a few exact solutions have been reported.[2-4] Recently, some constants of motion associated with the separation of variables in the Dirac equation in curved space-times have been characterized. Also, there has been a great interest in the study of the behavior of probe particles in the vicinity of black holes and space-time singularities.[5,6] The spinning field in the Kerr metric has also been a subject of detailed analysis and some criteria of separability of variables have been obtained by Teukolsky,[7] Carter,[8] McLennaghan,[9] and Güven,[10] among others. Chandrasekhar[11] has shown that the Dirac equation admits a separable solution in the Kerr metric, and, therefore, the Dirac equation in Minkowski space admits separable solution in oblate spheroidal coordinates. Bagrov *et al.*[12] have found new exact solutions to the Dirac equation, in the presence of electromagnetic fields in curvilinear coordinates. Such coordinates were obtained in the analysis of separability of variables in the Klein–Gordon equation.

It should be mentioned that, for a free particle, exact solutions to the Dirac equation, after separation of variables, have been reported only in Cartesian, cylindrical, and spherical coordinates.[13] Chandrasekhar's result has originated a series of works devoted to analyze the conditions of separability of variables and the peculiar nature of spheroidal coordinates. Here, the work of Cook[14] and the more recent article of Kalnins *et al.*[15] should be mentioned. The method of separation of variables, based on a complete set of first-order differential operators[16,17] has been applied in the search of new exact solutions to the Dirac equation in curvilinear coordinates in the presence of external vector fields, and the authors have paid particular attention to the problem of separability in orthogonal curvilinear coordinates when the Lamé metric functions depend on two variables, i.e., parabolic and elliptic cylinder coordinates and spheroidal coordinates. The above results were obtained in a formalism of first-order differential operators where no null tetrads are required in order to write the resulting equations.

In the present paper, we show how the method of separation of variables[16,17] enables us to find new exact solutions to the Dirac equation, for a free particle, in parabolic and elliptic cylinder coordinates. In Sec. II, the separation of variables for the Dirac equation for a free particle in the above-mentioned coordinates is achieved. In Sec. III, the equations obtained in Sec. II are solved, and the asymptotic behavior of the wave spinor is presented.

II. SEPARATION OF VARIABLES

Orthogonal cylindrical curvilinear coordinates are defined by

$$x=f(\mu,\nu),\quad y=g(\mu,\nu),\quad z=z,\quad t=t, \tag{2.1}$$

where the functions f and g satisfy the conditions

$$\frac{\partial f}{\partial \mu}=\frac{\partial g}{\partial \nu},\quad \frac{\partial f}{\partial \nu}=-\frac{\partial g}{\partial \mu}. \tag{2.2}$$

Therefore the metric form is

$$g_{\alpha\beta}=\mathrm{diag}(-1;(f_{,\mu}^2+f_{,\nu}^2);(f_{,\mu}^2+f_{,\nu}^2);1), \tag{2.3}$$

where the comma indicates partial differentiation.

Let us obtain the Dirac equation in the curvilinear coordinates (2.1), (2.2) referred to both fixed tetrads (Cartesian gauge tetrad) and rotating tetrads (diagonal gauge tetrad). In the first case, the Dirac equation takes the form

$$\{\tilde{\gamma}^{\alpha}\partial_{\alpha}+m\}\psi_c=0. \tag{2.4}$$

The connection between the Dirac matrices in the Cartesian gauge ($\tilde{\gamma}^{\alpha}$) and the constant Dirac matrices (γ^{β}) is established by

$$\tilde{\gamma}^{\alpha}=h^{\alpha}_{\beta}\gamma^{\beta}, \tag{2.5}$$

with

$$\begin{aligned}\tilde{\gamma}^1&=(\sqrt{f_{,\mu}^2+f_{,\nu}^2})^{-1}[f_{,\mu}\gamma^1+g_{,\mu}\gamma^2],\\ \tilde{\gamma}^2&=(\sqrt{f_{,\mu}^2+f_{,\nu}^2})^{-1}[f_{,\nu}\gamma^1+g_{,\nu}\gamma^2],\\ \tilde{\gamma}^3&=\gamma^3,\quad \tilde{\gamma}^0=\gamma^0,\end{aligned} \tag{2.6}$$

a) Postal address: Carmelitas 4282, Caracas 1010-A, Venezuela.

 0022-2488/90/112702-06$03.00

where α takes the values 0, 1, 2, 3 which correspond to coordinates t, μ, ν, and z, respectively.

The constant Dirac matrices (γ^β) satisfy the relation

$$\gamma^\alpha\gamma^\beta + \gamma^\beta\gamma^\alpha = 2\eta^{\alpha\beta}; \quad \eta^{\alpha\beta} = \mathrm{diag}(-1,1,1,1). \quad (2.7)$$

If the differentials are evaluated with respect to a rotating tetrad, the Dirac equation reads

$$\{\bar{\gamma}^\alpha \nabla_\alpha + m\}\psi_d = 0, \quad (2.8)$$

where ∇_α denotes the covariant differentiation defined by the relation

$$\nabla_\alpha = \partial_\alpha - \Gamma_\alpha \quad (2.9)$$

with the spinor connection Γ_α given by

$$\Gamma_\alpha = \tfrac{1}{4} g_{\mu\lambda}\{\partial_\alpha h^\kappa_\partial h^\lambda_\kappa - \Gamma^\lambda_{\partial\alpha}\} S^{\alpha\partial}, \quad (2.10)$$

where

$$S^{\alpha\partial} = \tfrac{1}{2}(\bar{\gamma}^\alpha\bar{\gamma}^\partial - \bar{\gamma}^\partial\bar{\gamma}^\alpha). \quad (2.11)$$

In the diagonal gauge, the γ matrices associated with the coordinates (2.1) when expressed in terms of the constant Dirac matrices, take the form

$$\bar{\gamma}^0 = \gamma^0, \quad \bar{\gamma}^1 = \gamma^1/h, \quad \bar{\gamma}^2 = \gamma^2/h, \quad \bar{\gamma}^3 = \gamma^3, \quad (2.12)$$

where

$$h = \sqrt{f_{,\mu}^2 + f_{,\nu}^2} \quad (2.13)$$

is the Lamé metric function.

It is easy to see that, the set of matrices (2.6) and (2.12) satisfy the same algebra

$$[\bar{\gamma}^\alpha,\bar{\gamma}^\beta]_+ = [\bar{\gamma}^\partial,\bar{\gamma}^\beta]_+ = 2g_{\alpha\beta}, \quad (2.14)$$

therefore, they are related by a transformation S

$$S^{-1}\bar{\gamma}S = \bar{\gamma}, \quad \psi_c = S\psi_d, \quad (2.15)$$

$$S = \exp\left(-\frac{1}{2}\phi\gamma^1\gamma^2\right) = \cos\frac{\phi}{2} - \gamma^1\gamma^2\sin\frac{\phi}{2}, \quad (2.16)$$

where

$$\phi = \phi(\mu,\nu) = \arctan(g_{,\mu}/f_{,\mu}).$$

In order to separate variables in Eq. (2.8), it is convenient to define a new spinor Φ

$$\Phi = (f_{,\mu}^2 + f_{,\nu}^2)^{1/4}\cdot\psi_d. \quad (2.17)$$

Then, the Dirac equation, in the diagonal gauge, for the spinor Φ reads

$$\{\gamma^0\partial_0 + (\gamma^1/h)\partial_1 + (\gamma^2/h)\partial_2 + \gamma^3\partial_3 + m\}\Phi = 0. \quad (2.18)$$

Notice that the matrices which appear in Eq. (2.18) are the constant Dirac matrices satisfying relation (2.7).

Applying the method of separation of variables, Eq. (2.18) can be written as a sum of two first-order commuting differential operators as follows:

$$\{\hat{k}_1 + \hat{k}_2\}\bar{\phi} = 0, \quad [\hat{k}_1,\hat{k}_2] = 0, \quad (2.19)$$

with

$$\hat{k}_1 = (1/h)\{\gamma^1\partial_1 + \gamma^2\partial_2\}\gamma^3\gamma^0, \quad (2.20)$$

$$\hat{k}_2 = \{\gamma^3\partial_3 + \gamma^0\partial_0 + m\}\gamma^3\gamma^0, \quad (2.21)$$

$$\bar{\phi} = \gamma^3\gamma^0\phi, \quad (2.22)$$

$$\hat{k}_1\bar{\phi} = -\hat{k}_2\bar{\phi} = iR\bar{\phi}, \quad (2.23)$$

where R is a constant of separation.

It should be noticed that the equation $\hat{k}_2\bar{\phi} = -iR\bar{\phi}$ does not depend on the variables μ and ν, and it commutes with the "energy" operator $i\partial_0$ and linear momentum operator, $i\partial_3$, whose eigenvalues are E and P_z, respectively. Therefore, the effect of Eq. (2.22) for k_2 is to relate the different components of the spinor ϕ among themselves. Choosing to work in the Jauch and Röhrlich[18] representation for Dirac matrices

$$\gamma^0 = \begin{pmatrix} -i & 0 \\ 0 & i \end{pmatrix}, \quad \gamma^\kappa = \begin{pmatrix} 0 & \sigma^\kappa \\ \sigma^\kappa & 0 \end{pmatrix}, \quad \kappa = 1,2,3, \quad (2.24)$$

and substituting (2.24) into (2.21), we find that, Φ takes the form

$$\bar{\phi} = \begin{pmatrix} \phi_1(\mu,\nu) \\ [(m+E)/(R+iP_z)]\sigma^3\phi_1(\mu,\nu) \end{pmatrix}\exp i(P_z z - Et), \quad (2.25)$$

where the constant of separation R satisfies the relation

$$R^2 = E^2 - P_z^2 - m^2. \quad (2.26)$$

The equations for $\phi_1(\mu,\nu)$ and $\phi_2(\mu,\nu)$ can be obtained by substituting (2.24) into (2.20) and considering (2.23). Then, we obtain for ϕ the following two systems of equations:

$$(\sigma^2\partial_1 - \sigma^1\partial_2 + ihR)\phi_1 = 0, \quad (2.27)$$

$$(\sigma^2\partial_1 - \sigma^1\partial_2 + ihR)\phi_2 = 0, \quad \bar{\phi} = \begin{pmatrix} \phi_1 \\ \phi_2 \end{pmatrix}. \quad (2.28)$$

It is clear from (2.25), that the problem of finding a solution to Eq. (2.18) reduces to solving any of the above two equations. In particular, we choose to work with Eq. (2.27).

It should be noticed that the presence of the Lamé metric function h in Eq. (2.27) does not allow us to separate the variables μ and ν in a straightforward way; i.e., we cannot write (2.27) as a sum of two first-order commuting differential operators $\hat{k}_1(\mu)$ and $\hat{k}_2(\nu)$. Therefore, we shall introduce a similarity transformation $T(\mu,\nu)$ acting on the Pauli matrices and on the spinor ϕ_1 in order to write (2.27) in a more simple way, where the functional dependence on μ and becomes ν additive.

Let T be a matrix of the form

$$T = e^{\beta}e^{i\alpha\sigma^1}, \quad (2.29)$$

with

$$\alpha = \alpha(\mu,\nu), \quad \beta = \beta(\mu,\nu).$$

Substituting (2.29) into (2.27), and imposing the relations between α and β

$$\alpha_{,1} = \beta_{,2}, \quad \beta_{,1} = -\alpha_{,2} \quad (2.30)$$

it is easy to see that Eq. (2.27) becomes

$$\{\sigma^2\partial_1 - \sigma^1\partial_2 + ihRe^{2i\alpha\sigma^1}\}Y = 0, \quad (2.31)$$

where Y is related to ϕ_1 by the expression

$$TY = \bar{\phi}_1, \quad Y = \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix}. \quad (2.32)$$

The variables μ and in (2.31) can be separated if we impose

$$e^{2\alpha i\sigma^1} = [a(\mu) + ib(\nu)\sigma^3]/h, \quad (2.33)$$

with $a^2 + b^2 = h^2$.

Among the orthogonal curvilinear coordinates with cylindrical symmetry satisfying conditions (2.30) and (2.33), we find the parabolic and elliptic cylinder coordinates given by

$$x=(\mu^2-\nu^2)/2, \quad y=\mu\nu, \tag{2.34}$$

$$x=\alpha \sin\mu \cosh v$$

and

$$y=\alpha \cos\mu \sinh v \tag{2.35}$$

where the values of α and β were obtained in Refs. 17 and 19.

The equation resulting from the substitution of (2.33) into (2.31) reads

$$(\hat{L}_1+i\hat{L}_2\sigma^3)Y=0, \tag{2.36}$$

where

$$\hat{L}_1=\sigma^2\partial_1+iRa, \quad \hat{L}_2=\sigma^2\partial_2+iRb.$$

It is easy to see that $\hat{L}_1$ and $\hat{L}_2$ commute but not $\hat{L}_1$ and $\hat{L}_2\sigma^3$. Therefore, it is not possible to achieve separation of μ and ν by using two first-order commuting operators. It is, then, necessary to work in a second-order formalism. Let Z be a new auxiliary spinor

$$Y=(i\sigma^3\hat{L}_1+\hat{L}_2)Z. \tag{2.37}$$

Substituting (2.37) into Eq. (2.34), we get

$$[(\partial_1^2+iR\sigma^2\partial_1 a+R^2a^2)$$
$$+(\partial_2^2+iR\sigma^2\partial_2 b+R^2b^2)]Z=0. \tag{2.38}$$

The presence of σ^2 in (2.36) mixes the components Z_1 and Z_2 of the spinor Z. A further simplification of Eq. (2.36) can be accomplished by performing the unitary transformation U on the spinor Z.

$$UZ=(1/\sqrt{2})(1-i\sigma^1)Z=w, \tag{2.39}$$

which allows us to write Eq. (2.36) as follows:

$$(\partial_1^2+iR\sigma^3\partial_1 a+R^2a^2)w=\lambda^2 w, \tag{2.40}$$

$$(\partial_2^2+iR\sigma^3\partial_2 b+R^2a^2)w=-\lambda^2 w, \tag{2.41}$$

where λ^2 is a constant of separation. Due to the form of Eqs. (2.38) and (2.39), the spinor solution w can be written as

$$w=\begin{pmatrix}\alpha(\mu)A(v)\\ \beta(\mu)B(v)\end{pmatrix}, \tag{2.42}$$

where α, β, A, and B satisfy the equations

$$(\partial_1^2+iR\partial_1 a+R^2a^2-\lambda^2)\alpha(\mu)=0, \tag{2.43}$$

$$(\partial_1^2-iR\partial_1 a+R^2a^2-\lambda^2)\beta(\mu)=0, \tag{2.44}$$

$$(\partial_2^2+iR\partial_2 b+R^2b^2+\lambda^2)A(v)=0, \tag{2.45}$$

$$(\partial_2^2-iR\partial_2 b+R^2b^2+\lambda^2)B(v)=0. \tag{2.46}$$

It should be noticed that the above second-order differential equations can be rewritten as two systems of coupled equations as follows:

$$(\partial_1+iRa)\alpha(\mu)=\lambda\beta(\mu), \tag{2.47}$$

$$(\partial_1-iRa)\beta(\mu)=\lambda\alpha(\mu), \tag{2.48}$$

$$(\partial_2+iRb)A(\nu)=i\lambda B(\nu), \tag{2.49}$$

$$(\partial_2-iRb)B(\nu)=i\lambda A(\nu). \tag{2.50}$$

Substituting (2.42) into (2.39) and taking into account the auxiliary condition (2.37), with the relation between α, β and A, B established by the system (2.47)–(2.48) and (2.49)–(2.50), we obtain the expression for the spinor Y:

$$Y=\frac{1}{\sqrt{2}}\begin{pmatrix}(\alpha B-\beta A)(1+i)\\ (\beta A-\alpha B)(1-i)\end{pmatrix}. \tag{2.51}$$

The expression for $T(\mu,\nu)$ Eq. (2.29), can be written in parabolic cylinder and elliptic cylinder coordinates by substituting the corresponding values of α and β [obtained from the system (2.30)]; then we obtain for the parabolic cylinder coordinates[16,17] that $a=\mu$, $b=\nu$,

$$e^{2i\alpha\sigma^1}=(\mu+i\nu\sigma^3)/h \tag{2.52}$$

and, therefore,

$$T=(1/\sqrt{2})\{\sqrt{\mu+h}+i\sigma^3\sqrt{h-\mu}\}, \tag{2.53}$$

where

$$h=\sqrt{\mu^2+\nu^2}. \tag{2.54}$$

On the other hand, for the elliptical cylinder coordinates we obtain[16,17]

$$a=\bar{\alpha}\cos\mu, \quad b=\bar{\alpha}\sinh v,$$

and

$$e^{2i\bar{\alpha}\sigma^1}=(\cos\mu+i\sinh v\sigma^3)/h. \tag{2.55}$$

From (2.29), we have

$$T=[(\cosh v+\sin\mu)^{1/2}/\sqrt{2}h]$$
$$\times\{\sqrt{\cos\mu+h}+i\sigma^3\sqrt{h-\cos\mu}\}, \tag{2.56}$$

with

$$h=(\cos^2\mu+\sinh^2 v)^{1/2}. \tag{2.57}$$

Then, substituting (2.53) and (2.57) into (2.32), we can write the solution to Dirac equation (2.18) in the parabolic cylinder coordinates (2.34), and elliptic cylinder coordinates (2.35), respectively. In Sec. III, we will find the explicit form of the solution to systems (2.47)–(2.48) and (2.49)–(2.50).

III. EXACT SOLUTIONS

The system of equations (2.47)–(2.48) and (2.49)–(2.50) for the parabolic cylinder coordinates reads

$$(d\mu+iR\mu)\alpha=\lambda\beta, \tag{3.1}$$

$$(d\mu-iR\mu)\beta=\lambda\alpha, \tag{3.2}$$

and

$$(d\nu+iR\nu)A=iB\lambda, \tag{3.3}$$

$$(d\nu-iR\nu)B=iA\lambda. \tag{3.4}$$

The solution of the above two systems can be obtained in terms of confluent hypergeometric functions. Substituting (3.2) into (3.1), and (3.1) into (3.2), we obtain two parabolic cylinder equations:

$$(d^2\mu+iR+R^2\mu^2-\lambda^2)\alpha=0, \tag{3.5}$$

$$(d^2\mu-iR+R^2\mu^2-\lambda^2)\beta=0, \tag{3.6}$$

notice that these equations can be obtained directly from (2.43) and (2.44) by setting $a=\mu$. The general solution of (3.5) is

$$\alpha(\mu) = \alpha_0 e^{-iz^2/4} M\left(-\frac{i\lambda^2}{4R}, \frac{1}{2}, \frac{iz^2}{2}\right) + \alpha_1 z e^{-iz^2/4} M\left(-\frac{i\lambda^2}{4R} + \frac{1}{2}, \frac{3}{2}, \frac{iz^2}{2}\right), \tag{3.7}$$

where $z = \sqrt{2R}\,\mu$.

In order to solve (3.6), we observe that Eq. (3.5) is the complex conjugate of (3.6), therefore

$$\beta = \beta_0 e^{+iz^2/4} M\left(\frac{i\lambda^2}{4R}, \frac{1}{2}, -\frac{iz^2}{2}\right) + \beta_1 z e^{iz^2/4} M\left(\frac{i\lambda^2}{4R}, +\frac{1}{2}, \frac{3}{2}, -\frac{iz^2}{2}\right). \tag{3.8}$$

Taking into account the relation[20]

$$M(a,b,z) = e^z M(b-a, b, -z), \tag{3.9}$$

and Eq. (3.1), we obtain

$$\beta = \beta_0 e^{-iz^2/4} M\left(-\frac{i\lambda^2}{4R} + \frac{1}{2}, \frac{1}{2}, \frac{iz^2}{2}\right) + \beta_1 z e^{-iz^2/4} M\left(-\frac{i\lambda^2}{4R} + 1, \frac{3}{2}, \frac{iz^2}{2}\right), \tag{3.10}$$

where the coefficients α_0, α_1, β_0, and β_1 are related as follows:

$$\alpha_0(\lambda/\sqrt{2R}) = \beta_1, \quad \alpha_1 = \beta_0(\lambda/\sqrt{2R}). \tag{3.11}$$

In an analogous way, we obtain the solution of the system (3.3)–(3.4):

$$A = \overline{A}_0 e^{-iy^2/4} M\left(\frac{i\lambda^2}{4R}, \frac{1}{2}, i\frac{y^2}{2}\right) + \overline{A}_1 y e^{-iy^2/4} M\left(\frac{i\lambda^2}{4R} + \frac{1}{2}, \frac{3}{2}, i\frac{y^2}{2}\right), \tag{3.12}$$

$$B = \overline{B}_0 e^{-iy^2/4} M\left(\frac{i\lambda^2}{4R}, +\frac{1}{2}, \frac{1}{2}, i\frac{y^2}{2}\right) + \overline{B}_1 y e^{-iy^2/4} M\left(\frac{i\lambda^2}{4R} + 1, \frac{3}{2}, i\frac{y^2}{2}\right), \tag{3.13}$$

where $y = \sqrt{2R}\,\nu$, and

$$\overline{A}_0[(i\lambda)/\sqrt{2R}] = \overline{B}_1, \quad \overline{B}_0[(i\lambda)/\sqrt{2R}] = \overline{A}_1. \tag{3.14}$$

Now, substituting the expressions for α, β, A, and B, into (2.51), (2.32), we obtain the exact solution to Dirac equation in parabolic cylinder coordinates in the diagonal (rotating) tetrad gauge. The wave spinor solution to Dirac equation (2.4), can be obtained applying the unitary transformation S given by Eq. (2.15) as follows:

$$\psi_c = (\mu^2 + \nu^2)^{-1/4} S\gamma^3\gamma^0\overline{\phi}, \tag{3.15}$$

where S can be obtained from (2.16), giving[19]

$$S = (1/\sqrt{2h})\{(\mu + h)^{1/2} - i\Sigma_3(h - \mu)^{1/2}\}, \tag{3.16}$$

with $h = (\mu^2 + \nu^2)^{1/2}$ and Σ_3 as

$$\Sigma_3 = \begin{pmatrix} \sigma^3 & 0 \\ 0 & \sigma^3 \end{pmatrix}. \tag{3.17}$$

Then, substituting the explicit form of ϕ, into (3.15) obtained from (2.51) and (2.25), we get that, ψ_c reads

$$\psi_c = i \begin{pmatrix} \dfrac{m+E}{R+iP_z} \begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \\ \begin{pmatrix} -Y_1 \\ Y_2 \end{pmatrix} \end{pmatrix} \exp i(P_z z - Et), \tag{3.18}$$

where Y is given by (2.32) and (2.51).

Once we have derived the solution of (2.4), valid for any value of μ and ν, we are going to analyze the asymptotic behavior of ψ_c.

It is easy to see that, the limit $x \to \infty$ can be obtained in the coordinates for the parabolic cylinder (2.34) if we put $\nu = 0$ and make the limit $\mu \to +\infty$. Therefore, using the asymptotic expression for the confluent hypergeometric function for $|z| \gg 1$ (Ref. 20),

$$\frac{M(a,b,z)}{\Gamma(b)} \to \frac{e^{i\pi a} z^{-a}}{\Gamma(b-a)} + \frac{e^z z^{a-b}}{\Gamma(a)}, \tag{3.19}$$

we obtain that the functions y_1 and y_2 for $x \to \infty$ take the form

$$(c_1 + D_1/\sqrt{x})x^{i\lambda^2/4k} e^{-ikx} + (c_2 + D_2/\sqrt{x})x^{-i\lambda^2/4k} e^{ikx}, \tag{3.20}$$

where c_1, D_1, c_2, D_2 are constants. That is, the solution takes the asymptotic form of a modulated cylindrical wave.

Now, we are going to solve the Dirac equation in the elliptical cylinder coordinates defined by (2.35). Substituting the values of a and b obtained in Sec. II, Eq. (2.55) for these coordinates, we see that the system of equations (2.47)–(2.50) takes the form

$$(d_v + iR\overline{\alpha}\sinh v)A = i\lambda B, \tag{3.21}$$

$$(d_v - iR\overline{\alpha}\sinh v)B = i\lambda A. \tag{3.22}$$

Making the change of variable $v = 1w$, and substituting (3.21) into (3.22), and vice versa, we get

$$\left(dw^2 - iR\overline{\alpha}\cos w - \frac{R^2\overline{\alpha}^2}{2}\cos 2w + \frac{R^2\overline{\alpha}^2}{2} - \lambda^2\right)A = 0, \tag{3.23}$$

$$\left(dw^2 + iR\overline{\alpha}\cos w - \frac{R^2\overline{\alpha}^2}{2}\cos 2w + \frac{R^2\overline{\alpha}^2}{2} - \lambda^2\right)B = 0. \tag{3.24}$$

These are Hill's equations with three terms or Whittaker–Hill equations which were studied and discussed by Arscott.[21] After reducing (3.23) and (3.24) to Ince's equations, we have that

$$A = e^{-iR\overline{\alpha}\cos w} Y_1 + e^{iR\overline{\alpha}\cos w} X_2,$$

$$B = e^{-iR\overline{\alpha}\cos w} X_1 + e^{iR\overline{\alpha}\cos w} Y_2,$$

where Y_1, Y_2, X_1, X_2 satisfy the following differential equations:

$$\left[\frac{d^2}{dw^2} \pm 2iR\overline{\alpha}\sin w\frac{d}{dw} - \lambda^2 \pm 2iR\overline{\alpha}\cos w\right]X_{1,2} = 0, \tag{3.25}$$

$$\left[\frac{d^2}{dw^2} \pm 2iR\overline{\alpha}\sin w\frac{d}{dw} - \lambda^2\right]Y_{1,2} = 0, \tag{3.26}$$

where the plus and minus correspond, respectively, to the indices 1 and 2 of in X and Y.

Using Eq. (3.21), and the general solution to (3.25) and

(3.26), we obtain that the functions A and B are given by the expressions

$$A = e^{-iR\bar{\alpha}\cosh v}\left[-i\sum_{r=0}^{\infty} C_r \sinh(rv)\right] + e^{iR\bar{\alpha}\cosh v}\left[-\frac{1}{2\lambda}\sum_{r=1}^{\infty} rG_r \cosh(rv)\right], \quad (3.27)$$

$$B = e^{-iR\bar{\alpha}\cosh v}\left[\frac{i}{2\lambda}\sum_{r=1}^{\infty} rC_r \sinh rv\right] + e^{iR\bar{\alpha}\cosh v}\left[\sum_{r=0}^{\infty} G_r \cosh rv\right], \quad (3.28)$$

where the coefficients C_r and G_r satisfy the recurrence relations

$$2(1+\lambda^2)C_1 + C_2 = 0,$$
$$-(i/2)(r-1)R\bar{\alpha}C_{r-1} + (r^2+4\lambda^2)C_r + (i/2)k\bar{\alpha}(r+1)C_{r+1} = 0, \quad (3.29)$$
$$2(1+\lambda^2)G_1 + G_2 = 0,$$
$$(i/2)(r-1)R\bar{\alpha}G_{r-1} + (r^2+4\lambda^2)G_r - (i/2)k\bar{\alpha}(r+1)G_{r+1} = 0. \quad (3.30)$$

The study of the convergence of the solutions (3.27), (3.28) was carried out by Urwin and Arscott.[22] They show that both series are absolutely and uniformly convergent for any value of the argument v.

The system of equations (2.47), (2.48) for the value of "a" given by Eq. (2.55) can be solved, after some minor changes, using the expressions (3.27) and (3.28). Then, the functions $\alpha(\mu)$ and $\beta(\mu)$ are

$$\alpha = e^{-iR\bar{\alpha}\cos 2\theta}\left\{\sum_{r=0}^{\infty} C_r \sin 2r\theta\right\} + e^{iR\bar{\alpha}\cos 2\theta}\left\{-\frac{1}{2\lambda}\sum_{r=1}^{\infty} rG_r \cos 2r\theta\right\}, \quad (3.31)$$

$$\beta = e^{-iR\bar{\alpha}\cos 2\theta}\left\{-\frac{1}{2\lambda}\sum_{r=1}^{\infty} rC_r \sin 2r\theta\right\} + e^{iR\bar{\alpha}\cos 2\theta}\left\{\sum_{r=0}^{\infty} G_r \cos 2r\theta\right\}, \quad (3.32)$$

where $\mu = \pi/2 - 2\theta$ and the relation between the coefficients is given by (3.29) and (3.30).

Then, an exact solution to Eq. (2.8) in the coordinates (2.35) can be obtained substituting the expressions (3.27), (3.28), (3.31), and (3.32) into (2.51), (2.32), and (2.25). In elliptical cylinder coordinates, the unitary matrix transformation (2.15), which relates the solution ψ_c and ψ_d, is[19]

$$S = (1/\sqrt{2h})\{(\cos\mu\cosh v + h)^{1/2} + i\Sigma_3(h - \cos\mu\cosh v)^{1/2}\}, \quad (3.33)$$

with

$$h = (\cosh^2 v + \sin^2\mu)^{1/2},$$

and Σ_3 is given by (3.17).

Therefore, the solution to Dirac equation (2.4) in the Cartesian (fixed) gauge tetrad reads

$$\psi_c = i\frac{(\cosh v + \sin\mu)^{1/2}}{2h^2}\begin{pmatrix} \dfrac{m+E}{R+iP_z}\,P\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \\ P\begin{pmatrix} -Y_1 \\ Y_2 \end{pmatrix} \end{pmatrix} \times \exp i(P_z z - Et), \quad (3.34)$$

where P is

$$\{[(\cos\mu\cosh v + h)(\cos\mu + h)]^{1/2} - [(h - \cos\mu\cosh v)(h - \cos\mu)]^{1/2}\} + i\sigma^3\{[(\cos\mu\cosh v + h)(h - \cos\mu)]^{1/2} + [(h - \cos\mu\cosh v)(\cos\mu + h)]^{1/2}\}. \quad (3.35)$$

Finally, we can analyze the asymptotic behavior of the Dirac spinor ψ_c when y goes to infinity and x is zero ("y axis"). In the coordinates of the elliptical cylinder, it is equivalent to consider $v \to \infty$ and $\mu = 0$. On the y axis, ψ_c takes the simple form

$$\psi_c(\mu = 0) = i\begin{pmatrix} \dfrac{m+E}{R+iP_z}\begin{pmatrix} Y_1 \\ Y_2 \end{pmatrix} \\ \begin{pmatrix} -Y_1 \\ Y_2 \end{pmatrix} \end{pmatrix} \exp i(P_z z - Et), \quad (3.36)$$

where Y is given by (2.51).

Using the results of Arscott about the asymptotic behavior of paraboidal wave functions and solutions to Whittaker–Hill equations, it is straightforward to obtain the form that the expressions (3.27), (3.28) take when v goes to infinity. After substituting the asymptotic form of (3.27), (3.28) into (2.51), we get

$$Y_1 = (\cosh v)^{-1/2}\{C_1 e^{-iR\bar{\alpha}\cosh v} + C_2 e^{iR\bar{\alpha}\cosh v}\}, \quad (3.37)$$

$$Y_2 = (\cosh v)^{-1/2}\{C_3 e^{-iR\bar{\alpha}\cosh v} + C_4 e^{iR\bar{\alpha}\sinh v}\}, \quad (3.38)$$

where C_1, C_2, C_3, C_4 are constants.

From (3.37), (3.38), we note that, in the asymptotic region $y \to \infty$, the solution to the Dirac equation in the coordinates (2.35) behaves as a plane wave modulated by a factor $(\cosh v)^{-1/2}$ (which for great distances is proportional to $y^{-1/2}$).

As a final remark, we have to say that the new exact solutions obtained in the present work could be applied in the analysis of scattering processes of relativistic electrons by a dispersion center with symmetry associated to elliptical cylinder or parabolic cylinder coordinates.

[1] W. Miller, Jr., *Symmetry and Separation of Variables* (Addison-Wesley, Reading, MA, 1977).
[2] P. J. Redmond, J. Math. Phys **6**, 1163 (1965).
[3] L. Lam, Phys. Lett. A **31**, 406 (1970).
[4] G. N. Stanciu, J. Math. Phys. **8**, 2043 (1967).
[5] S. W. Hawking, Commun. Math. Phys. **43**, 199 (1975).
[6] G. W. Gibbons and S. W. Hawking, Phys. Rev. D **15**, 2738 (1977).
[7] S. A. Teukolsky, Astrophys. J. **185**, 635 (1973).
[8] B. Carter and R. G. McLennaghan, Phys. Rev. D **19**, 1093 (1979).
[9] R. G. McLennaghan and Ph. Spindel, Phys. Rev. D **20**, 409 (1979).

[10] R. Guven, Proc. R. Soc. London Ser. A **356**, 465 (1977).
[11] S. Chandrasekhar, *The Mathematical Theory of Black Holes* (Oxford U.P., Oxford, 1983).
[12] V. G. Bagrov, D. M. Gitman, I. M. Ternov, V. P. Khalilov, and V. N. Shapovalov, *Exact Solutions of Relativistic Wave Equations* (Nauka, Novosibirsk, 1982).
[13] A. S. Davydov, *Quantum Mechanics* (Pergamon, Oxford, 1965).
[14] A. H. Cook, Proc. R. Soc. London Ser. A **383**, 247 (1982).
[15] E. G. Kalnins, W. Miller, Jr., and G. C. Williams, J. Math. Phys. **27**, 1893 (1986).
[16] G. V. Shishkin and V. M. Villalba, J. Math. Phys. **30**, 2132 (1989).
[17] V. M. Villalba and G. V. Shishkin, Vestsi Akad. Navuk B. SSR Ser. Fiz. Mat. Navuk **6**, 87 (1988).
[18] J. M. Jauch and F. Rohrlich, *The Theory of Photons and Electrons* (Springer, New York, 1976).
[19] G. V. Shishkin and V. M. Villalba, Vestn. Beloruss. Gos. Univ. Ser. 1 **3**, 5 (1987).
[20] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).
[21] F. M. Arscott, Proc. R. Soc. Edinburgh Sec. A **67**, 265 (1967).
[22] K. M. Urwin and F. M. Arscott, Proc. R. Soc. Edinburgh Sec. A **69**, 28 (1969).

An Sp(2)-covariant version of generalized canonical quantization of dynamical systems with linearly dependent constraints

I. A. Batalin
Lebedev Physical Institute, Moscow 117924, USSR

P. M. Lavrov
Tomsk State Pedagogical Institute, Tomsk 634041, USSR

I. V. Tyutin
Moscow Institute of Electronic Engineering, Moscow 109028, USSR

(Received 14 February 1990; accepted for publication 23 May 1990)

An Sp(2)-covariant version of the method of generalized canonical quantization of dynamical systems with linearly dependent first-class constraints is proposed. The existence theorem for solutions of generating equations of a gauge algebra is proved and the natural arbitrariness in these solutions is described. The scheme proposed is shown to be equivalent to the standard version of generalized canonical quantization.

I. INTRODUCTION

In our previous papers[1,2] we considered an Sp(2)-symmetric version of the method of generalized canonical quantization and the corresponding covariant (Lagrangian) formulation of dynamical systems with linearly independent gauge generators. However, the requirements of locality and explicit relativistic convariance of the dynamic description are known to lead in the general case to the necessity of using linearly dependent (reducible) generators of gauge transformations, of which antisymmetric tensor fields are a typical example.

Being linearly dependent, gauge generators possess zero vectors which, in turn, may have their own zero vectors, etc. So, in this case there naturally arises an exact sequence whose length, by definition, is the stage of theory reducibility.

In the framework of the standard version of generalized canonical formalism, the problem of quantization of reducible gauge theories was solved in Refs. 3–5. The corresponding Lagrangian formulation was proposed in Refs. 6–8.

The present paper is aimed at formulating an Sp(2)-symmetric version of the method of generalized canonical quantization for dynamical systems with linearly dependent first-class constraints. Thereby the results of our previous paper[1] are extended here to gauge theories of any stage of reducibility. Note that exceedingly instructive for us were Refs. 9 and 10.

The present paper is organized as follows. The basic definitions concerning the reducibility properties of the theory are given in Sec. II. In Sec. III an extended phase space is introduced, the Sp(2)-covariant generating equations of the gauge algebra along with the necessary boundary conditions are formulated, and the unitarizing Hamiltonian of the theory is constructed. The existence theorem for the solution of Sp(2)-covariant generating equations is proved in Sec. IV. The natural arbitrariness in the solution of the generating equations is considered in Sec. V, after which physical equivalence of the Sp(2) symmetric and the standard versions of the method of generalized canonical quantization in the reducible case is established. Finally, Sec. VII presents a brief discussion of the alternative version of the Sp(2)-symmetric Hamiltonian formulation that generalizes the results of Ref. 5.

We use here the following notation. The Poisson superbracket in the phase space (P_A, Q^A) is standardly defined as[11]

$$\{G,F\} = \frac{\delta G}{\delta Q^A}\frac{\delta F}{\delta P_A} - \frac{\delta F}{\delta Q^A}\frac{\delta G}{\delta P_A}(-1)^{\epsilon(G)\epsilon(F)},$$

where $\epsilon(G)$ denotes the Grassmann parity of the quantity G. The Grassmann parities of the canonical variables P_A and Q^A coincide: $\epsilon(P_A) = \epsilon(Q^A) = \epsilon_A$. Derivatives with respect to generalized momenta are always understood as left and those with respect to generalized coordinates as right. Left derivatives with respect to Q^A are labeled by the letter "l," i.e., $\delta_l/\delta Q^A$. The Jacobi identity for the superbracket has the form

$$\{\{F,G\},H\}(-)^{\epsilon(F)\epsilon(H)} + \text{cycl. perm.}(FGH) = 0.$$

The ghost number of the quantity G is denoted as usual by $\mathrm{gh}(G)$. Besides, we further define the so-called new ghost number that we denote by $\mathrm{ngh}(G)$. Both the types of the ghost number obey the additional composition law:

$$\mathrm{gh}(GH) = \mathrm{gh}(G) + \mathrm{gh}(H),$$

$$\mathrm{ngh}(GH) = \mathrm{ngh}(G) + \mathrm{ngh}(H).$$

The indices of the global symplectic group Sp(2) are labeled by lower-case letters $a, b, c, \ldots$ and assume two values $a = 1,2$. The invariant tensor of the Sp(2) group is defined as

$$\epsilon^{ab} = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix}, \quad \epsilon^{ac}\epsilon_{cb} = \delta^a_b.$$

Symmetrization over the indices of the Sp(2) group is given by

$$B^{\{ab\}} = B^{ab} + B^{ba}.$$

Ranks of supermatrices are characterized by a pair of numbers $r = (r_+, r_-)$, where $r_+ (r_-)$ is the rank of a bose–bose (fermi–fermi) block of the supermatrix. Similarly, the

 0022-2488/90/112708-10$03.00

number of constraints will be characterized by a pair of numbers $m=(m_+,m_-)$, where m_+ (m_-) is the number of bose (fermi) constraints. We use the following conventions:

$$(m_+,m_-)=(n_+,n_-)\Leftrightarrow m_\pm=n_\pm,$$
$$(m_+,m_-)<(n_+,n_-)\Leftrightarrow m_+<n_+,\quad m_-=n_-,$$
$$\text{or}\quad m_+=n_+,\quad m_-<n_-,$$
$$\text{or}\quad m_+<n_+,\quad m_-<n_-,$$
$$(m_+,m_-)+(n_+,n_-)=(m_++n_+,m_-+n_-).$$

II. REDUCIBLE GAUGE THEORIES

We shall consider, in the Hamiltonian formalism, a dynamical system with a finite number of degrees of freedom. Suppose in the phase space of initial canonical variables $\eta=(p_i,q^i)$, $i=1,2,...,n$; $\epsilon(p_i)=\epsilon(q^i)=\epsilon_i$ this system is described by the Hamiltonian $H_0=H_0(p,q)$, $\epsilon(H_0)=0$, and by the set of first-class constraints $T_{\alpha_0}=T_{\alpha_0}(p,q)$, $\epsilon(T_{\alpha_0})=\epsilon_{\alpha_0}$, $\alpha_0=1,2,...,m_0=m_{0+}+m_{0-}$. We assume that the constraints $T_{\alpha_0}\sim\eta+O(\eta^2)$. By the definition of first-class constraints, there hold the following involution relations:

$$\{T_{\alpha_0},T_{\beta_0}\}=T_{\gamma_0}U^{\gamma_0}_{\alpha_0\beta_0},\quad \{H_0,T_{\alpha_0}\}=T_{\beta_0}V^{\beta_0}_{\alpha_0}. \tag{1}$$

Consider the Jacobian supermatrix of the initial constraints

$$\Sigma=\{T,\eta\}\big|_{T=0} \tag{2}$$

and let its rank be equal to

$$\text{rank}\,\Sigma\equiv(r_{0+},r_{0-}). \tag{3}$$

If the condition $r_{0\pm}=m_{0\pm}$ is fulfilled, the constraints T_{α_0} are linearly independent and we are dealing with the so-called irreducible gauge theories. The procedure of quantization of such theories in the standard version of generalized canonical formalism was proposed in Refs. 12 and 13. Then the recent paper[1] by the authors gave formulation of the corresponding Sp(2)-covariant modification of this procedure. If $(r_{0+},r_{0-})<(m_{0+},m_{0-})$, then the constraints T_{α_0} are dependent. In this case the gauge theories are called reducible. The linear dependence of T_{α_0} means that on the constraint surface $T_{\alpha_0}=0$ there exist vectors $Z^{\alpha_0}_{\alpha_1}=Z^{\alpha_0}_{\alpha_1}(p,q)$, such that

$$T_{\alpha_0}Z^{\alpha_0}_{\alpha_1}=T_{\alpha_0}T_{\beta_0}K^{\alpha_0\beta_0}_{\alpha_1},\quad \alpha_1=1,...,m_1=m_{1+}+m_{1-}. \tag{4}$$

Redefining

$$Z^{\alpha_0}_{\alpha_1}(Z^{\alpha_0}_{\alpha_1}\to\overline{Z}^{\alpha_0}_{\alpha_1}=Z^{\alpha_0}_{\alpha_1}-T_{\beta_0}K^{\alpha_0\beta_0}_{\alpha_1}),$$

we can always choose $Z^{\alpha_0}_{\alpha_1}$ such that the relations of the first stage of reducibility (4) are of the form

$$T_{\alpha_0}Z^{\alpha_0}_{\alpha_1}=0. \tag{5}$$

There exist such numbers $\epsilon_{\alpha_1}=0,1$ that $\epsilon(Z^{\alpha_0}_{\alpha_1})=\epsilon_{\alpha_0}+\epsilon_{\alpha_1}$. Let, on the constraint surface, the rank of the supermatrix $Z^{\alpha_0}_{\alpha_1}$ be equal to

$$\text{rank}\quad Z^{\alpha_0}_{\alpha_1}\big|_{T_{\alpha_0}=0}\equiv(r_{1+},r_{1-}).$$

If the condition $m_{1\pm}=r_{1\pm}\equiv m_{0\pm}-r_{0\pm}$ is fulfilled, the gauge theories are said to have the first stage of reducibility. In the general case

$$(m_{1+},m_{1-})>(r_{1+},r_{1-}),$$

the set $Z^{\alpha_0}_{\alpha_1}$ itself is linearly dependent, and therefore there exists a set of vectors $Z^{\alpha_1}_{\alpha_2}=Z^{\alpha_1}_{\alpha_2}(p,q)$:

$$Z^{\alpha_0}_{\alpha_1}Z^{\alpha_1}_{\alpha_2}=T_{\beta_0}A^{\beta_0\alpha_0}_{\alpha_2},\quad \alpha_2=1,...,m_2=m_{2+}+m_{2-}, \tag{6}$$

and numbers $\epsilon_{\alpha_2}=0,1$, such that $\epsilon(Z^{\alpha_1}_{\alpha_2})=\epsilon_{\alpha_1}+\epsilon_{\alpha_2}$. Let, in turn

$$\text{rank}\quad Z^{\alpha_1}_{\alpha_2}\big|_{T_{\alpha_0}=0}\equiv(r_{2+},r_{2-}).$$

Be redefining $Z^{\alpha_1}_{\alpha_2}$, one can choose the coefficients $A^{\alpha_0\beta_0}_{\alpha_2}$ in (6) to be antisymmetric. We assume that this has been done in a way that

$$A^{\alpha_0\beta_0}_{\alpha_2}=-(-1)^{\epsilon_{\alpha_0}\epsilon_{\beta_0}}A^{\beta_0\alpha_0}_{\alpha_2}. \tag{7}$$

If $m_{2\pm}=r_{2\pm}\equiv m_{1\pm}-r_{1\pm}$, we are dealing with gauge theories of the second stage of reducibility. In the general case the set $Z^{\alpha_1}_{\alpha_2}$ can be overfull, i.e., $(m_{2+},m_{2-})>(r_{2+},r_{2-})$, etc. Thus there appears a sequence of equations of reducibility

$$Z^{\alpha_{s-2}}_{\alpha_{s-1}}Z^{\alpha_{s-1}}_{\alpha_s}=T_{\beta_0}A^{\beta_0\alpha_{s-2}}_{\alpha_s},$$
$$\alpha_s=1,...,\quad m_s=m_{s+}+m_{s-},\quad s=1,...,L, \tag{8}$$

where

$$Z^{\alpha_{-1}}_{\alpha_0}\equiv T_{\alpha_0},\quad A^{\beta_0\alpha_{-1}}_{\alpha_1}=0,\quad \epsilon(Z^{\alpha_{s-1}}_{\alpha_s})=\epsilon_{\alpha_{s-1}}+\epsilon_{\alpha_s}$$

and

$$\text{rank}\quad Z^{\alpha_{s-1}}_{\alpha_s}\Big|_{T_{\alpha_0}=0}\equiv(r_{s+},r_{s-}). \tag{9}$$

The order of reducibility L is determined by the last of the s values for which $(m_{s+},m_{s-})=(r_{s+},r_{s-})$. Irreducible theories are characterized by $L=0$.

Note that the constraints T_{α_0} are not uniquely defined. The constraints $\overline{T}_{\alpha_0}=0$,

$$\overline{T}_{\alpha_0}=T_{\beta_0}M^{\beta_0}_{\alpha_0}, \tag{10}$$

where the supermatrix $M^{\beta_0}_{\alpha_0}$ is invertible, define the same surface as $T_{\alpha_0}=0$. Obviously, the quantities $Z^{\alpha_{s-1}}_{\alpha_s}$ in (8) are not uniquely defined either. The arbitrariness that exists here can be described by the following relations

$$\overline{Z}^{\alpha_0}_{\alpha_1}=Z^{\alpha_0}_{\beta_1}M^{\beta_1}_{\alpha_1}+T_{\beta_0}D^{\beta_0\alpha_0}_{\alpha_1}, \tag{11}$$
$$\overline{Z}^{\alpha_{s-1}}_{\alpha_s}=Z^{\alpha_{s-1}}_{\beta_s}M^{\beta_s}_{\alpha_s}+T_{\beta_0}D^{\beta_0\alpha_{s-1}}_{\alpha_s},\quad s=2,...,L. \tag{12}$$

In (11) and (12) the supermatrices $M^{\beta_s}_{\alpha_s}$, $s=1,...,L$, are invertible $D^{\beta_0\alpha_0}_{\alpha_1}$ and $D^{\beta_0\alpha_1}_{\alpha_2}$ possess the properties

$$D^{\beta_0\alpha_0}_{\alpha_1}=-(-1)^{\epsilon_{\alpha_0}\epsilon_{\beta_0}}D^{\alpha_0\beta_0}_{\alpha_1},$$
$$Z^{\alpha_0}_{\alpha_1}D^{\beta_0\alpha_1}_{\alpha_2}=-(-1)^{\epsilon_{\alpha_0}\epsilon_{\alpha_1}+\epsilon_{\alpha_0}\epsilon_{\beta_0}+\epsilon_{\beta_0}\epsilon_{\alpha_1}}Z^{\beta_0}_{\alpha_1}D^{\alpha_0\alpha_1}_{\alpha_2}, \tag{13}$$

and $D^{\beta_0\alpha_{s-1}}_{\alpha_s}$, $s=3,...,L$ are arbitrary. We can make use of the arbitrariness (10)–(13) in the definition of the constraints T_{α_0} and the quantities $Z^{\alpha_{s-1}}_{\alpha_s}$ to bring locally T_{α_0}, $Z^{\alpha_{s-1}}_{\alpha_s}$ to the simplest possible form. Denote the set of linearly independent constraints contained among the initial constraints T_{α_0} by T_{A_0}. To this convention there corresponds the dividing $\alpha_0=(A_0,\bar{\alpha}_0)$. Then we have

$$T_{\bar{\alpha}_0} = T_{A_0} K^{A_0}_{\bar{\alpha}_0}. \tag{14}$$

We now redefine the dependent constraints $\overline{T}_{\bar{\alpha}_0} = T_{\bar{\alpha}_0} - T_{A_0} K^{A_0}_{\bar{\alpha}_0}$ and go over from the constraints T_{α_0} to an equivalent set $\overline{T}_{\alpha_0}$:

$$\overline{T}_{\alpha_0} = T_{\beta_0} M^{\beta_0}_{\alpha_0} = (T_{A_0}, 0), \quad M = \begin{pmatrix} 1 & 0 \\ -K & -1 \end{pmatrix}, \tag{15}$$

where the supermatrix M is obviously nonsingular. If constraints are chosen in the form (15), Eqs. (5) become

$$T_{A_0} \tilde{Z}^{A_0}_{\alpha_1} = 0, \quad \tilde{Z}^{\alpha_0}_{\alpha_1} = (M^{-1})^{\alpha_0}_{\beta_0} Z^{\beta_0}_{\alpha_1}. \tag{16}$$

Since the constraints T_{A_0} are linearly independent on the constraint surface, then from (16) (omitting the tilde over $\tilde{Z}^{A_0}_{\alpha_1}$) we conclude that

$$Z^{A_0}_{\alpha_1} = T_{B_0} M^{B_0 A_0}_{\alpha_1}, \tag{17}$$

with a certain matrix $M^{B_0 A_0}_{\alpha_1}$ possessing the properties

$$M^{B_0 A_0}_{\alpha_1} = -(-1)^{\epsilon_{B_0}\epsilon_{A_0}} M^{A_0 B_0}_{\alpha_1}. \tag{18}$$

Defining $\overline{Z}^{A_0}_{\alpha_1} : \overline{Z}^{A_0}_{\alpha_1} = Z^{A_0}_{\alpha_1} - T_{B_0} M^{B_0 A_0}_{\alpha_1}$, we can always choose $Z^{A_0}_{\alpha_1} = 0$ as a solution of Eq. (16), and

$$\operatorname{rank} \left. Z^{\alpha_0}_{\alpha_1} \right|_{T_{\alpha_0}=0} = \operatorname{rank} \left. Z^{\bar{\alpha}_0}_{\alpha_1} \right|_{T_{\alpha_0}=0}. \tag{19}$$

In turn, Eq. (19) allows us to divide the index α_1 into two groups $\alpha_1 = (A_1, \bar{\alpha}_1)$ in such a way that the matrix $Z^{\bar{\alpha}_0}_{A_1}, A_1 = 1,...,m_{1+} + m_{1-}$, is reversible on the constraint surface and the columns of the matrix $Z^{\bar{\alpha}_0}_{\bar{\alpha}_1}$ are linearly dependent

$$Z^{\bar{\alpha}_0}_{\bar{\alpha}_1} = Z^{\bar{\alpha}_0}_{A_1} M^{A_1}_{\bar{\alpha}_1}. \tag{20}$$

Redefining now $Z^{\bar{\alpha}_0}_{\bar{\alpha}_1} : \overline{Z}^{\bar{\alpha}_0}_{\bar{\alpha}_1} = Z^{\bar{\alpha}_0}_{\bar{\alpha}_1} - Z^{\bar{\alpha}_0}_{A_1} M^{A_1}_{\bar{\alpha}_1}$, we may assume, when solving Eq. (5), that $Z^{\bar{\alpha}_0}_{\bar{\alpha}_1} = 0$. Finally, making allowance for reversibility of the matrix $Z^{\bar{\alpha}_0}_{A_1}$ we can obtain, by making use of a linear combination, $Z^{\bar{\alpha}_0}_{A_1} = \delta^{\bar{\alpha}_0}_{A_1}$. With similar arguments applied to the quantities $Z^{\alpha_{s-1}}_{\alpha_s}$ at each stage of reducibility, we obtain

$$Z^{\alpha_{s-1}}_{\alpha_s} = -\left(\begin{array}{c|c} 0 & I \\ \hline 0 & 0 \end{array}\right)^{A_{s-1}}_{\bar{\alpha}_s}{}^{\bar{\alpha}_{s-1}}_{A_s}, \quad s = 1,...,L,$$
$$Z^{\alpha_{L-1}}_{\alpha_L} = (0, \delta^{\bar{\alpha}_{L-1}}_{A_L}). \tag{21}$$

Further on we shall say that T_{α_0} and $Z^{\alpha_{s-1}}_{\alpha_s}, s = 1,...,L$, are chosen in the standard basis if they have the forms of (15) and (21). It should be emphasized that passing over to the standard basis is realized using two operations, namely, multiplying by a nonsingular matrix and adding the terms vanishing on the constraints surface.

III. Sp(2)-COVARIANT QUANTIZATION OF REDUCIBLE GAUGE THEORIES

We shall consider the gauge theory of the L th stage of reducibility. The extended phase space Γ of such a theory can be parametrized by the following set of canonical variables:

$$\Gamma = (P_A, Q^A) = (p_i, q^i; \mathscr{P}_{\alpha_0|a}, c^{\alpha_0|a}; \lambda_{\alpha_0}, \pi^{\alpha_0};$$
$$\mathscr{P}_{\alpha_s|aa_1\cdots a_s}, c^{\alpha_s|aa_1\cdots a_s}; \lambda_{\alpha_s|a_1\cdots a_s},$$
$$\pi^{\alpha_s|a_1\cdots a_s}; s = 1,...,L), \tag{22}$$

where the momenta $\mathscr{P}_{\alpha_0|a}, \mathscr{P}_{\alpha_s|aa_1\cdots a_s}, \lambda_{\alpha_s|a_1\cdots a_s}$, and the coordinates $c^{\alpha_0|a}, c^{\alpha_s|aa_1\cdots a_s}, \pi^{\alpha_s|a_1\cdots a_s}$, form symmetric tensors of corresponding ranks relative to the symplectic group Sp(2). The Grassmann parities of momenta and coordinates coincide, and for the newly introduced variables they are defined as follows:

$$\epsilon(c^{\alpha_0|a}) = \epsilon_{\alpha_0} + 1, \quad \epsilon(\pi^{\alpha_0}) = \epsilon_{\alpha_0},$$
$$\epsilon(c^{\alpha_s|aa_1\cdots a_s}) = \epsilon_{\alpha_s} + s + 1 (\operatorname{mod} 2), \tag{23}$$
$$\epsilon(\pi^{\alpha_s|a_1\cdots a_s}) = \epsilon_{\alpha_s} + s (\operatorname{mod} 2), \quad s = 1,...,L.$$

Variables of the extended phase space are assigned ghost numbers by the rule

$$\begin{aligned} &\operatorname{gh}(p_i) = \operatorname{gh}(q^i) = 0, \\ &\operatorname{gh}(c^{\alpha_0|a}) = -\operatorname{gh}(\mathscr{P}_{\alpha_0|a}) = 3 - 2a, \\ &\operatorname{gh}(\lambda_{\alpha_0}) = \operatorname{gh}(\pi^{\alpha_0}) = 0, \\ &\operatorname{gh}(c^{\alpha_s|aa_1\cdots a_s}) = -\operatorname{gh}(\mathscr{P}_{\alpha_s|aa_1\cdots a_s}) \\ &\qquad = 3 - 2a + \sum_{s'=1}^{s} (3 - 2a_{s'}), \\ &\operatorname{gh}(\pi^{\alpha_s|a_1\cdots a_s}) = -\operatorname{gh}(\lambda_{\alpha_s|a_1\cdots a_s}) \\ &\qquad = \sum_{s'=1}^{s} (3 - 2a_{s'}). \end{aligned} \tag{24}$$

In connection with the definitions (24) we recall that the group indices $a, a_1,...,a_s$ assume only two values, 1 and 2.

It would be reasonable to present here an explicit form of the normalization of elementary Poisson brackets for canonical variables realizing symmetric representations of the Sp(2) group:

$$\{\pi^{\alpha_s|a_1\cdots a_s}, \lambda_{\beta_r|b_1\cdots b_r}\} = \delta^s_r S^{a_1\cdots a_s}_{b_1\cdots b_s} \delta^{\alpha_s}_{\beta_s}, \tag{25}$$
$$\{c^{\alpha_s|a_1\cdots a_{s+1}}, \mathscr{P}_{\beta_r|b_1\cdots b_{r+1}}\} = \delta^s_r S^{a_1\cdots a_{s+1}}_{b_1\cdots b_{s+1}} \delta^{\alpha_s}_{\beta_s}. \tag{26}$$

The symmetrizer $S^{a_1\cdots a_s}_{b_1\cdots b_s}$ in the rhs of (25) and (26) is defined as

$$S^{a_1\cdots a_s}_{b_1\cdots b_s} \equiv \frac{1}{s!} \frac{\delta}{\delta X^{b_1}} \cdots \frac{\delta}{\delta X^{b_s}} X^{a_s} \cdots X^{a_1}, \tag{27}$$

where X^a are auxiliary boson variables. The definition (27) immediately implies the properties

$$S^{a_1\cdots a_s}_{b_1\cdots b_s} S^{b_1\cdots b_s}_{c_1\cdots c_s} = S^{a_1\cdots a_s}_{c_1\cdots c_s}, \tag{28}$$
$$S^{a_1\cdots a_s}_{b_1\cdots b_s} = \frac{1}{s} \sum_{i=1}^{s} S^{a_1\cdots a_{i-1}a_{i+1}\cdots a_s}_{b_2\cdots b_s} \delta^{a_i}_{b_1}. \tag{29}$$

The key role in the procedure of extended BRST quantization of reducible gauge theories is played by the generating functions Ω^a and $\mathscr{H}$ defined on the extended phase space Γ (22). The fermion functions Ω^a are solutions of the Sp(2)-covariant generating equations

$$\{\Omega^a, \Omega^b\} = 0, \tag{30}$$

which also satisfy the boundary conditions

$$\left.\frac{\delta \Omega^a}{\delta c^{\alpha_0|b}}\right|_{c=\mathscr{P}=\pi=\lambda=0} = T_{\alpha_0} \delta^a_b,$$
$$\left.\frac{\delta \Omega}{\delta \pi^{\alpha_0}}\right|_{c=\pi=\mathscr{P}_{\varnothing}=\lambda=0} = \epsilon^{ab} \mathscr{P}_{\alpha_0|b}, \tag{31}$$

where the symbol $\mathscr{P}_{\varnothing}$ implies that from the set of momenta $\mathscr{P}_{\alpha_0|a}$, $\mathscr{P}_{\alpha_s|aa_1\cdots a_s}, s=1,...,L$, one should discard $\mathscr{P}_{\alpha_0|a}$.

The boson function $\mathscr{H}$ satisfies generating equations of the form

$$\{\mathscr{H},\Omega^a\}=0 \tag{32}$$

with the boundary condition

$$\mathscr{H}|_{c=\pi=\mathscr{P}=\lambda=0}=H_0. \tag{33}$$

The total unitarizing Hamiltonian H is now defined in terms of $\mathscr{H}$ and Ω^a by the formula

$$H=\mathscr{H}+\tfrac{1}{2}\epsilon_{ab}\{\{\phi,\Omega^b\},\Omega^a\}, \tag{34}$$

where ϕ is the boson function fixing a concrete choice of admissible gauge, $\mathrm{gh}(\phi)$ being equal to zero. An essential property of the unitarizing Hamiltonian H (34) is its invariance under transformation of the canonical variables (22)

$$\delta\Gamma=\{\Gamma,\Omega^a\}\mu_a, \tag{35}$$

namely,

$$\delta H=\{H,\Omega^a\}\mu_a=0. \tag{36}$$

In (35) and (36) the quantity μ_a is an Sp(2) doublet of constant Grassmann parameters. Invariance of (36) is obvious from Eqs. (30), (32) and the Jacobi identity for Ω^a.

In the theory with the Hamiltonian H, the generating functional of the Green's functions is given in the usual manner by the functional integral

$$Z_\phi(J)=\int D\Gamma\exp\left\{\frac{i}{\hbar}\int dt(P_A\dot{Q}^A-H+J\Gamma)\right\}. \tag{37}$$

The generating equations (30) and (32) for Ω^a and $\mathscr{H}$, the definition (34) of the unitarizing Hamiltonian H, and the definition (37) of the generating functional of the Green's functions $Z_\phi(J)$ coincide formally with the case of irreducible gauge theories.[1] Therefore, the corollaries of the invariance of H (36), obtained in Ref. 1, are also valid for reducible gauge theories in the quantum region. In particular, we state that the S matrix corresponding to (37) does not depend on the choice of gauge function ϕ in (34). The invariance (36) of the total Hamiltonian H allows the Ward identities to also be derived for the functional (37). We omit the corresponding arguments and calculations and refer the reader to Ref. 1.

IV. THE EXISTENCE THEOREM FOR THE GENERATING EQUATIONS OF GAUGE ALGEBRA

As in the case of irreducible gauge theories,[1] the question of the existence of solutions of the generating equations (30) and (32) satisfying the boundary conditions (31) and (32) is the most essential for the quantization scheme. Here, we shall prove the existence of solutions of equations (30) and (32) in the form of power series of canonical variables $c^{\alpha_0|a}$, π^{α_0}, $c^{\alpha_s|aa_1\cdots a_s}$, $\pi^{\alpha_s|a_1\cdots a_s}, s=1,...,L$. As in Ref. 1, it is convenient to control the arising series using the so-called "new ghost number"— ngh. To do so, we ascribe the new ghost number to all the variables of the extended phase space (22) by the rule

$$\begin{aligned}
&\mathrm{ngh}(p_i)=\mathrm{ngh}(q^i)=0,\\
&\mathrm{ngh}(c^{\alpha_0|a})=-\mathrm{ngh}(\mathscr{P}_{\alpha_0|a})=1,\\
&\mathrm{ngh}(\pi^{\alpha_0})=-\mathrm{ngh}(\lambda_{\alpha_0})=2,\\
&\mathrm{ngh}(c^{\alpha_s|aa_1\cdots a_s})=-\mathrm{ngh}(\mathscr{P}_{\alpha_s|aa_1\cdots a_s})=s+1,\\
&\mathrm{ngh}(\pi^{\alpha_s|a_1\cdots a_s})=-\mathrm{ngh}(\lambda_{\alpha_s|a_1\cdots a_s})=s+2,\\
&\qquad\qquad s=1,...,L,
\end{aligned} \tag{38}$$

and require fulfillment of the conditions

$$\mathrm{ngh}(\mathscr{H})=0,\quad \mathrm{ngh}(\Omega^a)=1. \tag{39}$$

We shall seek the solutions of equations (30) and (32) in the form of the following expansions in power series of $c^{\alpha_0|a}$, π^{α_0}, $c^{\alpha_s|aa_1\cdots a_s}$, $\pi^{\alpha_s|a_1\cdots a_s}$, $s=1,...,L$:

$$\Omega^a=\sum_{n=1}^{\infty}\Omega^a_n,\quad \mathrm{ngh}(\Omega^a_n)=1,\quad \Omega^a_n\sim(c,\pi)^n, \tag{40}$$

$$\mathscr{H}=H_0+\sum_{n=1}^{\infty}\mathscr{H}_n,\quad \mathrm{ngh}(\mathscr{H}_n)=0,\quad \mathscr{H}_n\sim(c,\pi)^n. \tag{41}$$

We shall first prove the existence of solutions of equations (30). The first approximation Ω^a_1 can be represented in the form

$$\Omega^a_1=\sum_{l=1}^{L+1}\overset{(l)}{\Omega}{}^a_1,\quad \overset{(l)}{\Omega}{}^a_1\sim(\mathscr{P},\lambda)^l,\quad l=2,...,L+1, \tag{42}$$

and $\overset{(1)}{\Omega}{}^a_1$ is a first-order polynomial in momenta $\mathscr{P}_{\alpha_0|a}$, λ_{α_0}, $\mathscr{P}_{\alpha_s|aa_1\cdots a_s}$, $\lambda_{\alpha_s|a_1\cdots a_s}$, $s=1,...,L$. The property $\mathrm{ngh}(\overset{(1)}{\Omega}{}^a_1)=1$, the boundary conditions (31) and the requirement of Sp(2) covariance admit the following most general form for $\overset{(1)}{\Omega}{}^a_1$:

$$\begin{aligned}
\overset{(1)}{\Omega}{}^a_1=&T_{\alpha_0}c^{\alpha_0|a}+\epsilon^{ab}\mathscr{P}_{\alpha_0|b}\pi^{\alpha_0}\\
&+\sum_{s=1}^{L}(\epsilon^{ab}\mathscr{P}_{\alpha_s|ba_1\ldots a_s}N^{\alpha_s}_{s\beta_s}\pi^{\beta_s|a_1\cdots a_s}\\
&+\mathscr{P}_{\alpha_{s-1}|a_1\ldots a_s}Q^{\alpha_{s-1}}_{s\alpha_s}c^{\alpha_s|aa_1\cdots a_s}\\
&+\lambda_{\alpha_{s-1}|a_1\cdots a_{s-1}}M^{\alpha_{s-1}}_{s\alpha_s}\pi^{\alpha_s|aa_1\cdots a_{s-1}}),
\end{aligned} \tag{43}$$

where $N^{\alpha_s}_{s\beta_s}$, $Q^{\alpha_{s-1}}_{s\alpha_s}$, $M^{\alpha_{s-1}}_{s\alpha_s}$, $s=1,...,L$, are some unknown functions of the variables of the initial phase space $\eta=(p_i,q^i)$, and the matrices $N_s\equiv(N^{\alpha_s}_{s\beta_s})$ are assumed to be nonsingular.

One can simplify the structure of $\overset{(1)}{\Omega}{}^a_1$ (43) by making use of the fact that the canonical transformations are a natural arbitrariness in the solution of equations (30). We shall make a canonical transformation with the generating function

$$\begin{aligned}
X(P',Q)=&X_0(p',q;\mathscr{P}',c;\lambda'_0,\pi^0)\\
&+\sum_{s=1}^{L}\lambda'_{\alpha_s|a_1\cdots a_s}N^{\alpha_s}_{s\beta_s}\pi^{\beta_s|a_1\cdots a_s},
\end{aligned} \tag{44}$$

where X_0 is the generating function of the identity canonical transformation. Note that the canonical transformation

(44), first, does not change the structure of the boundary conditions (31) and, second, does not change the structure of $\overset{(1)}{\Omega}{}^a_1$, except that in the transformed $\overset{(1)}{\Omega}{}^a_1$ we have

$$N^{\alpha_s}_{s\beta_s} = \delta^{\alpha_s}_{\beta_s}. \tag{45}$$

Consequently, in considering the structure of $\overset{(1)}{\Omega}{}^a_1$ (43), we may assume without loss of generality that Eq. (45) is fulfilled.

Next we shall find out what restrictions on the matrices Q_s and M_s in (43) follow from Eqs. (30). Consider Eqs. (30) in the first order. Note that in this approximation the summands containing no momenta and those linear in momenta must vanish separately. This results in the following system of equations containing the matrices Q_s and M_s:

$$T_{\alpha_0} Q^{\alpha_0}_{1\alpha_1} = 0, \tag{46}$$

$$M^{\alpha_{s-1}}_{s\alpha_s} = -\frac{s}{s+1} Q^{\alpha_{s-1}}_{s\alpha_s} + T_{\beta_0} X^{\beta_0\alpha_{s-1}}_{s\alpha_s}, \quad s = 1,...,L, \tag{47}$$

$$Q^{\alpha_{s-2}}_{s-1\alpha_{s-1}} Q^{\alpha_{s-1}}_{s\alpha_s} = T_{\beta_0} Y^{\beta_0\alpha_{s-2}}_{s\alpha_s}, \quad s = 2,...,L, \tag{48}$$

with some functions $X^{\beta_0\alpha_{s-1}}_{s\alpha_s}$ and $Y^{\beta_0\alpha_{s-2}}_{s\alpha_s}$, and

$$X^{\beta_0\alpha_0}_{1\alpha_1} = -(-1)^{\epsilon_{\alpha_0}\epsilon_{\beta_0}} X^{\alpha_0\beta_0}_{1\alpha_1}, \quad Y^{\beta_0\alpha_0}_{2\alpha_2} = -(-1)^{\epsilon_{\alpha_0}\epsilon_{\beta_0}} Y^{\alpha_0\beta_0}_{2\alpha_2}. \tag{49}$$

First consider Eqs. (46). From (5) and (11) it follows that

$$Q^{\alpha_0}_{1\alpha_1} = Z^{\alpha_0}_{\beta_1} R^{\beta_1}_{\alpha_1} + T_{\beta_0} D^{\beta_0\alpha_0}_{\alpha_1}, \tag{50}$$

where $R^{\beta_1}_{\alpha_1}$ is a nonsingular supermatrix, and the functions $D^{\beta_0\alpha_0}_{\alpha_1}$ possess the properties (13). We shall make a canonical transformation with the generating function:

$$\begin{aligned} X(P',Q) = {} & X_0(p',q;\mathscr{P}'_1,c^1;\lambda'_1,\pi^1) \\ & + \mathscr{P}'_{\alpha_1|aa_1} R^{\alpha_1}_{\beta_1} c^{\beta_1|aa_1} + \lambda'_{\alpha_1|a} R^{\alpha_1}_{\beta_1} \pi^{\beta_1|a} \\ & - \frac{1}{2}(-1)^{\epsilon_{\alpha_0}} \mathscr{P}'_{\alpha_0|a_1} \mathscr{P}'_{\beta_0|a} D^{\beta_0\alpha_0}_{\alpha_1} c^{\alpha_1|aa_1}, \end{aligned} \tag{51}$$

where X_0 is the generating function of the identity transformation and the symbols $\mathscr{P}_1$, c^1, λ_1, π^1 imply that in the sets of momenta and coordinates one should, respectively, omit $\mathscr{P}_{\alpha_1|aa_1}$, $c^{\alpha_1|aa_1}$, $\lambda_{\alpha_1|a_1}$, $\pi^{\alpha_1|a_1}$. The canonical transformation (51) affects neither the structure of the boundary conditions (31), nor the structure of the terms $\mathscr{P}\pi$ in $\overset{(1)}{\Omega}{}^a_1$ [recall that $N^{\alpha_s}_{s\beta_s} = \delta^{\alpha_s}_{\beta_s}$ (45)], nor the structure of the terms $\mathscr{P}c$ and $\lambda\pi$ in $\overset{(1)}{\Omega}{}^a_1$, except that in the transformed $\overset{(1)}{\Omega}{}^a_1$ one should put

$$Q^{\alpha_0}_{1\alpha_1} = Z^{\alpha_0}_{\alpha_1}. \tag{52}$$

Consequently, without loss of generality, one can choose (52) as a solution of Eq. (46). In the next step, Eqs. (48) and (47) with account taken of (52) give

$$Z^{\alpha_0}_{\alpha_1} Q^{\alpha_1}_{2\alpha_2} = T_{\beta_0} Y^{\beta_0\alpha_0}_{2\alpha_2}, \tag{53}$$

$$M^{\alpha_0}_{1\alpha_1} = -\tfrac{1}{2} Z^{\alpha_0}_{\alpha_1} + T_{\beta_0} X^{\beta_0\alpha_0}_{1\alpha_1}. \tag{54}$$

The solution of Eqs. (53) can be represented in the form [see (8) and (12)]

$$Q^{\alpha_1}_{2\alpha_2} = Z^{\alpha_1}_{\beta_2} R^{\beta_2}_{\alpha_2} + T_{\beta_0} D^{\beta_0\alpha_1}_{\alpha_2}, \tag{55}$$

where $R^{\beta_2}_{\alpha_2}$ is a reversible supermatrix. We shall make a canonical transformation with the generating function

$$\begin{aligned} X(P',Q) = {} & X_0(p',q;\mathscr{P}'_2,c^2;\lambda'_2,\pi^2) \\ & + \mathscr{P}'_{\alpha_2|aa_1a_2} R^{\alpha_2}_{\beta_2} c^{\beta_2|aa_1a_2} + \lambda'_{\alpha_2|a_1a_2} R^{\alpha_2}_{\beta_2} \pi^{\beta_2|a_1a_2} \\ & + (-1)^{\epsilon_{\alpha_1}} \mathscr{P}'_{\alpha_1|a_1a_2} \mathscr{P}'_{\beta_0|a} D^{\beta_0\alpha_1}_{\alpha_2} c^{\alpha_2|aa_1a_2} \\ & + (-1)^{\epsilon_{\alpha_0}} \lambda'_{\alpha_0} \mathscr{P}'_{\beta_0|a} X^{\beta_0\alpha_0}_{1\alpha_1} \pi^{\alpha_1|a}. \end{aligned} \tag{56}$$

The canonical transformation (56) affects neither the boundary conditions (31), nor the condition (52), nor the structure of $\overset{(1)}{\Omega}{}^a_1$, except that in the transformed $\overset{(1)}{\Omega}{}^a_1$ one should put

$$Q^{\alpha_1}_{2\alpha_2} = Z^{\alpha_1}_{\alpha_2}, \quad M^{\alpha_0}_{1\alpha_1} = -(1/2) Z^{\alpha_0}_{\alpha_1}. \tag{57}$$

Continuing this process, we finally establish that in solving Eqs. (31) one can always choose for $\overset{(1)}{\Omega}{}^a_1$ the following expression:

$$\begin{aligned} \overset{(1)}{\Omega}{}^a_1 = {} & T_{\alpha_0} c^{\alpha_0|a} + \epsilon^{ab} \mathscr{P}_{\alpha_0|b} \pi^{\alpha_0} \\ & + \sum_{s=1}^{L} \Big(\epsilon^{ab} \mathscr{P}_{\alpha_s|ba_1\cdots a_s} \pi^{\alpha_s|a_1\cdots a_s} \\ & + \mathscr{P}_{\alpha_{s-1}|a_1\cdots a_s} Z^{\alpha_{s-1}}_{\alpha_s} c^{\alpha_s|aa_1\cdots a_s} \\ & - [s/(s+1)] \lambda_{\alpha_{s-1}|a_1\cdots a_{s-1}} Z^{\alpha_{s-1}}_{\alpha_s} \pi^{\alpha_s|aa_1\cdots a_{s-1}} \Big). \end{aligned} \tag{58}$$

The expression (58) will be called the canonical form for $\overset{(1)}{\Omega}{}^a_1$.

We have proved, in fact, that if the solution of Eqs. (30) with the boundary conditions (31) exists then it is canonically equivalent to the solution for which $\overset{(1)}{\Omega}{}^a_1$ has the canonical form (58) for arbitrary choice of the vectors $Z^{\alpha_{s-1}}_{\alpha_s}$, $s = 1,...,L$. Finally, let us consider the canonical transformation with generating function:

$$X = \mathscr{P}'_{\alpha_0|a} (M^{-1})^{\alpha_0}_{\beta_0} c^{\alpha_0|a} + \lambda'_{\alpha_0} (M^{-1})^{\alpha_0}_{\beta_0} \pi^{\beta_0} + X_0,$$

where X_0 is the generating function of identity canonical transformation for all the variables excepting $\mathscr{P}_{\alpha_0|a}$, $c^{\alpha_0|a}$, $\lambda_{\alpha_0}, \pi^{\alpha_0}$. Under such a transformation the solution of Eq. (30) with the boundary conditions (31) becomes the solution with the boundary condition of the form (31), but $T'_{\alpha_0} = T_{\beta_0} M^{\beta_0}_{\alpha_0}$ appears instead of T_{α_0}. Let us choose the matrix $M^{\beta_0}_{\alpha_0}$ from (15). Then T'_{α_0} coincides with the choice of the constraints in standard basis. Therefore, any solution of Eq. (30) is canonically equivalent to that with the choice of constraints in standard basis. The solution of Eq. (30) will be called standard, Ω_{st}, if it has the canonical form with the constraints T_{α_0} and vectors $Z^{\alpha_{s-1}}_{\alpha_s}$, $s = 1,...,L$, in standard basis of (15) and (21). Above considerations mean that any solution of Eq. (30) with the boundary conditions (31) is canonically equivalent to some Ω_{st}.

The possibility of choosing $\overset{(1)}{\Omega}{}^a_1$ in the canonical form

(58) can be formulated as additional boundary conditions to equations (30):

$$\left.\frac{\delta\Omega^a}{\delta\pi^{\alpha_s|a_1\cdots a_s}}\right|_{c=\pi=\mathscr{P}_s=\lambda=0}=\epsilon^{ab}\mathscr{P}_{\alpha_s|ba_1\cdots a_s}, \tag{59}$$

$$\left.\frac{\delta\Omega^a}{\delta\mathscr{P}_{\alpha_{s-1}|a_1\cdots a_s}}\right|_{c^A=\pi=\mathscr{P}=\lambda=0}=Z^{\alpha_{s-1}}_{\alpha_s}c^{\alpha_s|aa_1\cdots a_s}, \tag{60}$$

$$\left.\frac{\delta\Omega^a}{\delta\lambda_{\alpha_{s-1}|a_1\cdots a_{s-1}}}\right|_{c=\pi^A=\mathscr{P}=\lambda=0}=-\frac{s}{s+1}Z^{\alpha_{s-1}}_{\alpha_s}\pi^{\alpha_s|aa_1\cdots\alpha_{s-1}},\ s=1,..,L. \tag{61}$$

We now turn directly to the proof of the existence theorem for solutions of equations (30) with the boundary conditions (31), (59)–(61). We shall make a canonical transformation of Ω^a to $\Omega^a_{\rm st}$ in the standard basis of (15) and (21) and choose

$$\Omega^a_{1\rm st}=\overset{(1)}{\Omega}{}^a_{1\rm st}=T_{A_0}c^{A_0|a}+\epsilon^{ab}\mathscr{P}_{A_0|b}\pi^{A_0}+\sum_{s=1}^{L}\Big(\epsilon^{ab}\mathscr{P}_{A_s|ba_1\cdots a_s}\pi^{A_s|a_1\cdots a_s}+\epsilon^{ab}\mathscr{P}_{A_s|ba_1\cdots a_{s-1}}\pi^{A_s|a_1\cdots a_{s-1}}$$
$$+\mathscr{P}_{A_s|a_1\cdots a_s}c^{A_s|aa_1\cdots a_s}-\frac{s}{s+1}\lambda_{A_s|a_1\cdots a_{s-1}}\pi^{A_s|aa_1\cdots a_{s-1}}\Big). \tag{62}$$

It is a remarkable fact that $\Omega^a_{1\rm st}$ (62) is in itself a solution of equations (30) in the first order. Starting now with (62), it is not difficult to prove the existence of solutions $\Omega^a_{\rm st}$ of (30) for Ω^a in the standard basis. Indeed, suppose we have quantities $\Omega^a_{n\rm st}$ such that Eqs. (30) are satisfied in the nth order. We shall find the expression for $\{\Omega^a_{\rm st},\Omega^b_{\rm st}\}$ in the $(n+1)$ order

$$\{\Omega^a_{\rm st},\Omega^b_{\rm st}\}_{n+1}=W^{\{a}\Omega^{b\}}_{n+1\,\rm st}+B^{ab}_{n+1}, \tag{63}$$

where the functions B^{ab}_{n+1} are constructed from $\Omega^a_{k\rm st}$, $k\leqslant n$, and possess the symmetry properties $B^{ab}_{n+1}=B^{ba}_{n+1}$. The operators W^a in (63) have the following form:

$$W^a=T_{A_0}\frac{\delta}{\delta\mathscr{P}_{A_0|a}}+\epsilon^{ab}\mathscr{P}_{A_0|b}\frac{\delta}{\delta\lambda_{A_0}}+\sum_{s=1}^{L}\Big(\epsilon^{ab}\mathscr{P}_{A_s|ba_1\cdots a_s}\frac{\delta}{\delta\lambda_{A_s|a_1\cdots a_s}}+\epsilon^{ab}\mathscr{P}_{A_s|ba_1\cdots a_{s-1}}\frac{\delta}{\delta\lambda_{A_s|a_1\cdots a_{s-1}}}$$
$$+\mathscr{P}_{A_s|a_1\cdots a_s}\frac{\delta}{\delta\mathscr{P}_{A_s|aa_1\cdots a_s}}-\frac{s}{s+1}\lambda_{A_s|a_1\cdots a_{s-1}}\frac{\delta}{\delta\lambda_{A_s|aa_1\cdots a_{s-1}}}\Big)+\epsilon^{ab}(-1)^{\epsilon_{A_0}}\pi^{A_0}\frac{\delta_l}{\delta c^{A_0|b}}$$
$$+\sum_{s=1}^{L}\Big(\epsilon^{ab}(-1)^{(\epsilon_{A_s}+s)}\pi^{A_s|a_1\cdots a_s}\frac{\delta_l}{\delta c^{A_s|ba_1\cdots a_s}}+\epsilon^{ab}(-1)^{(\epsilon_{A_s}+s)}\pi^{A_s|a_1\cdots a_{s-1}}\frac{\delta_l}{\delta c^{A_s|ba_1\cdots a_{s-1}}}$$
$$+(-1)^{(\epsilon_{A_s}+s+1)}c^{A_s|aa_1\cdots a_s}\frac{\delta_l}{\delta c^{A_s|a_1\cdots a_s}}-\frac{s}{s+1}(-1)^{(\epsilon_{A_s}+s)}\pi^{A_s|aa_1\cdots a_{s-1}}\frac{\delta_l}{\delta\pi^{A_s|a_1\cdots a_{s-1}}}\Big). \tag{64}$$

Through a direct verification we make sure that W^a (64) form a set of nilpotent anticommuting operators

$$W^{\{a}W^{b\}}=0. \tag{65}$$

The quantities B^{ab}_{n+1} satisfy the equations

$$W^aB^{bc}_{n+1}+\mathrm{perm}(abc)=0, \tag{66}$$

which follow from the Jacobi identities for $\Omega^a_{\rm st}$ calculated in the $(n+1)$ order [(65) should be taken into account].

All further steps in the proof are based on the following lemma.

Lemma: Any regular solution of equations

$$W^aX=0, \tag{67}$$

$$W^{\{a}X^{a_1\cdots a_s\}}=0, \tag{68}$$

which vanishes when $T=\mathscr{P}=\lambda=0$ has, respectively, the form

$$X=\tfrac{1}{2}\epsilon_{ab}W^aW^bY, \tag{69}$$

$$X^{a_1\cdots a_n}=W^{\{a_1}Y^{a_2\cdots a_n\}}, \tag{70}$$

where Y, $Y^{a_1\cdots a_{n-1}}$ are some functions of the variables of the extended phase space (22). Moreover, if the functions X and $X^{a_1\cdots a_n}$ are Sp(2) scalar and tensor, respectively, then the functions Y and $Y^{a_2\cdots a_n}$ also can be chosen as Sp(2) scalar and symmetric tensor. In (68) the functions $X^{a_1\cdots a_n}$ are assumed to be symmetric under permutation of any indices. The symbol $\{a_1a_2\cdots a_n\}$ in (68) and (70) implies cyclic permutation of the indices $a_1,a_2,\cdots,a_n$. The proof of the Lemma is given in Ref. 1 and is based on the possibility of constructing operators Γ_a which constitute together with W^a, an algebra of the form

$$\Gamma_{\{a}\Gamma_{b\}}=0,\quad W^a\Gamma_b+\Gamma_bW^a=\delta^a_bN, \tag{71}$$

where N is a scalar operator under the group Sp(2). The solution of equations (71) is existent, for example,

$$\Gamma_a=\mathscr{P}_{A_0|a}\frac{\delta_l}{\delta T_{A_0}}-\epsilon_{ab}\lambda_{A_0}\frac{\delta}{\delta\mathscr{P}_{A_0|b}}+\sum_{s=1}^{L}\Big(\mathscr{P}_{A_s|aa_1\cdots a_s}$$
$$\times\frac{\delta}{\delta\mathscr{P}_{A_s|a_1\cdots a_s}}-\frac{s}{s+1}\epsilon_{ab}\lambda_{A_s|a_1\cdots a_{s-1}}$$

$$\times\frac{\delta}{\delta\mathscr{P}_{A_s|ba_1\cdots a_{s-1}}}$$

$$-\epsilon_{ab}\lambda_{A_s|a_1\cdots a_s}\frac{\delta}{\delta\mathscr{P}_{A_s|ba_1\cdots a_s}}$$

$$-\lambda_{A_s|aa_1\cdots a_{s-1}}\frac{\delta}{\delta\lambda_{A_s|a_1\cdots a_{s-1}}}\Bigg). \tag{72}$$

The operator N, is, in fact, a "conformal generator:"

$$N=T_{A_0}\frac{\delta_l}{\delta T_{A_0}}+\sum_{s=0}^{L}\Bigg(\mathscr{P}_{A_s|a_1\cdots a_{s+1}}\frac{\delta}{\delta\mathscr{P}_{A_s|a_1\cdots a_{s+1}}}$$

$$+\lambda_{A_s|a_1\cdots a_s}\frac{\delta}{\delta\lambda_{A_s|a_1\cdots a_s}}\Bigg)+\sum_{s=1}^{L}\Bigg(\mathscr{P}_{A_s|a_1\cdots a_s}$$

$$\times\frac{\delta}{\delta\mathscr{P}_{A_s|a_1\cdots a_s}}+\lambda_{A_s|a_1\cdots a_{s-1}}\frac{\delta}{\delta\lambda_{A_s|a_1\cdots a_{s-1}}}\Bigg). \tag{73}$$

Returning to the proof of existence of solutions of equations (30) in the standard basis, we note that the functions B^{ab}_{n+1} possess the properties $B^{ab}_{n+1}=0$ when $T=\mathscr{P}=\lambda=0$ (the proof of this fact is quite analogous to the arguments presented in Ref. 1). Then by virtue of the lemma, (66) implies the existence of functions Y^a_{n+1} such that

$$B^{ab}_{n+1}=W^{\{a}Y^{b\}}_{n+1}. \tag{74}$$

Choosing

$$\Omega^a_{n+1\,\mathrm{st}}=-Y^a_{n+1}, \tag{75}$$

we obtain that Eqs. (30) in the standard basis are satisfied already in the $(n+1)$ order. Using induction we state that the solution of Eqs. (30) in the standard basis does exist. To complete the proof of the existence theorem for solutions of Eqs. (30) with the boundary conditions (31), (59)–(61), it suffices to note that passing over from the standard basis to the one in which $\overset{(1)}{\Omega}{}^a_1$ has the canonical form (58) is realized via a canonical transformation.

The situation with the existence of solution of Eqs. (32) with the boundary condition (33) is quite similar. Namely, the proof of the existence of solution of Eqs. (32) in the standard basis repeats literally the arguments presented in Ref. 1. After this, to make sure of the existence of solution of Eqs. (32) with the boundary condition (33), it suffices to make a canonical transformation and pass over from the standard basis to the one in which $\overset{(1)}{\Omega}{}^a_1$ has the canonical form (58). As Ω^a, in these equations there appear solutions of Eqs. (30) with the boundary conditions (31), (59)–(61).

V. ARBITRARINESS IN SOLUTIONS OF GENERATING EQUATIONS

We describe here the arbitrariness that exists in the solutions of the generating Eqs. (30) and (32). Let us first consider Eqs. (30). Let $\widetilde{\Omega}^a$ and Ω^a be any two solutions of these equations with the boundary conditions (59)–(61). Using canonical transformations, we pass over to the standard basis in the solutions $\widetilde{\Omega}^a$ and Ω^a; given this,

$$\widetilde{\Omega}^a\to\widetilde{\Omega}^a_{\mathrm{st}},\quad \Omega^a\to\Omega^a_{\mathrm{st}}.$$

$\widetilde{\Omega}^a_{\mathrm{st}}$ and Ω^a_{st} satisfy Eqs. (30) and

$$\overset{(1)}{\widetilde{\Omega}}{}^a_{\mathrm{st}}=\overset{(1)}{\Omega}{}^a_{\mathrm{st}}. \tag{76}$$

Suppose now that

$$\widetilde{\Omega}^a_{\mathrm{st}}=\Omega^a_{\mathrm{st}}+\Delta\Omega^a_{\mathrm{st}}, \tag{77}$$

where $\Delta\Omega^a_{\mathrm{st}}$ is the function whose expansion in power series of momenta starts at least with the second order. In the first order in c and π the function $\Delta\Omega^a_{1\mathrm{st}}$ satisfies the equations

$$w^{\{a}\Delta\Omega^{b\}}_{1\mathrm{st}}=0, \tag{78}$$

where the operators w^a act on any function X by the rule

$$w^aX=\{\Omega^a_{1\mathrm{st}},X\}_{\mathrm{gh}}, \tag{79}$$

where $\Omega^a_{1\mathrm{st}}$ is the first approximation for Ω^a_{st}, and the sign gh means that the Poisson superbracket is calculated with respect to the canonical variables that complement the initial phase space $\eta=(p_i,q^i)$ to the extended one (22). The Jacobi identity for Ω^a_{st}, Ω^b_{st}, X implies that the operators w^a possess the following nilpotence properties

$$w^{\{a}w^{b\}}=0. \tag{80}$$

We shall now show that any solution of the equations

$$w^{\{a}X^{b\}}=0 \tag{81}$$

vanishing when $T_{A_0}=\mathscr{P}=\lambda=0$ has the form

$$X^a=w^aY \tag{82}$$

with a certain function Y. Moreover, if the function X^a is Sp(2) vector, then the function Y can be chosen as Sp(2) scalar. With this purpose we represent the operators w^a in the form

$$w^a=W^a+\sum_{l=1}^{L}W^a_{(l)}$$

where W^a is defined in (64) and $W^a_{(l)}$ are operators raising by l the degree of momenta of the monomials on which they act. We shall solve Eqs. (81) in the form of power series of $\mathscr{P},\lambda,T_{A_0}$. We denote the lower order for X^a by X^a_0 and let it be equal to a certain number m. Then from (81) it follows that

$$W^{\{a}X^{b\}}_0=0, \tag{83}$$

where $X^b_0=0$ for $\mathscr{P}=\lambda=T_{A_0}=0$. Applying our lemma to (78), we have

$$X^a_0=W^aY_0. \tag{84}$$

Now

$$X^a=W^aY_0+\widetilde{X}^a_1=w^aY_0+X^a_1, \tag{85}$$

where X^a_1 has the order not less than $m+1$ with respect to the powers of variables $\mathscr{P},\lambda,T_{A_0}$. Thus in the mth order the solutions of Eqs. (81) can be represented in the form (82). Let X^a_{10} be the lowest order for X^a_1. Then

$$W^{\{a}X^{b\}}_{10}=0, \tag{86}$$

and, therefore,

$$X^a_{10}=W^aY_1. \tag{87}$$

One can see that

$$X^a = w^a Y_0 + W^a Y_1 + \tilde{X}^a_2 = w^a(Y_0 + Y_1) + X^a_2, \tag{88}$$

i.e., solutions of Eqs. (81) can be represented in the form (82) with an accuracy already to the terms of order $m+1$ in $\mathscr{P}$, λ, T_{A_0}. Continuing this process, we make sure that the statement expressed above is valid.

Returning to (78), we have

$$\Delta\Omega^a_{1st} = w^a X_{1st} \tag{89}$$

with a certain function X_{1st}. Repeating literally the corresponding arguments from Ref. 1, we first derive from (89) that in the first order the solutions $\tilde{\Omega}^a_{st}$ and Ω^a_{st} are related through a canonical transformation and then we conclude that the exact $\tilde{\Omega}^a_{st}$ and Ω^a_{st} are also related through a canonical transformation. Making then a canonical transformation from the standard basis to the one in which $\overset{(1)}{\Omega}{}^a_1$ has the canonical form (58), we come to the conclusion that any two solutions of Eqs. (30) with the boundary conditions (31), (59)–(61) are related to each other by means of a canonical transformation.

By analogous arguments one can establish that any two solutions $\mathscr{H}$ and $\tilde{\mathscr{H}}$ of Eqs. (32) with the boundary condition (33) are related as

$$\tilde{\mathscr{H}} = \mathscr{H} + (1/2)\epsilon_{ab}\{\Omega^b,\{\Omega^a,Y\}\}, \tag{90}$$

i.e., the arbitrariness in the solutions of Eqs. (32) (with given Ω^a) corresponds to the change of the gauge in the total unitarizing Hamiltonian.

We have hitherto discussed arbitrariness in solutions of the generating Eqs. (30) and (32) with fixed boundary conditions that are determined by the choice of T_{α_0}, $Z^{\alpha_{s-1}}_{\alpha_s}$, H_0. It is well known that classical dynamics do not depend on linear combination of constraints T_{α_0} and on addition to the Hamiltonian H_0 of a linear combination of constraints T_{α_0}: $H'_0 = H_0 + \Lambda^{\alpha_0} T_{\alpha_0}$. Besides, we have seen that $Z^{\alpha_{s-1}}_{\alpha_s}$ from Eqs. (5), (6), and (8) are not uniquely defined [see (11) and (12)]. We shall consider how the solutions of Eqs. (30) and (32) corresponding to the boundary conditions with T_{α_0}, $Z^{\alpha_{s-1}}_{\alpha_s}$ H_0, and T'_{α_0}, $Z'^{\alpha_{s-1}}_{\alpha_s}$, H'_0 are related:

$$T'_{\alpha_0} = T_{\beta_0} M^{\beta_0}_{\alpha_0}, \quad Z'^{\alpha_{s-1}}_{\alpha_s} = Z^{\alpha_{s-1}}_{\beta_s} M^{\beta_s}_{\alpha_s} + T_{\beta_0} D^{\beta_0 \alpha_{s-1}}_{\alpha_s},$$
$$H'_0 = H_0 + \Lambda^{\alpha_0} T_{\alpha_0}. \tag{91}$$

We denote these solutions by Ω^a, Ω'^a_a and $\mathscr{H}$, $\mathscr{H}'$, respectively. Note that Ω^a and Ω'^a are canonically equivalent to the solution of Eqs. (30) $\tilde{\Omega}^a_1$ in the standard basis with $\overset{(1)}{\Omega}{}^a_1 = \overset{(1)}{\Omega}{}^a_{1st}$ (62), and therefore one can state that Ω'^a and Ω^a are related to each other via a canonical transformation. Similarly, the solutions $\mathscr{H}$ and $\mathscr{H}'$ are related to each other (with fixed Ω^a) through a change in the gauge of the form (90).

Finally, the whole arbitrariness existing in the solutions of Eqs. (30) is described by a canonical transformation, and the arbitrariness in (32) is transformed into a change of the gauge in the unitarizing Hamiltonian.

Moreover, as in the irreducible case,[1] one can show that there exist functions Ξ_a, Θ which form, together with Ω^a, the complex

$$\{\Omega^a,\Omega^b\} = \{\Xi_a,\Xi_b\} = \{\Omega^a,\Theta\} = \{\Xi_a,\Theta\} = 0,$$
$$\{\Omega^a,\Xi_b\} = \delta^a_b \Theta.$$

VI. EQUIVALENCE OF THE STANDARD AND Sp(2)-COVARIANT QUANTIZATIONS OF REDUCIBLE THEORIES

We discuss here the connection between the formalism of quantization of reducible gauge theories, developed here, with the method of Refs. 3 and 4. Recall that the extended phase space Γ of the gauge theory of the L th stage of reducibility is defined in Refs. 3 and 4 in the form

$$\Gamma = (P_A,Q^A)$$
$$= (p_i,q^i;\overline{\mathscr{P}}^{s}_{\alpha_s},c^{\alpha_s}_s,s=0,1,...,L;\bar{c}^{s}_{\alpha_s},$$
$$\mathscr{P}^{\alpha_s}_s,s=0,1,...,L;\pi^{s}_{\alpha_s},\lambda^{\alpha_s}_s,s=0,1,...,L;\pi^{s's}_{s\alpha_s},\lambda^{s'\alpha_s}_s,$$
$$s'=1,...,L,s=s',s'+1,...,L;\bar{c}^{s's}_{s\alpha_s},\mathscr{P}^{s'\alpha_s}_s,s'=1,...,L,$$
$$s=s',s'+1,...,L), \tag{92}$$

where the Grassmann parities and ghost numbers for the newly introduced variables are, respectively, equal to

$$\epsilon(c^{\alpha_s}_s) = \epsilon(\overline{\mathscr{P}}^{s}_{\alpha_s}) = \epsilon_{\alpha_s} + s + 1 (\mathrm{mod}\ 2), s=0,1,...,L,$$
$$\mathrm{gh}(c^{\alpha_s}_s) = -\mathrm{gh}(\overline{\mathscr{P}}^{s}_{\alpha_s}) = s+1,$$
$$\epsilon(\bar{c}^{s}_{\alpha_s}) = \epsilon(\mathscr{P}^{\alpha_s}_s) = \epsilon_{\alpha_s} + s + 1 (\mathrm{mod}\ 2), s=0,1,...,L,$$
$$-\mathrm{gh}(\bar{c}^{s}_{\alpha_s}) = \mathrm{gh}(\mathscr{P}^{\alpha_s}_s) = s+1,$$
$$\epsilon(\pi^{s}_{\alpha_s}) = \epsilon(\lambda^{\alpha_s}_s) = \epsilon_{\alpha_s} + s (\mathrm{mod}\ 2), s=0,1,...,L,$$
$$-\mathrm{gh}(\pi^{s}_{\alpha_s}) = \mathrm{gh}(\lambda^{\alpha_s}_s) = s,$$
$$\epsilon(\pi^{s's}_{s\alpha_s}) = \epsilon(\lambda^{s'\alpha_s}_s) = \epsilon_{\alpha_s} + s - s' (\mathrm{mod}\ 2),$$
$$\mathrm{gh}(\pi^{s's}_{s\alpha_s}) = -\mathrm{gh}(\lambda^{s'\alpha_s}_s)$$
$$= -(s-s'), s'=1,...,L, s=s',s'+1,...,L,$$
$$\epsilon(\bar{c}^{s's}_{s\alpha_s}) = \epsilon(\mathscr{P}^{s'\alpha_s}_s) = \epsilon_{\alpha_s} + s - s' + 1 (\mathrm{mod}\ 2),$$
$$-\mathrm{gh}(\bar{c}^{s's}_{s\alpha_s}) = \mathrm{gh}(\mathscr{P}^{s'\alpha_s}_s)$$
$$= s - s' + 1, s'=1,...,L, s=s',s'+1,...,L. \tag{93}$$

Next, in the entire phase space (92), the fermion Ω and the boson $\mathscr{H}$ functions are defined as special solutions of the following generating equations:

$$\{\Omega,\Omega\} = 0, \tag{94}$$
$$\{\mathscr{H},\Omega\} = 0. \tag{95}$$

For our further purposes we do not need a detailed knowledge of the structure Ω and $\mathscr{H}$ and so we do not discuss them here but refer the reader to the original papers of Ref. 3 and 4 and the review, Ref. 14.

The first step in establishing the required correspondence consists in identification of the extended phase spaces (92) and (22). One of the possible ways of such an identification [with allowance for (23), (24), and (93)] is as follows:

(1) for s even

$$c^{a_1a_2\cdots a_{s+1}}=\Big(c^{11\cdots1}=c_s,\ldots,c^{\underbrace{11\cdots1}_{n}\underbrace{2\cdots2}}=\mathscr{P}_s^{2n},\ldots,$$

$$c^{\underbrace{11\ldots1}_{m}\underbrace{2\ldots2}}=\pi_s^{2s-2m+1},\ldots,c^{22\cdots2}=\bar{c}^s\Big),$$

$$\mathscr{P}_{a_1a_2\cdots a_{s+1}}=\Big(\mathscr{P}_{11\ldots1}=\bar{\mathscr{P}}^s,\ldots,\mathscr{P}_{\underbrace{11\cdots1}_{n}\underbrace{2\cdots2}}=\bar{c}_s^{2n},\ldots,$$

$$\mathscr{P}_{\underbrace{11\cdots1}_{m}\underbrace{2\cdots2}}=\lambda_s^{2s-2m+1},\ldots,\mathscr{P}_{22\cdots2}=\mathscr{P}_s\Big),$$

$$1\leqslant n\leqslant s/2,\quad s/2<m\leqslant s,$$

$$\pi^{a_1\cdots a_s}=\Big(\pi^{11\cdots1}=\lambda_s,\ldots,\pi^{\underbrace{11\cdots1}_{n}\underbrace{2\cdots2}}=\lambda_s^{2n},\ldots,$$

$$\pi^{\underbrace{11\cdots1}_{m}\underbrace{2\cdots2}}=\bar{c}_s^{2s-2m+1},\ldots,\pi^{22\cdots2}=\bar{c}_s^1\Big),$$

$$\lambda_{a_1\cdots a_s}=\Big(\lambda_{11\cdots1}=\pi^s,\ldots,\lambda_{\underbrace{11\cdots1}_{n}\underbrace{2\cdots2}}=\pi_s^{2n},\ldots,$$

$$\lambda_{\underbrace{11\cdots1}_{m}\underbrace{2\cdots2}}=\mathscr{P}_s^{2s-2m+1},\ldots,\lambda_{22\cdots2}=\mathscr{P}_s^1\Big),$$

$$1\leqslant n\leqslant s/2,\quad s/2<m\leqslant s-1; \tag{96}$$

(2) for s odd

$$c^{a_1a_2\cdots a_{s+1}}=\Big(c^{11\cdots1}=c_s,\ldots,c^{\underbrace{11\cdots1}_{n}\underbrace{2\cdots2}}=\mathscr{P}_s^{2n},\ldots,$$

$$c^{\overbrace{1\cdots1}^{(s+1)/2}\overbrace{2\cdots2}^{(s+1)/2}}=\lambda_{\cdot}^{s},\ldots,c^{\underbrace{11\cdots1}_{m}\underbrace{2\cdots2}}=\pi_{\cdot}^{2s-2m+1},$$

$$c^{22\cdots2}=\bar{c}^s\Big),$$

$$\mathscr{P}_{a_1a_2\cdots a_{s+1}}=\Big(\mathscr{P}_{11\cdots1}=\bar{\mathscr{P}}^s,\ldots,\mathscr{P}_{\underbrace{11\cdots1}_{n}\underbrace{2\cdots2}}=\bar{c}^{2n},\ldots,$$

$$\mathscr{P}_{\underbrace{1\ldots1}_{(s+1)/2}\underbrace{2\ldots2}_{(s+1)/2}}=\pi_s^s,\ldots,\mathscr{P}_{\underbrace{11\cdots1}_{m}\underbrace{2\cdots2}}=\lambda_s^{2s-2m+1},\ldots,$$

$$\mathscr{P}_{22\cdots2}=\mathscr{P}_s\Big), \tag{97}$$

$$1\leqslant n\leqslant(s-1)/2,(s+1)/2<m\leqslant s,$$

$$\pi^{a_1\cdots a_s}=\Big(\pi^{11\cdots1}=\lambda_s,\ldots,\pi^{\underbrace{11\cdots1}_{n}\underbrace{2\cdots2}}=\lambda_s^{2n},\ldots,$$

$$\pi^{\underbrace{11\cdots1}_{m}\underbrace{2\cdots2}}=\bar{c}^{2s-2m+1},\ldots,\pi^{22\cdots2}=\bar{c}_s^1\Big),$$

$$\lambda_{a_1\cdots a_s}=\Big(\lambda_{11\cdots1}=\pi_s,\ldots,\lambda_{\underbrace{11\cdots1}_{n}\underbrace{2\cdots2}}=\pi_s^{2n},\ldots,$$

$$\lambda_{\underbrace{11\cdots1}_{m}\underbrace{2\cdots2}}=\mathscr{P}_s^{2s-2m+1},\ldots,\lambda_{22\cdots2}=\mathscr{P}_s^1\Big),$$

$$1\leqslant n\leqslant(s-1)/2,(s+1)/2\leqslant m\leqslant s-1.$$

Note that any other identification of the extended phase spaces (22) and (92) differs from (96) and (97) by a canonical transformation.

Next, one can see that the function Ω^1 satisfies Eq. (94) in the same extended phase space as Ω. But the solutions of Eqs. (94) are related to each other via a canonical transformation. Consequently, one can state that Ω^1 is canonically equivalent to Ω. Then the boson function $\mathscr{H}$ constructed as a special solution of Eq. (95) can differ from the boson function $\mathscr{H}$ satisfying (32) only by the change of the gauge in the unitarizing Hamiltonian. The general conclusion is that the Sp(2)-covariant quantization of reducible theories is a particular case of the standard version that corresponds to a special choice of the basis of canonical variables and the gauge.

VII. AN ALTERNATIVE APPROACH TO THE Sp(2)-COVARIANT QUANTIZATION OF REDUCIBLE GAUGE THEORIES

In addition to the formalism described above, one can construct an alternative quantization scheme based on the approach to reducible theories developed in Ref. 5. The crucial point of the paper was to represent, by adding new constraints, the initial dynamical system in terms of an equivalent system with linearly independent constraints. With this purpose it was proposed to extend the initial phase space $\eta=(p_i,q^i)$ up to $\tilde{\Gamma}_0$ by way of introducing new (nonghost!) canonical variables:

$$\tilde{\Gamma}_0=(p_i,q^i;p_{\alpha_{2s+1}},q^{\alpha_{2s+1}};s=0,1,\ldots,[(L-1)/2]), \tag{98}$$

where $[a]$ denotes the integral part of the number a, and

$$\epsilon(p_{\alpha_{2s+1}})=\epsilon(q^{\alpha_{2s+1}})\equiv\epsilon_{\alpha_{2s+1}},$$

$$\mathrm{gh}(p_{\alpha_{2s+1}})=\mathrm{gh}(q^{\alpha_{2s+1}})=0,\quad s=0,1,\ldots,[(L-1)/2]. \tag{99}$$

In the phase space (98), the Hamiltonian of the system $\tilde{H}_0$ is defined as

$$\tilde{H}_0=H_0, \tag{100}$$

i.e., it simply coincides with the initial Hamiltonian. We introduce a set of new constraints $\tilde{T}_{\alpha_{2s}}$, $s=0,1,\ldots,[(L-1)/2]$, by the rule $(\epsilon(\tilde{T}_{\alpha_{2s}})=\epsilon_{\alpha_{2s}})$:

$$\tilde{T}_{\alpha_0}=T_{\alpha_0}+p_{\alpha_1}(Z^T)^{\alpha_1}_{\alpha_0},$$

$$\tilde{T}_{\alpha_{2s}}=p_{\alpha_{2s-1}}Z^{\alpha_{2s-1}}_{\alpha_{2s}}+p_{\alpha_{2s+1}}(Z^T)^{\alpha_{2s+1}}_{\alpha_{2s}},s=1,\ldots,[(L-1)/2]. \tag{101}$$

Here, the superscript "T" implies matrix transposition, and the functions $Z^{\alpha_{s-1}}_{\alpha_s}$, as functions of the initial canonical variables, satisfy Eqs. (5) and (8). In Ref. 5 it is shown that the set of constraints (101) in the space (98) is linearly independent, and it is also established that the standard quantization of the system (100) and (101) is equivalent to quantization of the initial system, when from among the set of constraints T_{α_0} one chooses linearly independent ones.

The dynamical system described in the phase space (98) by the Hamiltonian (100) and by the set of first-class constraints (101) is irreducible. To quantize this system in the framework of the Sp(2)-covariant approach, one should introduce an extended phase space:

$$\tilde{\Gamma}=(P_A,Q^A)=(p_i,q^i;p_{\alpha_{2s+1}},q^{\alpha_{2s+1}};\mathscr{P}_{\alpha_0|a},c^{\alpha_0|a};$$

$$\mathscr{P}_{\alpha_{2s}|a},c^{\alpha_{2s}|a};\lambda_{\alpha_0},\pi^{\alpha_0};\lambda_{\alpha_{2s}},\pi^{\alpha_{2s}}), \tag{102}$$

where

$$\epsilon(c^{\alpha_0|a})=\epsilon_{\alpha_0}+1,\quad \epsilon(c^{\alpha_{2s}|a})=\epsilon_{\alpha_{2s}}+1,$$

$$\epsilon(\pi^{\alpha_0})=\epsilon_{\alpha_0},\quad \epsilon(\pi^{\alpha_{2s}})=\epsilon_{\alpha_{2s}},$$

$$\mathrm{gh}(c^{\alpha_0|a})=-\mathrm{gh}(\mathscr{P}_{\alpha_0|a})=3-2a,$$

$$
\begin{aligned}
&\mathrm{gh}(c^{\alpha_{2s}|a}) = -\mathrm{gh}(\mathscr{P}_{\alpha_{2s}|a}) = 3 - 2a,\\
&\mathrm{gh}(\lambda_{\alpha_0}) = \mathrm{gh}(\pi^{\alpha_0})\\
&\qquad = \mathrm{gh}(\pi^{\alpha_{2s}}) = \mathrm{gh}(\lambda_{\alpha_{2s}}) = 0,\\
&\mathrm{ngh}(c^{\alpha_0|a}) = -\mathrm{ngh}(\mathscr{P}_{\alpha_0|a})\\
&\qquad = \mathrm{ngh}(c^{\alpha_{2s}|a}) = -\mathrm{ngh}(\mathscr{P}_{\alpha_{2s}|a}) = 1,\\
&\mathrm{ngh}(\pi^{\alpha_0}) = -\mathrm{ngh}(\lambda_{\alpha_0})\\
&\qquad = \mathrm{ngh}(\pi^{\alpha_{2s}}) = -\mathrm{ngh}(\lambda_{\alpha_{2s}}) = 2.
\end{aligned}
\tag{103}
$$

Then one should use the formalism that was considered in detail in Ref. 1.

[1] I. A. Batalin, P. M. Lavrov, and I. V. Tyutin, J. Math. Phys. **31**, 6 (1990).
[2] I. A. Batalin, P. M. Lavrov, and I. V. Tyutin, J. Math. Phys. **31**, 1487 (1990).
[3] I. A. Batalin and E. S. Fradkin, Phys. Lett. B **122**, 157 (1983).
[4] I. A. Batalin and E. S. Fradkin, Lett. Nuovo Cimento **38**, 393 (1983).
[5] P. M. Lavrov and I. V. Tyutin, Izv. Vuzov Fiz. **28**(7), 64 (1985).
[6] I. A. Batalin and G. A. Vilkovisky, Phys. Lett. B **120**, 166 (1983).
[7] I. A. Batalin and G. A. Vilkovisky, Phys. Rev. D **28**, 2567 (1983).
[8] P. M. Lavrov and I. V. Tyutin, Izv. Vuzov Fiz. **26**(1), 124 (1983).
[9] M. Henneaux and C. Teitelboim, Comm. Math. Phys. **115**, 213 (1988).
[10] J. Fisch, M. Henneaux, J. Stasheff, and C. Teitelboim, Comm. Math. Phys. **120**, 379 (1989).
[11] E. S. Fradkin, Proceedings of the X Winter School of Theoretical Physics in Karpacs, **207**, 93 (1973).
[12] E. S. Fradkin and G. A. Vilkovisky, Phys. Lett. B **55**, 224 (1975).
[13] I. A. Batalin and G. A. Vilkovisky, Phys. Lett. B **69**, 309 (1977).
[14] M. Henneaux, Phys. Rep. **126**, 2 (1985).

Solutions of the classical SU(2) Yang–Mills theory in (2+1) dimensions with the Chern–Simons term: Ansatz building

Rosy Teh[a)]
International Centre for Theoretical Physics, Trieste, Italy

(Received 6 October 1989; accepted for publication 25 April 1990)

A general way of writing the gauge potentials A^a_μ for which the SU(2) Yang–Mills equations of motion can be simplified and become solvable is shown. A number of exact solutions can be obtained from these simplified equations of motion.

I. INTRODUCTION

The Chern–Simons term was first proposed by Deser, Jackiw, and Templeton,[1] Schonfeld,[2] and Jackiw and Templeton.[3] The inclusion of the Chern–Simons term to the $(2+1)$-dimensional theory exhibit a number of interesting phenomena. With the presence of the Chern–Simons term, both electrically charged and neutral vortices acquire finite energy both in Abelian and non-Abelian gauge theories.[4–7] However, only numerical solutions are discussed and no exact solutions have yet been found. Also, $(2+1)$-dimensional theories are the high temperature limit of four-dimensional theories[8,9] and hence have physical applications at high temperatures.[10]

In this paper, I would like to report on a general way of writing the gauge potentials A^a_μ, so that the SU(2) Yang–Mills equation of motion can be simplified and made solvable. Some of the solutions of these simplified equations are also being discussed.

In Sec. II of this paper, I will briefly review the SU(2) Yang–Mills theory with the Chern–Simons term with the necessary notations. Section III will be divided into three parts. In this section I will show the general way of writing the ansatz after which I will discuss the ansatz in spherical coordinates in Sec. III A. Cylindrical coordinates will be dealt with in Sec. III B. Here, some new exact solutions both in the Minkowski and Euclidean space will be discussed. These solutions are complex, possess zero energy, and have finite action. Section III C will be on rectilinear coordinate and both Euclidean and Minkowski space solutions can be obtained from the simplified equations of motion. New exact Minkowski space solutions are also obtained. These complex solutions are wavelike in nature. I will end with some comments in Sec. IV.

II. NOTATIONS

The SU(2) Yang–Mills equations of motion with the Chern–Simons term are

$$\partial^\mu F^a_{\mu\nu} + \epsilon^{abc} A^\mu_b F_{\mu\nu c} + \tfrac{1}{2}\xi\epsilon_{\nu\mu\alpha} F^{\mu\alpha a} = 0, \tag{1a}$$

$$F^a_{\mu\nu} = \partial_\mu A^a_\nu - \partial_\nu A^a_\mu + \epsilon^{abc} A^b_\mu A^c_\nu. \tag{1b}$$

The Chern–Simons constant ξ is real in Minkowski space and it is replaced by $-i\xi$ in Euclidean space. The group indices $a,b,c = 1, 2$, and 3. The space indices $\mu, \nu, \alpha = 1, 2$, and 3 in Euclidean space and $\mu, \nu, \alpha = 0$, 1, and 2 in Minkowski space. The gauge coupling constant $g = 1$. The non-Abelian electric and magnetic field are given, respectively, by

$$E^a_i = F^a_{0i} \tag{2a}$$

and

$$B^a_i = -\tfrac{1}{2}\epsilon_{ij} F^{aij}; \quad i,j = 1,2. \tag{2b}$$

The energy-momentum tensor is written as

$$T^{\mu\nu} = F^{\mu\alpha a} F^\nu_{\alpha a} + g^{\mu\nu}\mathcal{L}_{\rm YM}, \tag{3}$$

where the Yang–Mills Lagrangian density is

$$\mathcal{L}_{\rm YM} = -\tfrac{1}{4} F^a_{\mu\nu} F^{\mu\nu}_a. \tag{4}$$

The Chern–Simons Lagrangian density is

$$\mathcal{L}_{\rm CS} = -\tfrac{1}{2}\xi\epsilon^{\mu\nu\alpha}(A^a_\mu \partial_\nu A^a_\alpha + \tfrac{1}{3}\epsilon^{abc} A^a_\mu A^b_\nu A^c_\alpha) \tag{5}$$

and the Yang–Mills action with the Chern–Simons term is

$$S = \int d^3x (\mathcal{L}_{\rm YM} + \mathcal{L}_{\rm CS}). \tag{6}$$

Hence the energy density T^{00} is given by

$$T^{00} = \tfrac{1}{2}(E^a_i E^a_i + B^a B^a). \tag{7}$$

III. THE GENERAL ANSATZ

By writing the Euclidean space gauge potentials as

$$A^a_\mu = (\hat{\alpha}^a \psi_1 + \hat{\gamma}^a \psi_2)\hat{\alpha}_\mu + (\hat{\alpha}^a A_1 + \hat{\gamma}^a \hat{A}_2)\hat{\gamma}_\mu + \hat{\beta}^a \hat{\beta}_\mu \chi, \tag{8}$$

where $\hat{\alpha}^a$, $\hat{\beta}^a$, and $\hat{\gamma}^a$ are the same as the orthogonal unit vectors $\hat{\alpha}_\mu$, $\hat{\beta}_\mu$, and $\hat{\gamma}_\mu$, the equations of motion (1) can be greatly simplified. In the Minkowski space, the gauge potentials take the form,

$$A^a_i = (\hat{\alpha}^a \psi_1 + \delta^a_3 \psi_2)\hat{\alpha}_i + \hat{\beta}^a \hat{\beta}_i \chi, \tag{9a}$$

$$A^a_0 = (\hat{\alpha}^a A_1 + \delta^a_3 A_2), \quad \hat{\alpha}^3 = \hat{\beta}^3 = 0. \tag{9b}$$

Here $\hat{\alpha}^a$, $\hat{\beta}^a$ and $\hat{\alpha}_i$, $\hat{\beta}_i$ are similar orthogonal unit vectors and $i = 1,2$.

One restriction on the ansatz (8) is that the corresponding variable of the unit vector $\hat{\beta}_\mu$ must have the dimension of length.

A. Spherical coordinate

In spherical coordinate, the Euclidean space gauge potentials take the form,

[a)] Permanent address: Institute of Technology MARA, P. O. Box 41, 01700 Kangar, Perlis, Malaysia.

 0022-2488/90/112718-04$03.00

$$A^a_\mu = (\hat{\phi}^a\psi_1 + \hat{\theta}^a\psi_2)\hat{\phi}_\mu + (\hat{\phi}^a A_1 + \hat{\theta}^a A_2)\hat{\theta}_\mu + \hat{r}^a\hat{r}_\mu\chi, \tag{10}$$

where

$$\hat{\phi}_\mu = \epsilon_{3\mu\nu}(x^\nu/\rho), \quad \rho = (x_1^2 + x_2^2)^{1/2}, \tag{11a}$$

$$\hat{\theta}_\mu = (x_3 x_i \delta^i_\mu - \rho^2\delta_{,\mu})/\rho r, \quad r = (x_1^2 + x_2^2 + x_3^2)^{1/2}, \tag{11b}$$

and

$$\hat{r}_\mu = x_\mu/r. \tag{11c}$$

By writing

$$\psi_1 = (1/r)\Phi_2(r), \quad \psi_2 = (1/r)\Phi_1(r), \tag{12a}$$

$$A_1 = -(1/r)(\Phi_1(r) - 1)), \quad A_2 = (1/r)\Phi_2(r), \tag{12b}$$

and

$$\chi = A(r), \tag{12c}$$

the Yang–Mills equations of motion (1) reduce to

$$(-\Phi_1' - A\Phi_2)' - A(\Phi_2' - A\Phi_1) - (1/r^2)\Phi_1(1 - \Phi_1^2 - \Phi_2^2) - i\xi(\Phi_2' - A\Phi_1) = 0, \tag{13a}$$

$$(-\Phi_2' + A\Phi_1)' + A(\Phi_1' + A\Phi_2) - (1/r^2)\Phi_2(1 - \Phi_1^2 - \Phi_2^2) - i\xi(-\Phi_1' - A\Phi_2) = 0, \tag{13b}$$

$$(\Phi_1\Phi_2' - \Phi_1'\Phi_2) - A(\Phi_1^2 + \Phi_2^2) + (i/2)\xi(1 - \Phi_1^2 - \Phi_2^2) = 0. \tag{13c}$$

Putting $\Phi_1 = \Phi$ and $\Phi_2 = 0$ (Ref. 11), Eqs. (13) reduce to

$$A = (-i/2)\xi(1 - 1/\Phi^2), \tag{14a}$$

$$\Phi'' + \frac{1}{4}\xi^2\frac{(1 - \Phi^4)}{\Phi^3} + \frac{\Phi}{r^2}(1 - \Phi^2) = 0. \tag{14b}$$

Here prime means d/dr.

D'Hoker and Vinet[11] had solved for Eqs. (14). However, only numerical solutions are being discussed. This numerical solution is complex. So far no exact solution has been found for the differential equations of (14).

B. Cylindrical coordinate

Here, the Euclidean space gauge potentials become

$$A^a_\mu = (\hat{\phi}^a\psi_1 + \delta^a_3\psi_2)\hat{\phi}_\mu + (\hat{\phi}^a A_1 + \delta^a_3 A_2)\delta_{\mu 3} + \hat{\rho}^a\hat{\rho}_\mu\chi, \tag{15}$$

where $\hat{\rho}_\mu = x_i\delta^i_\mu/\rho$ and ψ_1, ψ_2, A_1, and A_2 are functions of ρ and x_3. Here, χ is a function of ρ only. Substituting Eq. (15) into Eq. (1), the Yang–Mills equations of motion becomes, after setting,

$$\psi_1 = \psi, \quad A_1 = A, \tag{16a}$$

$$\psi_2 = i\psi + 1/\rho, \quad A_2 = iA, \tag{16b}$$

and replacing χ by $i\chi$,

$$\nabla^2 A - 2\chi A' - A\chi' - (1/\rho)\chi A + \chi^2 A + i\xi[\psi' + (1/\rho)\psi - \chi\psi] = 0, \tag{17a}$$

$$\nabla^2\psi + \partial_3^2\psi - (1/\rho^2)\psi - 2\chi\psi' - \psi\chi' - (1/\rho)\chi\psi + \chi^2\psi - i\xi(A' - \chi A) = 0, \tag{17b}$$

$$\partial_3(A' - \chi A) = -i\xi\partial_3\psi. \tag{17c}$$

Here, prime means $d/d\rho$ and ∇^2 is

$$\frac{1}{\rho}\frac{d}{d\rho}\left(\rho\frac{d}{d\rho}\right).$$

By letting the integration constant be zero, Eq. (17c) becomes

$$A' = \chi A - i\xi\psi. \tag{18}$$

Equation (18) solves Eq. (17a) exactly. To simplify Eq. (17b), I write

$$\psi(\rho, x_3) = e^{-\xi x_3}\Psi(\rho) \tag{19}$$

and Eq. (17b) reduced to

$$\Psi'' + \frac{1}{\rho}\Psi' - \frac{1}{\rho^2}\Psi - 2\chi\Psi' - \Psi\chi' - \frac{1}{\rho}\chi\Psi + \chi^2\Psi = 0. \tag{20}$$

As a result of Eq. (19), the function A becomes,

$$A(\rho, x_3) = e^{-\xi x_3}R(\rho), \tag{21}$$

and Eq. (18) then reduced to,

$$R' = \chi R - i\xi\Psi. \tag{22a}$$

At this point, it is clear that the function $\chi(\rho)$ is a completely arbitrary function and Eq. (20) can be solved in a linear fashion. One simple solution is

$$\Psi' = (\chi + 1/\rho)\Psi. \tag{22b}$$

Hence, together with Eq. (22a), Eq. (19), and Eq. (21), the explicit solution is

$$\psi = ic_2\rho\exp\left(\int\chi\,d\rho - \xi x_3\right), \tag{23a}$$

$$A = \left(c_3 + \frac{1}{2}\xi c_2\rho^2\right)\exp\left(\int\chi\,d\rho - \xi x_3\right). \tag{23b}$$

With the solutions of Eq. (23) the Yang–Mills action $S_0 = \int d^3x\mathcal{L}_{\rm YM}$ is zero. Hence, the action S is just the Cherns–Simons action,

$$S = S_0 + \int d^3x\mathcal{L}_{\rm CS} = \frac{-1}{2}\xi^2 c_2\int\exp\left(\int\chi\,d\rho - \xi x_3\right)d^3x. \tag{24}$$

When χ is just a negative nonzero constant, that is, $\chi = -b$, the solution of Eq. (23) has finite action,

$$S = (-c_2\xi/b^2)\pi, \tag{25}$$

when $x_3 > 0$. The solution has zero energy and momentum density.

In Minkowski space, the gauge potentials take the form

$$A^a_i = (\hat{\phi}^a\psi_1 + \delta^a_3\psi_2)\hat{\phi}_i + \hat{\rho}^a\hat{\rho}_i\chi, \tag{26a}$$

$$A^a_0 = \hat{\phi}^a A_1 + \delta^a_3 A_2. \tag{26b}$$

From Eq. (1) the simplified equations of motion can be solved in a similar manner and the solution obtained is

$$\psi = a_2\rho \sin(\xi x_0)\exp\left(\int \chi\, d\rho\right), \tag{27a}$$

$$A = (a_3 + \tfrac{1}{2}\xi a_2\rho^2)\sin(\xi x_0)\exp\left(\int \chi\, d\rho\right), \tag{27b}$$

where ψ and A are related to ψ_1, ψ_2, A_1, and A_2 by Eq. (16) and χ being replaced by $i\chi$.

A second cylindrical ansatz can also be written,

$$A^a_\mu = (\hat{\phi}^a\psi_1 + \hat{\rho}^a\psi_2)\hat{\phi}_\mu + (\hat{\phi}^a A_1 + \hat{\rho}^a A_2)\hat{\rho}_\mu + \delta^a_3\delta_{3\mu}\chi. \tag{28}$$

By writing,

$$A_1 = i\psi, \quad A_2 = -\psi, \tag{29a}$$

$$\psi_1 = \psi, \quad \psi_2 = i\psi, \tag{29b}$$

$$\chi = i\xi(n-1) \text{ or } i\xi n, \tag{29c}$$

n being a constant and solving the equations of motion (1), this Euclidean space solution is found to be

$$\psi = (a/\rho^2)e^{-n\xi x_3}, \quad a = \text{const.} \tag{30}$$

The Minkowski space version of this solution (30) can be similarly found by using the same substitution of Eqs. (29a) and (29b) and $\chi = -n\xi$ or $\xi(-n+i)$. Here, ψ is again given by Eq. (30).

C. Rectilinear coordinate

The rectilinear coordinate ansatz first appeared in Refs. 12 and 13. As in Ref. 12,

$$A^a_\mu = (\delta^a_2\psi_1 + \delta^a_3\psi_2)\delta_{\mu 2} + (\delta^a_2 A_1 + \delta^a_3 A_2)\delta_{\mu 3} + \delta^a_1\delta_{\mu 1}\chi, \tag{31}$$

and by setting,

$$\psi = \psi_2 = -\psi_3 = (i\sqrt{2}/4)\xi, \tag{32a}$$

$$A(x_2) = A_2 = A_3, \tag{32b}$$

$$\chi = \sqrt{2}\,A(x_2), \tag{32c}$$

the Euclidean space Yang–Mills equation of motion (1) is just

$$\partial_2^2 A(x_2) - \tfrac{1}{4}\xi^2 A(x_2) - 2A^3(x_2) = 0. \tag{33}$$

Equation (33) is just the differential equations of the Jacobi elliptic functions.[14] The detail analysis of the Minkowski as well as the Euclidean space solutions have already been done in Ref. 12.

Time-dependent Minkowski space solution can also be obtained when,

$$\psi_2 = A_2 = \psi(x_2 - x_0), \tag{34a}$$

$$\psi_3 = A_3 = -i\psi(x_2 - x_0), \tag{34b}$$

$$\chi = i\xi. \tag{34c}$$

The gauge potentials then become,

$$A^a_\mu = (\delta^a_2 - i\delta^a_3)(\delta^2_\mu + \delta^0_\mu)\psi(x_2 - x_0) + i\delta^a_1\delta^1_\mu\,\xi. \tag{35}$$

Here, ψ is a completely arbitrary function of $(x_2 - x_0)$. The electric and magnetic field are given by

$$E^a_i = \xi(\delta^a_2 - i\delta^a_3)\delta_{i1}\psi(x_2 - x_0), \tag{36a}$$

$$B^a = \xi(\delta^a_2 - i\delta^a_3)\psi(x_2 - x_0). \tag{36b}$$

The energy density, momentum density, and action all vanishes.

IV. COMMENTS

(1) The Bessel functions solutions of Ref. 12 can be obtained from the cylindrical coordinate ansatz of Eqs. (15) and (26) in the Euclidean and Minkowski space, respectively, when the functions $A_2 = \chi = 0$ and $\psi_2 = 1/\rho$.

(2) In arbitrary χ function solutions, the equations of motion (17) are nonlinear although they can be solved in a linear way. As a result, a number of explicit solutions can be obtained by using the ansatz of Eqs. (15) and (26), other than the solutions given by Eqs. (23) and (27) in the Euclidean and Minkowski space, respectively.

Examples of the other solutions are when Eq. (20) can be solved by writing,

$$\Psi'(\rho) = (\chi(\rho) - 1/\rho)\Psi(\rho) + i\xi m R(\rho). \tag{37}$$

When $m = 0$, the solution is given by

$$\psi = \frac{ic_1}{\rho}\exp\left(\int \chi\, d\rho - \xi x_3\right), \tag{38a}$$

$$A = (c_3 + \xi c_1 \ln\rho)\exp\left(\int \chi\, d\rho - \xi x_3\right). \tag{38b}$$

When $m = 1$, the solution becomes more complicated, as the solution now is

$$A = \frac{y}{(1+y^2)^{1/2}}\exp\left(\int\left(\chi - \frac{1}{\rho(1+y^2)}\right)d\rho - \sqrt{2}\,\xi x_3\right), \tag{39a}$$

$$\psi = \frac{1}{(1+y^2)^{1/2}}\exp\left(\int\left(\chi - \frac{1}{\rho(1+y^2)}\right)d\rho - \sqrt{2}\,\xi x_3\right) \tag{39b}$$

where

$$y = [aK_0(\xi\rho) + bI_0(\xi\rho)]/[aK_1(\xi\rho) + bI_1(\xi\rho)];$$

a,b = const and K_0, K_1, I_0, and I_1 are the modified Bessel functions. When m takes other values, different solutions will be obtained.

The Minkowski space version of solutions (38) and (39) can be obtained in a similar manner with some slight modification.

(3) In contrast to the ansatz of Eq. (15) for which many solutions can be obtained, the second cylindrical ansatz of Eq. (28) has limited solution.

(4) Since the electric and magnetic fields of the ansatz (15) and (16) point in the direction of the null vector $(\hat{\phi}^a + i\delta^a_3)$ in group space, the energy and momentum density have to be zero. Although the Yang–Mills action is zero, the total action can be a finite quantity by suitable choice of χ and for $x_3 > 0$. Hence, the solutions obtained do not correspond to any physical states. At most they may represent processes when x_3 runs from zero to infinity.

(5) Similarly, the electric and magnetic fields of the gauge potentials of Eq. (35) are in the direction of the null vector $(\delta_2^a - i\delta_3^a)$ in group space and the energy density, momentum density, and action are zero as a result of this group space null vector.

ACKNOWLEDGMENTS

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

She also acknowledges the Government of Japan for financial support.

[1]S. Deser, R. Jackiw, and S. Templeton, Phys. Rev. Lett. **48**, 975 (1982); Ann. Phys. **140**, 372 (1982).
[2]J. Schonfeld, Nucl. Phys. B **185**, 157 (1981).
[3]R. Jackiw and S. Templeton, Phys. Rev. D **23**, 2291 (1981).
[4]H. J. de Vega and F. A. Schaposnik, Phys. Rev. Lett. **56**, 2564 (1986); Phys. Rev. D **34**, 3206 (1986).
[5]S. K. Paul and A. Khare, Phys. Lett. B **191**, 389 (1987); Phys. Lett. B **174**, 420 (1986).
[6]G. Lozano, M. V. Manias, and F. A. Schaposnik, Phys. Rev. D **38**, 601 (1988).
[7]C. N. Kumar and A. Khare, Phys. Lett. B **178**, 395 (1986).
[8]L. Dolan and R. Jackiw, Phys. Rev. D **9**, 3320 (1974).
[9]S. Weinberg, Phys. Rev. D **9**, 3357 (1974).
[10]A. M. Polyakov, Mod. Phys. Lett. A **3**, 325 (1988).
[11]E. D'Hoker and L. Vinet, Ann. Phys. **162**, 413 (1985).
[12]R. Teh, J. Phys. G: Nucl. Part. Phys. **16**, 175 (1990).
[13]C. H. Oh, L. S. Sia, and R. Teh, Phys. Rev. D **40**, 601 (1989).
[14]I. S. Gradshteyn and I. W. Ryzhik, *Table of Integrals, Series and Products* (Academic, New York, 1980).

δ expansion for a quantum field theory in the nonperturbative regime

Carl M. Bender
Department of Physics, Washington University, St. Louis, Missouri 63130

Kimball A. Milton
Department of Physics and Astronomy, University of Oklahoma, Norman, Oklahoma 73019

Stephen S. Pinsky
Department of Physics, The Ohio State University, Columbus, Ohio 43210

L. M. Simmons, Jr.
Theoretical Division, Los Alamos National Laboratory, Los Alamos, New Mexico 87545 and The Santa Fe Institute, 1120 Canyon Road, Santa Fe, New Mexico 87501

(Received 6 February 1990; accepted for publication 5 July 1990)

The δ expansion, a recently proposed nonperturbative technique in quantum field theory, is used to calculate the dimensionless renormalized coupling constant of a $\lambda(\varphi^2)^{1+\delta}$ quantum field theory in d-dimensional space-time at the critical point defined by $\lambda \to \infty$ with the renormalized mass held fixed. The calculation is performed to leading order in δ and compared with previous lattice strong-coupling calculations. The numerical results are good and provide new evidence that the theory in four dimensions is free for all δ.

I. INTRODUCTION

It was our objective in introducing the δ expansion to devise a scheme that uses perturbative techniques to obtain nonperturbative information from a quantum field theory. The δ expansion relies on an artificial expansion parameter δ to organize the content of a quantum field theory in the form of a conventional perturbation series $\Sigma a_n \delta^n$. The parameter δ is introduced in a self-interacting scalar quantum field theory in the exponent of the self-interaction term: $\lambda(\varphi^2)^{\delta+1}$. Observe that when $\delta = 0$ the theory is free. As δ increases, the nonlinear interactions smoothly turn on. Thus the artificial parameter δ interpolates smoothly between a free theory and a fully nontrivial quantum field theory like $\lambda\varphi^4$.

In previous papers we have explained how to apply the δ expansion to a variety of physical problems. We have shown how to expand the Green's functions of a quantum field theory as series in powers of δ. The diagrammatic techniques are explained in Refs. 1–6. The δ expansion is particularly suited for quantum field theories having a global supersymmetry invariance.[7,8] The problem of renormalization has also been examined.[9–12] Moreover, δ-expansion methods have been used to solve ordinary differential equations describing classical physical systems.[13,14]

The purpose of this paper is to demonstrate clearly that the δ expansion probes the nonperturbative structure of a quantum field theory. We examine a $\lambda\varphi^4$ quantum field theory in d-dimensional space-time in the limit of an infinitely deep double well. Specifically, we consider a Lagrangian of the form

$$\mathscr{L} = \tfrac{1}{2}(\partial\varphi)^2 + \tfrac{1}{2}\mu^2\varphi^2 + \lambda M^2\varphi^2(M^{2-d}\varphi^2)^{\delta} \qquad (1)$$

in d-dimensional Euclidean space-time, where M is regarded as a fixed parameter having dimensions of mass and λ is the dimensionless coupling constant. We will compute the quantity G_∞ defined as the limit of the renormalized dimensionless coupling constant G_R as the unrenormalized coupling constant λ approaches infinity, subject to the constraint that the renormalized mass M_R be fixed. This limit is known as the Ising limit. It is a physically significant limit because it defines the critical point of the theory. The behavior of the theory at this critical point is universal and depends only on dimension d of space-time. Furthermore, the neighborhood of the critical point is a deeply nonperturbative region of the theory because holding M_R fixed as $\lambda \to \infty$ drives μ^2, the square of the unrenormalized mass, to $-\infty$. This limit of the theory cannot be obtained from a conventional weak-coupling series in powers of λ.

II. RENORMALIZED COUPLING

We define the renormalized quantities for the Lagrangian $\mathscr{L}$ in (1) in the following manner. From $\mathscr{L}$ we construct the vacuum persistence function $Z[J]$:

$$Z[J] \equiv \int D\varphi \exp\left\{-\int d^d x[\mathscr{L} + J(x)\varphi(x)]\right\}, \qquad (2)$$

where $J(x)$ is an external c-number source function. Next we define the n-point connected Green's functions by

$$W_n(x_1,\ldots,x_n) \equiv \frac{\delta}{\delta J(x_1)} \cdots \frac{\delta}{\delta J(x_n)} \ln Z[J]\Big|_{J=0}. \qquad (3)$$

We construct the renormalized mass from the Fourier transform of the two-point function $\tilde{W}_2(p^2)$:

$$M_R^2 \equiv Z_3 \tilde{W}_2^{-1}(p^2)|_{p^2=0}, \qquad (4)$$

where the wave-function renormalization constant Z_3 is given by

$$Z_3^{-1} \equiv \frac{d\tilde{W}_2^{-1}(p^2)}{dp^2}\Big|_{p^2=0}. \qquad (5)$$

These are the conventional definitions of the renormalization constants in the intermediate renormalization scheme. The dimensionless renormalized coupling constant G_R is constructed by amputating the legs from the four-point

 0022-2488/90/112722-04$03.00

function evaluated at zero momentum, multiplying by two powers of Z_3 to carry out the wave-function renormalization, multiplying by M_R^{d-4} to make the result dimensionless, and multiplying by -1 to give a positive quantity:

$$G_R \equiv -M_R^{d-4} Z_3^2 [\tilde{W}_2^{-1}(0)]^4 \tilde{W}_4(0,0,0). \tag{6}$$

To the order in which we will be working in this paper $Z_3 = 1$ and therefore this formula simplifies to

$$G_R = -M_R^{d-4} \Gamma_4, \tag{7}$$

where Γ_4 is the amputated four-point Green's function evaluated at zero momentum,

$$\Gamma_4 = M_R^8 \tilde{W}_4(0,0,0). \tag{8}$$

Let us make the dimension of space-time d explicit. It is interesting that one can calculate

$$G_\infty(d) = \lim_{\lambda \to \infty} G_R(d) \tag{9}$$

exactly in two simple cases: When the dimension of space-time d is 0, $G_\infty(0) = 2$, and when d is 1, $G_\infty(1) = 6$. We give the zero-dimensional argument here. The connected dimensionless amputated four-point function for fixed λ is given by

$$\frac{3[\int dx\, x^2 e^{-\mathscr{L}}]^2 - \int dx\, e^{-\mathscr{L}} \int dx\, x^4 e^{-\mathscr{L}}}{[\int dx\, x^2 e^{-\mathscr{L}}]^2}, \tag{10}$$

where

$$\mathscr{L} = (\mu^2/2)x^2 + \lambda(M^2 x^2)^{\delta+1}. \tag{11}$$

In the limit as $\lambda \to \infty$ the contribution to these integrals is localized at the saddle points $\pm x_0$, where x_0 satisfies

$$\mu^2 + 2\lambda M^2(1+\delta)(M^2 x_0^2)^\delta = 0. \tag{12}$$

Thus, as $\lambda \to \infty$, the complicated-looking expression in (10) simplifies to

$$G_\infty(0) = 2. \tag{13}$$

The calculation of $G_\infty(1)$ is slightly more complicated. The proof that $G_\infty(1) = 6$ is given in Ref. 15.

For dimensions other than 0 or 1, one can determine $G_\infty(d)$ numerically for all values of d between 0 and 4 using a strong-coupling lattice calculation. This calculation is described in Ref. 15. From a plot of $G_\infty(d)$ vs d given in this reference (see Fig. 1) one sees that $G_\infty(d)$ rises monotonically, taking on the approximate values $G_\infty(\frac{1}{2}) \simeq 3$, $G_\infty(\frac{3}{2}) \simeq 10$, $G_\infty(2) \simeq 15$, and $G_\infty(3) \simeq 30$. Note that $G_\infty(d)$ reaches a maximum near $d = 3.5$ and then decreases. The numerical calculation becomes less accurate as d increases; however, the results are consistent with the conclusion that $G_\infty(4) = 0$. Our goal in this paper is to show how one might reproduce these results using the δ expansion.

III. EXPANSION CALCULATION

Our starting point for the calculation is the δ expansion to second order of the renormalized mass M_R in d dimensions taken from Refs. 2 and 3:

$$M_R^2 = m^2 + \delta 2\lambda M^2 S + \lambda\delta^2 M^2 (S^2 - 1 + \psi(\tfrac{3}{2})) + \lambda^2\delta^2(4M^4/m^2)\Delta(0)(J_1 - J_2), \tag{14}$$

where

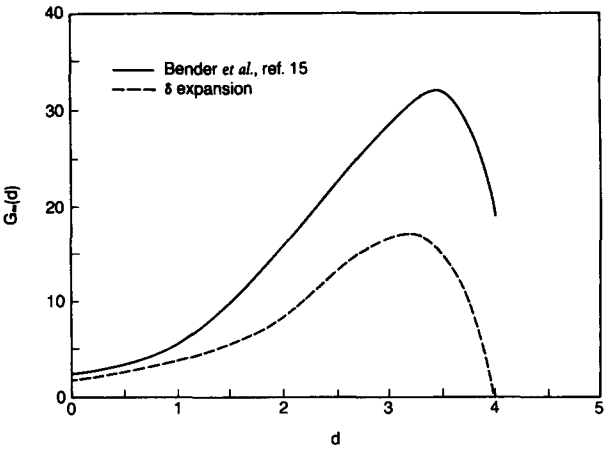


FIG. 1. The values of $G_\infty(d)$ found in Ref. 15 (solid line) compared with those found here (dashed line).

$$S = \psi(\tfrac{3}{2}) + 1 + \ln(2M^{2-d}\Delta(0)/m^{2-d}), \quad m^2 = \mu^2 + 2\lambda M^2, \tag{15a}$$

and

$$J_1 = \int d^d x \left(1 + \frac{1}{8}x^2\right)\left(\frac{\Delta(x)}{\Delta(0)}\right) \times \int_0^1 \frac{dt}{t}\sqrt{1-t}\,\ln\left(1 - t\left(\frac{\Delta(x)}{\Delta(0)}\right)^2\right), \tag{15b}$$

$$J_2 = \int d^d x \left\{ S\left(\frac{\Delta(x)}{\Delta(0)}\right)^2 + \int \frac{dt}{t^2}\sqrt{1-t} \times \left[\ln\left(1 - t\left(\frac{\Delta(x)}{\Delta(0)}\right)^2\right) + t\left(\frac{\Delta(x)}{\Delta(0)}\right)^2\right]\right\}. \tag{15c}$$

In (15b) we have included the effects of wavefunction renormalization. Here $\psi(\frac{3}{2}) = 2 - \gamma - 2\ln 2 = 0.036\,49$ is the digamma function evaluated at $\frac{3}{2}$ and $\Delta(x)$ is the dimensionless propagator in d-dimensional coordinate space. In terms of the dimensionless coordinate x,

$$\Delta(x) = (2\pi)^{-d/2}(x)^{1-d/2}K_{1-d/2}(x), \tag{16}$$

and, for $0 \leqslant d < 2$,

$$\Delta(0) = (4\pi)^{-d/2}\Gamma(1 - d/2). \tag{17}$$

When $d \geqslant 2$, $\Delta(0)$ is infinite, and (14) must be regulated, for example, by a momentum cutoff. The four-point function Γ_4 in second order in δ is given in Refs. 2 and 3:

$$\Gamma_4 = \frac{-4\lambda\delta M^2}{\Delta(0)(m^2)^{d-2}}\left[1 + \delta S + \frac{\lambda\delta M^2}{m^2}\Delta(0) \times (3I_1 + 8I_2 - 2I_3)\right], \tag{18}$$

where

$$I_1 = \int d^d x \int_0^1 dt \frac{1}{t\sqrt{1-t}} \ln\left(1 - t\left(\frac{\Delta(x)}{\Delta(0)}\right)^2\right), \tag{19}$$

$$I_2 = \int d^d x \left(\frac{\Delta(x)}{\Delta(0)}\right)^3 \int_0^1 dt \frac{\sqrt{1-t}}{1 - t(\Delta(x)/\Delta(0))^2}, \tag{20}$$

$$I_3 = \int d^d x \left[-S\left(\frac{\Delta(x)}{\Delta(0)}\right)^2 + \left(\frac{\Delta(x)}{\Delta(0)}\right)^4 \times \int_0^1 \frac{dt\sqrt{1-t}}{1 - t(\Delta(x)/\Delta(0))^2}\right]. \tag{21}$$

In preparation for taking the limit $\lambda \to \infty$, we rewrite (14) and (18) in the forms

$$M_R^2 = m^2\left[1 + \frac{2\lambda\delta M^2 S/m^2}{1 - \delta(S^2 - 1 + \psi(\frac{3}{2})/(2S) - 2\delta\lambda M^2\Delta(0)(J_1 - J_2)/(m^2 S)}\right], \tag{22}$$

$$\Gamma_4 = \frac{-4\lambda\delta M^2 m^{2-d}\Delta(0)^{-1}}{1 - \delta S - \delta\lambda M^2\Delta(0)(3I_1 + 8I_2 - 2I_3)/m^2}. \tag{23}$$

Forming the (1,1) Padé approximant in this manner is a standard procedure that is conventionally used to take the continuum (zero-lattice-spacing) limit of a lattice strong-coupling expansion.[16,17] We now take the $\lambda \to \infty$ limit of (22) and (23) and find, using (7), that

$$M_R^2 = m^2[1 - S^2/(\Delta(0)(J_1 - J_2))], \tag{24a}$$

and

$$G_\infty(d) = \frac{4(1 - S^2/[\Delta(0)(J_1 - J_2)])^{d/2-2}}{(\Delta(0))^2(-3I_1 - 8I_2 + 2I_3)}. \tag{24b}$$

Note that this result is independent of δ, which is consistent with the phenomenon of universality at the critical point.

We observe that the scale mass M is not determined. Of course, when $\delta = 1$ the Lagrangian $\mathscr{L}$ in (1) depends on the dimensional parameter $g = \lambda M^{4-d}$. However, to any finite order in δ, the δ expansion at $\delta = 1$ depends on M^2 as well as g. Therefore a condition is required that determines the value of M^2. Recently, Jones and Monoyios[18] applied the principle of minimal sensitivity originally enunciated by Stevenson[19] to the δ expansion at $\delta = 1$, and obtained good numerical results in low dimensions. The principle of minimal sensitivity determines M^2 by setting the derivative of the δ expansion with respect to M^2 equal to zero. Here, we find by explicit calculation that for $d = 0$ and 1, G_∞ is flat as a function of S in the vicinity of $S = 0$. Therefore we fix M^2 by the very simple condition $S = 0$. This is a very reasonable result, for then from (24a)

$$M_R^2 = m^2 = \mu^2 + 2\lambda M^2, \tag{25}$$

which forces $\mu^2 \to -\infty$, consistent with the double-well limit. For $d \geqslant 2$ the condition $S = 0$ is required to make M_R^2 finite as the momentum cutoff goes to infinity.

In zero dimensions, the integrals are rather easy to evaluate, and we find that

$$G_\infty(0) = 4/3\psi'(\tfrac{3}{2}) = 1.426\,33, \tag{26}$$

which agrees with the exact answer to 29%. In one dimension, the evaluation is a little more difficult:

$$G_\infty(1) = 16[-\tfrac{21}{2}\zeta(3) + 3\pi^2 \ln 2 + 64G - \pi^2/2 - 58]^{-1} = 4.458\,48, \tag{27}$$

where ζ is the Riemann zeta function and $G = 0.9159\,66...$is Catalan's constant. This result agrees with the exact answer to 26%. The calculation at $d = 2$ is performed by allowing d to approach 2 from below. In this limit the contributions from I_2 and I_3 vanish, and the contribution from I_1 comes from the first term in the expansion of the logarithm in the integrand. The result is

$$G_\infty(2) = 8\pi/3 = 8.377\,58, \tag{28}$$

which is about 40% low.

More generally, for $4 > d \geqslant 2$ this feature persists. In particular,

$$\Delta(0)^2 I_1 = -2\int d^d x \Delta(x)^2 = -\frac{4}{\Gamma(d/2)2^d\pi^{d/2}}\int_0^\infty dx\, x K_{d/2-1}(x)^2. \tag{29}$$

Evaluating (29) and substituting this into (24) we find that

$$G_\infty(d) = \tfrac{2}{3}\Gamma(d/2)2^d\pi^{d/2} \times [\sin((d/2 - 1)\pi)/((d/2 - 1)\pi)], \quad 4 > d \geqslant 2. \tag{30}$$

For $d = 2$ this reproduces our previous result, $8\pi/3$. For $d = 3$ we find $G_\infty(3) = 16\pi/3$, about 40% low. For $d = 4$ we see that $\Delta(0)^2 I_1$ diverges and therefore $G_\infty(4) = 0$. Thus we see that the δ expansion has provided additional evidence for the triviality in four dimensions of all theories of the class (1).

IV. CONCLUSION

These results are extremely gratifying and provide strong motivation for further calculations that should be done and that we shall try to do in the future. The next step is, of course, to improve the numerical results by performing a higher-order calculation of $G_\infty(d)$. This would require calculating M_R^2 and Γ_4 to at least third order in δ. This has never been done. Such calculations are extremely difficult except in zero-dimensional space-time.

ACKNOWLEDGMENTS

KAM thanks the Department of Physics, Ohio State University for its hospitality during calendar 1989, while this paper was written.

We thank the U. S. Department of Energy for partial financial support.

[1] C. M. Bender, K. A. Milton, M. Moshe, S. S. Pinsky, and L. M. Simmons, Jr., Phys. Rev. Lett. **58**, 2615 (1987).
[2] C. M. Bender, K. A. Milton, M. Moshe, S. S. Pinsky, and L. M. Simmons, Jr., Phys. Rev. D **37**, 1472 (1988).
[3] C. M. Bender and H. F. Jones, J. Math. Phys. **29**, 2659 (1988).
[4] S. S. Pinsky and L. M. Simmons, Jr., Phys. Rev. D **38**, 2518 (1988).
[5] N. Brown, Phys. Rev. D **38**, 723 (1988).
[6] H. T. Cho, K. A. Milton, S. S. Pinsky, and L. M. Simmons, Jr., J. Math. Phys. **30**, 2143 (1989).
[7] C. M. Bender, K. A. Milton, S. S. Pinsky, and L. M. Simmons, Jr., Phys. Lett. B **205**, 493 (1988).
[8] C. M. Bender and K. A. Milton, Phys. Rev. D **38**, 1310 (1988).
[9] C. M. Bender and H. F. Jones, Phys. Rev. D **38**, 2526 (1988).
[10] I. Yotsuyanage, Phys. Rev. D **39**, 485 (1989).
[11] J. Cline, S. S. Pinsky, and L. M. Simmons, Jr., Phys. Rev. D **39**, 3043 (1989).

[12] H. T. Cho, J. Cline, K. A. Milton, S. S. Pinsky, and L. M. Simmons, Jr., Nucl. Phys. B **329**, 574 (1990).
[13] S. M. Wandzura, J. Opt. Soc. Am. **71**, 321 (1981). The authors wish to thank S. Wandzura for bringing this article to their attention.
[14] C. M. Bender, K. A. Milton, S. S. Pinsky, and L. M. Simmons, Jr., J. Math. Phys. **30**, 1447 (1989).
[15] C. M. Bender, F. Cooper, G. S. Guralnik, H. Moreno, R. Roskies, and D. H. Sharp, Phys. Rev. Lett. **45**, 501 (1980).
[16] See Ref. 15 and references therein.
[17] It is also apparent from earlier studies of the δ expansion [for example, Refs. 2 and 6] that Padé approximants are effective in summing the δ expansion.
[18] H. F. Jones and M. Monoyios, Int. J. Mod. Phys. A **4**, 1735 (1989).
[19] P. M. Stevenson, Phys. Rev. D **23**, 2916 (1981).

Characteristic functions and invariants of supermatrices

Yuji Kobayashi
Department of Mathematics, Tokushima University, Tokushima 770, Japan
Shigeaki Nagamachi
Technical College, Tokushima University, Tokushima 770, Japan

(Received 27 July 1989; accepted for publication 16 May 1990)

The characteristic functions and the invariants of supermatrices are studied. It is shown that the Euclidean algorithm is useful in obtaining a system of invariants.

I. INTRODUCTION

The theories of linear algebra and analysis over a Grassmann algebra have been developed and are a base of the theory of supermanifolds, Lie supergroups, and Lie superalgebras,[1-5] which are extensively used in modern physics.[6,7] Especially, representation of supergroups and superalgebras plays an essential role in supersymmetric field theory. As a foundation of the representation theory, we study the eigenvalue problem of supermatrices. In our previous paper,[8] we introduced the notions of the eigenvalue and the eigenvector of supermatrices and studied their basic properties. In this paper, we study the characteristic functions and the invariants of supermatrices. The invariants of some supermatrices have found an application in statistical mechanics.[9]

In Sec. II, we recall the results of Ref. 8. In Sec. III, we define the characteristic function $h_M(X)$ of a supermatrix M, a super equivalent of the characteristic polynomial of an ordinary matrix. The function $h_M(X)$ is a rational function in X and its zeros and poles are the eigenvalues of M of the first kind and of the second kind, respectively. Therefore, if we can get an irreducible expression $f(X)/g(X)$ of $h_M(X)$ then the coefficients of $f(X)$ and $g(X)$ are the fundamental symmetric functions of the eigenvalues of the first kind and the second kind, respectively, and they are all invariant under the transformation $M \to UMU^{-1}$ by invertible supermatrix U. In Sec. IV, we give an efficient method to compute the invariants by employing the Euclidean algorithm.

II. PRELIMINARIES

We recall some basic definitions and results given in Kobayashi and Nagamachi.[8] Let $\Lambda = \Lambda_0 \oplus \Lambda_1$ be a Grassmann algebra over the complex numbers $\mathbb{C}$, where Λ_0 (resp. Λ_1) is the even (resp. odd) part of Λ. As a subalgebra, $\mathbb{C}$ is naturally contained in Λ and is called the body of Λ. Any element $a \in \Lambda$ is a sum of the body $\tilde{a} \in \mathbb{C}$ and the nilpotent element $s(a)$ called the soul. Then the mapping $\sim$ is a homomorphism of Λ to $\mathbb{C}$.

Let p and q be non-negative integers and put $n = p + q$. A (p,q) supermatrix M is an $n \times n$ matrix over Λ such that $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$, where A (resp. D) is a $p \times p$ matrix (resp. $q \times q$ matrix) over Λ_0 and B (resp. C) is a $p \times q$ matrix (resp. $q \times p$ matrix) over Λ_1. The body $\tilde{M}$ of $M = (m_{ij})$ is the matrix $(\tilde{m}_{ij})$ over $\mathbb{C}$. An even (resp. odd) vector is a column $x = (x_1,...,x_p,x_{p+1},...,x_n)^T$, where x_i is in Λ_0 (resp. Λ_1) for $i = 1,...,p$ and in Λ_1 (resp. Λ_0) for $i = p+1,...,n$. The body $\tilde{x}$ of x is the vector $(\tilde{x}_1,...,\tilde{x}_p,\tilde{x}_{p+1},...,\tilde{x}_n)^T$.

If $Mx = \lambda x$ for $\lambda \in \Lambda_0$ and x is a vector such that $\tilde{x} \neq 0$, λ is called an eigenvalue of M and x the eigenvector corresponding to λ. If x is even (resp. odd) we say λ is an eigenvalue of the first (resp. second) kind.

The main result of Ref. 8 is the following theorem, which is an elaboration of a theorem by Berezin.[2]

Theorem 2.1: Let $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ be a (p,q) supermatrix such that the eigenvalues $\alpha_1,...,\alpha_p$ of $\tilde{A}$ and the eigenvalues $\delta_1,...,\delta_q$ of $\tilde{D}$ are all different. Then M has eigenvalues $\beta_1,...,\beta_p$ and $\gamma_1,...,\gamma_q$ such that $\tilde{\beta}_1 = \alpha_1,...,\tilde{\beta}_p = \alpha_p$ and $\tilde{\gamma}_1 = \delta_1,...,\tilde{\gamma}_q = \delta_q$. Moreover, the eigenvalues $\beta_1,...,\beta_p$ (resp. $\gamma_1,...,\gamma_q$) are of the first (resp. second) kind, and there exists an invertible supermatrix U such that $UMU^{-1} = \mathrm{diag}(\beta_1,...,\beta_p,\gamma_1,...,\gamma_q)$.

III. CHARACTERISTIC FUNCTIONS

In this section we consider the super equivalent of the characteristic polynomial. In the super case it is not a polynomial but is expressed as a ratio of polynomials. Because our ring $\Lambda_0[X]$ of polynomials is not an integral domain, the ratio is an element of the quotient ring, the localization of $\Lambda_0[X]$ at the minimal prime ideal. For a polynomial

$$f(X) = a_0 X^n + a_1 X^{n-1} + \cdots + a_n \in \Lambda_0[X],$$

the body $\tilde{f}$ is defined by

$$\tilde{f}(X) = \tilde{a}_0 X^n + \tilde{a}_1 X^{n-1} + \cdots + \tilde{a}_n.$$

Proposition 3.1: The set $\mathcal{P} = s(\Lambda_0)[X]$ is the smallest prime ideal of $\Lambda_0[X]$. It is the set of nilpotent elements of $\Lambda_0[X]$ and its complement $\Lambda_0[X] - \mathcal{P}$ is the set of non-zero divizers of $\Lambda_0[X]$.

Proof: The proof is easy.

Let

$$\Lambda(X) = \Lambda[X]_{\mathcal{P}} = \{f/g \mid f \in \Lambda[X], g \in \Lambda_0[X], g \notin \mathcal{P}\}$$

be the ring of quotients of $\Lambda[X]$. The addition and the multiplication of $\Lambda(X)$ are defined in a usual way, and $f_1/g_1 = f_2/g_2$ in $\Lambda(X)$ if and only if $f_1 g_2 = g_1 f_2$ in $\Lambda[X]$. A polynomial $f \in \Lambda[X]$ is invertible if $\tilde{f} \neq 0$, because $f - \tilde{f}$ is nilpotent. The even part

$$\Lambda_0(X) = \Lambda_0[X]_{\mathcal{P}} = \{f/g \mid f, g \in \Lambda_0[X], g \notin \mathcal{P}\}$$

 0022-2488/90/112726-05$03.00

of the Z_2-graded algebra $\Lambda(X)$ is the localization of $\Lambda_0[X]$ at $\mathscr{P}$. Let $h(X)\in\Lambda_0(X)$ and $\alpha\in\Lambda_0$. If there is an expression $h(X)=f(X)/g(X)$ such that $g(\alpha)^{\sim}\neq 0$, then the value $h(\alpha)$ of h at α is defined to be $f(\alpha)/g(\alpha)$, otherwise the value $h(\alpha)$ is not defined. The value $h(\alpha)$ does not depend on the expression of $h(X)$. In particular, we say that $\alpha\in\Lambda_0$ is a zero of h if $h(\alpha)=0$. When h is invertible, a zero of h^{-1} is called a pole of h.

For a supermatrix M, the superdeterminant $\mathrm{sdet}(XE-M)$ of $XE-M$, which is an element of $\Lambda_0(X)$, is called the characteristic (rational) function of M and is denoted by $h_M(X)$. We call M *separable* if it satisfies the assumption of Theorem 2.1.

Theorem 3.2: Let M be a separable supermatrix. Then we have

$$h_M(X)=(X-\beta_1)\cdots(X-\beta_p)/(X-\gamma_1)\cdots(X-\gamma_q),$$

where $\beta_1,\ldots,\beta_p$ (resp. $\gamma_1,\ldots,\gamma_q$) are the eigenvalues of M of the first (resp. second) kind.

Proof: From the multiplicative property of superdeterminants, we have

$$\mathrm{sdet}(XE-M)=\mathrm{sdet}(XE-UMU^{-1}),$$

for an invertible supermatrix U. Hence, by Theorem 2.1 we may assume M is a diagonal matrix, and the assertion is clear.

Let $M=\left(\begin{smallmatrix}A & B\\ C & D\end{smallmatrix}\right)$. Then

$$\begin{aligned} h_M(X)&=\frac{\det(XE-A-B(XE-D)^{-1}C)}{\det(XE-D)}\\ &=\frac{\det((\det(XE-D))(XE-A)-B(\det(XE-D))(XE-D)^{-1}C)}{\det(XE-D)^{p+1}} \qquad (3.1)\\ &=\frac{\det(XE-A)}{\det(XE-D-C(XE-A)^{-1}B)}\\ &=\frac{\det(XE-A)^{q+1}}{\det((\det(XE-A))(XE-D)-C(\det(XE-A))(XE-A)^{-1}B)}. \qquad (3.2)\end{aligned}$$

Suppose M is separable. The eigenvalues of M of the first kind are obtained as the zeros of the numerator,

$$\det((\det(XE-D))(XE-A)-B(\det(XE-D))\times(XE-D)^{-1}C),$$

of (3.1) obtained by the Newton method starting with the zeros of $\det(XE-A)^{\sim}$. On the other hand, the eigenvalues of the second kind are obtained as the zeros of the denominator

$$\det((\det(XE-A))(XE-D)-C(\det(XE-A))\times(XE-A)^{-1}B)$$

of (3.2) obtained from the zeros of $\det(XE-D)^{\sim}$.

We say f and g in $\Lambda_0[X]$ are coprime and write as $(f,g)=1$, if the ideal generated by f and g is the whole ring $\Lambda_0[X]$. For $h\in\Lambda_0(X)$ an expression $h=f/g$ is called irreducible, if $(f,g)=1$. The expressions (3.1) and (3.2) are not irreducible. In the sequence of this section we will give the irreducible expression of the characteristic function h_M (Theorem 3.9).

Lemma 3.3: Let f and g be in $\Lambda_0[X]$. Then f and g are coprime in $\Lambda_0[X]$ if and only if $\tilde{f}$ and $\tilde{g}$ are coprime in $\mathbb{C}[X]$.

Proof: Suppose $\tilde{f}$ and $\tilde{g}$ are coprime and let p and q be in $\mathbb{C}[X]$ such that $p\tilde{f}+q\tilde{g}=1$. Then $pf+qg=1+r$ with $r\in\mathscr{P}$. Hence,

$$(p/(1+r))f+(q/(1+r))g=1,$$

where $p/(1+r)=p(1-r+r^2-\cdots)$ and $q/(1+r)$ are in $\Lambda_0[X]$, implying $(f,g)=1$. The converse is clear.

Lemma 3.4: Let f,g,f_1, and g_1 be in $\Lambda_0[X]$ and suppose $\tilde{f}=\tilde{f}_1,\tilde{g}=\tilde{g}_1$ and $(f,g)=1$. If $fg=f_1g_1$ then $f_1=(1+r)f$, $g_1=(1/(1+r))g$ with $r\in\mathscr{P}$. If moreover, f and f_1 are monic then $f=f_1$ and $g=g_1$.

Proof: Since $(f,g_1)=1$ by Lemma 3.3, we have $pf+qg_1=1$ for $p,q\in\Lambda_0[X]$. Hence,

$$f_1=pff_1+qg_1f_1=(pf_1+qg)f,$$

where $(pf_1+qg)^{\sim}=1$. Thus,

$$f_1=(1+r)f \text{ and } g_1=(1/(1+r))g$$

with

$$r=pf_1+qg-1\in\mathscr{P}.$$

Here, if f and f_1 are monic, then $r=0$ and consequently $f=f_1$ and $g=g_1$.

Corollary 3.5: If $h\in\Lambda_0(X)$ has an irreducible expression f/g with g monic, then this expression is unique.

The following is a key lemma for our discussion.

Lemma 3.6 (Hensel's lemma)*:* Let f and g be in $\Lambda_0[X]$ such that $(f,g)=1$ and let r be in $\mathscr{P}$. Then there exist r_1 and r_2 in $\mathscr{P}$ such that $fg+r=(f+r_1)(g+r_2)$.

Proof: We proceed by induction on the nilpotency of r. Let p and q be such that $pf+qg=1$. Then

$$fg+r=(f+rq)(g+rp)-r^2pq.$$

Since $(f+rq,g+rp)=1$ by Lemma 3.3 and the nilpotency of r^2pq is smaller than that of r, we have r_1' and r_2' in $\mathscr{P}$ satisfying

$$fg+r=(f+rq+r_1')(g+rp+r_2')$$

by the induction hypothesis.

Lemma 3.7: Let f be a monic polynomial in $\Lambda_0[X]$ and r be in $\mathscr{P}$. Then there is a monic polynomial $g\in\Lambda_0[X]$ and $t\in\mathscr{P}$ such that $f+r=g(1+t)$.

Proof: Let I be the ideal of Λ_0 generated by the coefficients of r. Since I is a nilpotent ideal we proceed by induction on the nilpotency m of I. By the division algorithm $r = qf + r_1$, q, $r_1 \in I[X]$ with $\deg(r_1) < \deg(f)$. Therefore, we have $f + r = (f + r_1 + r_2)(1 + q)$, where $r_2 = -r_1 q(1+q)^{-1}$. Here, $f + r_1$ is monic and the coefficients of $r_2 = -r_1 q(1 - q + q^2 - \cdots)$ are in I^2 and the nilpotency of the ideal generated by the coefficients of r_2 is smaller than m. By induction hypothesis we have $f + r_1 + r_2 = g(1 + t')$ with monic g and $t' \in \mathscr{P}$. Then letting $t = t' + q + t'q$, we have $f + r = g(1 + t)$.

By this lemma we get the following elaboration of Lemma 3.6.

Corollary 3.8: Let f and g be monic polynomials in $\Lambda_0[X]$ such that $(f,g) = 1$ and let r be in $\mathscr{P}$ such that $\deg(r) < \deg(fg)$. Then there exist r_1 and r_2 in $\mathscr{P}$ such that $fg + r = (f + r_1)(g + r_2)$, $\deg(r_1) < \deg(f)$, and $\deg(r_2) < \deg(g)$.

Theorem 3.9: Let $M = \begin{pmatrix} A & B \\ C & D \end{pmatrix}$ a supermatrix. Let f and g be the characteristic polynomials of A and D, respectively. Suppose $(f,g) = 1$, then the characteristic function h_M of M has a unique irreducible expression:

$$h_M = (f + r)/(g + t),$$

where r, $t \in \mathscr{P}$, $\deg(r) < \deg(f)$, and $\deg(t) < \deg(g)$.

Proof: The numerator of (3.1) is equal to

$$\det(g(X)(XE - A) - Bg(X)(XE - D)^{-1}C) = g^p f + u,$$

where $u \in \mathscr{P}$ and $\deg(u) < \deg(g^p f) = p(q + 1)$. Thus we have $h_M = (g^p f + u)/g^{p+1}$. Similarly, we have $h_M = f^{q+1}/(f^q g + v)$, where $v \in \mathscr{P}$ and $\deg(v) < q(p + 1)$. By Corollary 3.8,

$$h_M = (g^p + t')(f + r)/g^{p+1} = f^{q+1}/(f^q + r')(g + t),$$

where t', r, r', $t \in \mathscr{P}$, $\deg(t') < pq$, $\deg(r) < p$, $\deg(r') < pq$, and $\deg(t) < q$. Thus

$$f^{q+1} g^{p+1} = (f + r)(f^q + r')(g + t)(g^p + t').$$

By Lemma 3.4 we find $f^{q+1} = (f + r)(f^q + r')$ and $g^{p+1} = (g + t)(g^q + t')$, and consequently $h_M = (f + r)/(g + t)$.

IV. INVARIANTS OF SUPERMATRICES

In this section we suppose that Λ is the Grassmann algebra generated by an infinite number of generators v_1, v_2, ..., v_n, ... over $\mathbb{C}$. We consider the superspace $\mathscr{M}$ consisting of (p,q) supermatrices over Λ:

$$\begin{aligned}\mathscr{M} &= \Lambda_0^{p^2+q^2} \oplus \Lambda_1^{2pq} \\ &= \{(x_{ij}, y_{kl}, \xi_{il}, \eta_{kj}) \mid x_{ij}, y_{kl} \in \Lambda_0, \xi_{il}, \\ &\qquad \eta_{kj} \in \Lambda_1, 1 \leqslant i,j \leqslant p, 1 \leqslant k,l \leqslant q\}.\end{aligned}$$

Let $\Lambda[x,y,\xi,\eta]$ denote the ring of polynomials in $p^2 + q^2$ commutative indeterminates x_{ij}, y_{kl} and $2pq$ anticommutative indeterminates ξ_{il}, η_{kj} over Λ. A polynomial in $\Lambda[x,y,\xi,\eta]$ defines a polynomials function on $\mathscr{M}$ valued in Λ, and different polynomials define different functions because Λ is infinite dimensional. In this sense we identify polynomials in $\Lambda[x,y,\xi,\eta]$ and polynomial functions of $\mathscr{M}$. For a polynomial $\phi(x,y,\xi,\eta) \in \Lambda[x,y,\xi,\eta]$ the body $\tilde{\phi} \in \mathbb{C}[x,y]$ is the polynomial obtained from $\phi(x,y,0,0)$ by replacing all its coefficients by their bodies. Let

$$\begin{aligned}\Lambda(x,y,\xi,\eta) = \{&\phi/\psi \mid \phi \in \Lambda[x,y,\xi,\eta], \psi \in \Lambda_0[x,y], \\ &\psi \notin s(\Lambda_0)[x,y]\}\}\end{aligned}$$

be the ring of quotients of $\Lambda[x,y,\xi,\eta]$. Then $\Lambda(x,y,\xi,\eta)$ is canonically isomorphic to the Grassman algebra $\mathbb{C}(x,y)[v,\xi,\eta]$ generated by

$$v_1, v_2, \ldots, v_n, \ldots, \xi_{11}, \xi_{12}, \ldots, \xi_{pq}, \eta_{11}, \eta_{12}, \ldots, \eta_{qp}$$

over the rational function field

$$\mathbb{C}(x,y) = \mathbb{C}(x_{11}, x_{12}, \ldots, x_{pp}, y_{11}, y_{12}, \ldots, y_{qq}).$$

For $(x^0,y^0,\xi^0,\eta^0) \in \mathscr{M}$, if there is an expression $\Phi = \phi/\psi$ such that $\tilde{\psi}(x^0,y^0) = \psi(x^0,y^0)^{\sim} \neq 0$, then the value $\Phi(x^0,y^0,\xi^0,\eta^0)$ of Φ at (x^0,y^0,ξ^0,η^0) is defined to be $\phi(x^0,y^0,\xi^0,\eta^0)/\psi(x^0,y^0)$, otherwise $\Phi(x^0,y^0,\xi^0,\eta^0)$ is not defined. In this way we identify elements of $\Lambda(x,y,\xi,\eta)$ and rational functions on $\mathscr{M}$ valued in Λ.

A rational function Φ on $\mathscr{M}$ is called *invariant*, if for any $M \in \mathscr{M}$ such that $\Phi(M)$ is defined and for any invertible supermatrix U, $\Phi(UMU^{-1})$ is also defined and $\Phi(M) = \Phi(UMU^{-1})$ holds (see Chap. 1.5 of Ref. 10 for the invariants of ordinary matrices).

Let $M = \begin{pmatrix} x & \xi \\ \eta & y \end{pmatrix}$ be the matrix whose elements are the indeterminates x_{ij}, y_{kl}, ξ_{il}, η_{kj}. Then M is a (p,q) supermatrix over $\Lambda(x,y,\xi,\eta) = \mathbb{C}(x,y)[v,\xi,\eta]$. The discussion in the proofs of the lemmas in the preceding section is valid, even if we replace the field $\mathbb{C}$ by $\mathbb{C}(x,y)$ and the Grassmann algebra Λ by $\mathbb{C}(x,y)[v,\xi,\eta]$. Especially Theorem 3.9 is valid for supermatrices over $\Lambda(x,y,\xi,\eta)$. Since our supermatrix M satisfies the condition of Theorem 3.9, the characteristic function h_M of M has a unique irreducible expression $\bar{f}/\bar{g}$ with monic polynomials $\bar{f}$, $\bar{g} \in \Lambda(x,y,\xi,\eta)_0[X]$. Suppose $\bar{f} = X^p + \sigma_1 X^{p-1} + \cdots + \sigma_p$ and $\bar{g} = X^q + \sigma_{p+1} X^{q-1} + \cdots + \sigma_{p+q}$. These coefficients σ_i are elements of $\Lambda(x,y,\xi,\eta)_0$, rational functions on $\mathscr{M}$ of even grade. Let U be an invertible supermatrix over Λ. By the multiplicative property of superdeterminants we have

$$\begin{aligned}h_M &= \operatorname{sdet}(XE - M) \\ &= \operatorname{sdet}(UXE - M)U^{-1}) = h_{UMU^{-1}}.\end{aligned}$$

Therefore, all the coefficients $\sigma_1, \ldots, \sigma_{p+q}$ are invariant.

Theorem 4.1: A rational function Φ on $\mathscr{M}$ is invariant if and only if Φ is a rational function of $\sigma_1, \ldots, \sigma_{p+q}$.

Proof: Let $\Phi \in \Lambda(x,y,\xi,\eta)$ be an invariant rational function and $M \in \mathscr{M}$ be a separable supermatrix such that $\Phi(M)$ is defined. Then M is diagonalized to $\operatorname{diag}(\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)$ and we have

$$\begin{aligned}\Phi(M) &= \Phi(\operatorname{diag}(\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q)) \\ &= \Psi(\alpha_1, \ldots, \alpha_p, \beta_1, \ldots, \beta_q). \qquad (4.1)\end{aligned}$$

Here,

$$\Psi(z,w) = \Phi(\operatorname{diag}(z_1, \ldots, z_p, w_1, \ldots, w_q))$$

is a rational function in $z_1, \ldots, z_p$ and $w_1, \ldots, w_q$ over Λ. It follows from the invariance of Φ that for any point $(a,b) \in \mathbb{C}^{p+q}$ at which Ψ is defined and for any permutations π and τ

$$\Psi(a_1,...,a_p,b_1,...,b_q)=\Psi(a_{\pi(1)},...,a_{\pi(p)},b_{\tau(1)},...,b_{\tau(q)}), \tag{4.2}$$

holds. Since Ψ is an element of $\Lambda(z,w)=\mathbb{C}(z,w)[v]$, Ψ is expanded as

$$\Psi(z,w)=\Sigma_{i_1<\cdots<i_n}\Psi_{i_1\cdots i_n}(z,w)v_{i_1}\cdots v_{i_n},$$

with $\Psi_{i_1\cdots i_n}\in\mathbb{C}(z,w)$. Here, $\Psi_{i_1\cdots i_n}(z,w)$ are invariant under the permutations of $z_1,...,z_p$ and the permutations of $w_1,...,w_q$ by (4.2). Therefore, Ψ is a rational function $\bar{\Psi}$ of fundamental symmetric functions $\bar{\sigma}_1,...,\bar{\sigma}_p$, of $z_1,...,z_p$ and $\bar{\sigma}_{p+1},...,\bar{\sigma}_{p+q}$ of $w_1,...,w_q$ by the fundamental theorem of symmetric functions. By (4.1) we have

$$\Phi(M)=\Psi(\alpha,\beta)=\bar{\Psi}(\bar{\sigma}_1(\alpha,\beta),...,\bar{\sigma}_{p+q}(\alpha,\beta)).$$

Since $\bar{\sigma}_i(\alpha,\beta)=\sigma_i(M)$ by Theorem 3.2, we have

$$\Phi(M)=\bar{\Psi}(\sigma_1(M),...,\sigma_{p+q}(M)). \tag{4.3}$$

Because (4.3) holds for any separable supermatrix M at which Φ is defined, Φ is equal to $\bar{\Psi}(\sigma_1,...,\sigma_{p+q})$ as an element of $\Lambda(x,y,\xi,\eta)$.

The converse is obvious because $\sigma_1,...,\sigma_{p+q}$ are invariant.

The invariants $\sigma_1,...,\sigma_{p+q}$ are rather complicated rational functions on $\mathcal{M}$. We are going to find better invariants.

Let $\mathcal{K}=\Lambda(x,y,\xi,\eta)_0$ denote the ring of rational functions on $\mathcal{M}$ of even grade. As we have seen above, $\bar{f}$ and $\bar{g}$ are monic polynomials over $\mathcal{K}$.

Suppose $\deg(\bar{f})\geqslant\deg(\bar{g})$, then the Euclidean algorithm in $\mathcal{K}[X]$ goes as follows:

$$\begin{aligned}
&\bar{f}=t_0\bar{g}+\bar{r}_1, \quad \deg(\bar{r}_1)=\deg(\bar{g})-1=q-1,\\
&\bar{g}=t_1\bar{r}_1+\bar{r}_2, \quad \deg(\bar{r}_2)=\deg(\bar{r}_1)-1,\\
&\bar{r}_1=t_2\bar{r}_2+\bar{r}_3, \quad \deg(\bar{r}_3)=\deg(\bar{r}_2)-1,\\
&\cdots\\
&\bar{r}_{q-2}=t_{q-1}\bar{r}_{q-1}+\bar{r}_q, \quad \deg(\bar{r}_q)=0,\\
&\bar{r}_{q-1}=t_q\bar{r}_q,
\end{aligned} \tag{4.4}$$

where $\bar{r}_i$ is the r_i divided by its leading coefficient. Note that $\deg(t_i)=1$ for $i=1,...,q$, and $\bar{r}_q=1$ because r_q is a constant. Now we define $\delta_1,...,\delta_{p+q}$ as follows: For $1\leqslant i\leqslant p-q$, δ_i is the coefficient t_0 of degree $\deg(t_0)-i=p-q-i$, and for $j=1,...,q$, $\delta_{p-q+2j-1}$ is the leading coefficient of r_j, and δ_{p-q+2j} is the coefficient of t_j of degree 0. These are rational functions of $\sigma_1,...,\sigma_{p+q}$ and so they are invariants. In particular, $-\delta_1$ is the well-known supertrace of M.

Let $F=g^pf+u$ be the numerator and $G=g^{p+1}$ be the denominator of (3.1) [or $F=f^{q+1}$ be the numerator and $G=f^qg+v$ be the denominator of (3.2)], where f and g are the characteristic polynomials of A and D, respectively. Here, F and G are also considered to be elements of $\mathcal{K}[X]$. The proof of Theorem 3.9 shows the existence of monic polynomial H such that $F=H\cdot\bar{f}$ and $G=H\cdot\bar{g}$. Multiplying both sides of the equations in (4.4) by H, we have

$$\begin{aligned}
&F=t_0G+H\bar{r}_1,\\
&\deg(H\bar{r}_1)=\deg(G)-1=(p+1)q-1,\\
&G=t_1H\bar{r}_1+H\bar{r}_2, \quad \deg(H\bar{r}_2)=\deg(H\bar{r}_1)-1,\\
&H\bar{r}_1=t_2H\bar{r}_2+H\bar{r}_3, \quad \deg(H\bar{r}_3)=\deg(H\bar{r}_2)-1,\\
&\cdots\\
&H\bar{r}_{q-2}=t_{q-1}H\bar{r}_{q-1}+Hr_q, \quad \deg(Hr_q)=pq,\\
&H\bar{r}_{q-1}=t_qH\bar{r}_q.
\end{aligned}$$

This turns out to be the Euclidean algorithm starting with F and G, and we have the following theorem.

Theorem 4.2: Let F and G be as above, and apply the Euclidean algorithm to F and G:

$$\begin{aligned}
&F=T_0G+R_1,\\
&\deg(R_1)=\deg(G)-1=(p+1)q-1,\\
&G=T_1\bar{R}_1+R_2, \quad \deg(R_2)=\deg(R_1)-1,\\
&\bar{R}_1=T_2\bar{R}_2+R_3, \quad \deg(R_3)=\deg(R_2)-1,\\
&\cdots\\
&\bar{R}_{q-2}=T_{q-1}\bar{R}_{q-1}+R_q, \quad \deg(R_q)=pq,\\
&\bar{R}_{q-1}=T_q\bar{R}_q.
\end{aligned} \tag{4.5}$$

We define $\delta_1,...,\delta_{p+q}$ as follows: For $1\leqslant i\leqslant p-q$, δ_i is the coefficient T_0 of degree $\deg(T_0)-i=p-q-i$, and for $j=1,...,q$, $\delta_{p-q+2j-1}$ is the leading coefficient of R_j, and δ_{p-q+2j} is the coefficient of T_j of degree 0. Then $\delta_1,...,\delta_{p+q}$ are invariant rational functions, and every invariant rational function is a rational function of $\delta_1,...,\delta_{p+q}$.

Proof: Since $T_i=t_i$ and the leading coefficient of R_i is equal to that of r_i, the invariance of δ_i is obvious from the argument above. By (4.4), $\bar{f}$ and $\bar{g}$ are expressed as polynomials of $t_0=X^{p-q}+\delta_1X^{p-q-1}+\cdots+\delta_{p-q}$, $t_j=X+\delta_{p-q+2j}$, and $\delta_{p-q+2j-1}$ with $j=1,...,q$. Thus $\sigma_1,...,\sigma_{p+q}$ are all expressed as polynomials of $\delta_1,...,\delta_{p+q}$. Therefore, by Theorem 4.1 every invariant rational function is a rational function of $\delta_1,...,\delta_{p+q}$.

The functions σ_i, δ_i $(1\leqslant i\leqslant p+q)$ we have obtained are all rational functions but not all polynomials. Now, we define invariant polynomials γ_i as follows: First, define $\gamma_i=\delta_i$ for $i=1,...,p-q$. Next consider the following modified Euclidean algorithm:

$$\begin{aligned}
&F=T_0'G+R_1', \quad \deg(R_1')=\deg(G)-1,\\
&\gamma_{p-q+1}^2G=T_1'R_1'+R_2', \quad \deg(R_2')=\deg(R_1')-1,\\
&\gamma_{p-q+3}^2R_1'=T_2'R_2'+R_3', \quad \deg(R_3')=\deg(R_2')-1,\\
&\cdots\\
&\gamma_{p+q-3}^2R_{q-2}'=T_{q-1}'R_{q-1}'+R_q',\\
&\deg(R_q')=\deg(R_{q-1}')-1,\\
&\gamma_{p+q-1}^2R_{q-1}'=T_q'R_q',
\end{aligned} \tag{4.6}$$

where $\gamma_{p-q+2j-1}$ is the leading coefficient of R_j' and γ_{p-q+2j} is the coefficient of T_j' of degree 0 for $j=1,...,q$. Note that every γ_i is a polynomial in the entries of M.

Multiplying the second equation in (4.5) by γ_{p-q+1}^2 and the $j+1$th equation in (4.5) by $\gamma_{p-q+2j-1}^2\gamma_{p-q+2j-3}$ for $j=2,...,q$, we have

$$\begin{aligned}
&\gamma_{p-q+1}^2G=(\gamma_{p-q+1}T_1)(\gamma_{p-q+1}\bar{R}_1)+\gamma_{p-q+1}^2R_2,\\
&\gamma_{p-q+3}^2\gamma_{p-q+1}\bar{R}_1\\
&\quad=(\gamma_{p-q+3}\gamma_{p-q+1}T_2)(\gamma_{p-q+3}\bar{R}_2)\\
&\qquad+\gamma_{p-q+3}^2\gamma_{p-q+1}R_3,\\
&\cdots
\end{aligned} \tag{4.7}$$

$$\gamma_{p+q-3}^2\gamma_{p+q-5}\overline{R}_{q-2}$$
$$= (\gamma_{p+q-3}\gamma_{p+q-5}T_{q-1})$$
$$\times(\gamma_{p+q-3}\overline{R}_{q-1}) + \gamma_{p+q-3}^2\gamma_{p+q-5}R_q,$$
$$\gamma_{p+q-1}^2\gamma_{p+q-3}\overline{R}_{q-1}$$
$$= (\gamma_{p+q-1}\gamma_{p+q-3}T_q)(\gamma_{p+q-1}\overline{R}_q).$$

Note $R'_j = \gamma_{p-q+2j-1}\overline{R}_j$ and compare (4.7) with (4.6). Then we find $T'_0 = T_0$, $T'_1 = \gamma_{p-q+1}T_1$, $R'_1 = R_1$, $R'_2 = \gamma_{p-q+1}^2 R_2$, $T'_j = \gamma_{p-q+2j-1}\gamma_{p-q+2j-3}T_j$, and $R'_{j+1} = \gamma_{p-q+2j-1}^2\gamma_{p-q+2j-3}R_{j+1}$, for $j = 2,...,q$.

Thus we obtain the following relation between δ_i and γ_i:

$$\gamma_i = \delta_i \text{ for } i = 1,...,p-q+1,$$
$$\gamma_{p-q+2} = \delta_{p-q+2}\delta_{p-q+1},$$
$$\gamma_{p-q+3} = \delta_{p-q+3}\delta_{p-q+1}^2, \tag{4.8}$$
$$\gamma_{p-q+2j} = \delta_{p-q+2j}\gamma_{p-q+2j-3}\gamma_{p-q+2j-1},$$
$$\gamma_{p-q+2j+1} = \delta_{p-q+2j+1}\gamma_{p-q+2j-3}\gamma_{p-q+2j-1}^2,$$

for $j = 2,...,q-1$ and

$$\gamma_{p+q} = \delta_{p+q}\gamma_{p+q-3}\gamma_{p+q-1}.$$

Equation (4.8) shows that γ_i are expressed with δ_k and vice versa. Finally, we have the following theorem.

Theorem 4.3: The polynomials $\gamma_1,...,\gamma_{p+q}$ defined as above are invariant functions, and any invariant rational function is a rational function of $\gamma_1,...,\gamma_{p+q}$.

Example: Let $M = \left(\begin{smallmatrix} A & B \\ C & D \end{smallmatrix}\right)$ be a (2.1) supermatrix, where

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix}, \quad B = \begin{pmatrix} b_1 \\ b_2 \end{pmatrix}, \quad C = (c_1, c_2), \quad D = (d).$$

Then the characteristic function $h_M(X)$ is, by definition, equal to

$$\frac{\det(XE - A)}{\det(XE - D - C(XE - A)^{-1}B)} = \frac{f_A(X)^2}{f_A(X)(X-d) - k(X)},$$

where

$$f_A(X) = X^2 - (\operatorname{tr} A)X + \det A$$

and

$$k(X) = (X - a_{22})c_1 b_1 + a_{12}c_1 b_2 + a_{21}c_2 b_1 + (X - a_{11})c_2 b_2.$$

We apply the modified Euclidean algorithm to $F(X) = f_A(X)^2$ and $G(X) = f_A(X)(X-d) - k(X)$:

$$F(X) = (X + \gamma_1)G(X) + \gamma_2 X^2 + \omega_1 X + \omega_0,$$
$$\gamma_2^2 G(X) = (\gamma_2 X + \gamma_3)(\gamma_2 X^2 + \omega_1 X + \omega_0),$$

where

$$\gamma_1 = -\operatorname{str} M, \quad \gamma_2 = f_A(d) + c_1 b_1 + c_2 b_2,$$
$$\gamma_3 = -df_A(d) - k(2d),$$
$$\omega_1 = -f_A(d)\operatorname{tr} A + k(-\operatorname{str} M),$$
$$\omega_0 = f_A(d)\det A - k(0)\operatorname{str} M.$$

Thus we obtain the invariant polynomials $\gamma_1, \gamma_2, \gamma_3$. Then $h_M(X)$ is reduced to the irreducible expression:

$$h_M(X) = (X^2 + \sigma_1 X + \sigma_2)/(X + \sigma_3),$$

where $\sigma_3 = \gamma_3/\gamma_2 = -d - k(d)/\gamma_2$, $\sigma_1 = \gamma_1 + \sigma_3$, and $\sigma_2 = \gamma_2 + \gamma_1\sigma_3$.

Remark: It would be easily checked that γ_1, γ_2, and γ_3 in the above example are expressed with $\operatorname{str}(M)$, $\operatorname{str}(M^2)$, and $\operatorname{str}(M^3)$. This fact will be generalized to general supermatrices.

ACKNOWLEDGMENT

We express our thanks to the referee for his or her valuable suggestions.

[1] F. A. Berezin, *The Method of Second Quantization* (Academic, New York, 1966).
[2] F. A. Berezin, *Introduction to Superanalysis* (Reidel, Dordrecht, 1987).
[3] B. DeWitt, *Supermanifolds* (Cambridge U.P., London, 1984).
[4] Y. Kobayashi and S. Nagamachi, J. Math. Phys. **25**, 3367 (1984).
[5] A. Rogers, J. Math. Phys. **21**, 1352 (1980).
[6] Y. I. Manin, *Gauge Field Theory and Complex Geometry* (Springer-Verlag, Berlin, 1988).
[7] J. Wess and J. Bagger, *Supersymmetry and Supergravity* (Princeton U.P., Princeton, NJ, 1983).
[8] Y. Kobayashi and S. Nagamachi, Proc. Jpn. Acad. **64**, 249 (1988).
[9] F. Constantinescu, J. Stat. Phys. **50**, 1167 (1988); F. Constantinescu and H. F. de Groote, J. Math. Phys. **30**, 981 (1989).
[10] T. A. Springer, *Invariant Theory*, Lecture Notes in Mathematics, Vol. 585 (Springer-Verlag, Berlin, 1977).

S-matrix and Jost functions of Schrödinger Hamiltonian related to the Stark effect

Andrei A. Kvitsinsky[a)]
Physics Department, University of California, Los Angeles, California 90024

Vadim V. Kostrykin
Department of Mathematical and Computational Physics, Institute for Physics, Leningrad University, Leningrad SU-199034, USSR

(Received 30 January 1990; accepted for publication 23 May 1990)

The scattering theory for the Hamiltonian of the Stark effect is considered. A partial decomposition of the S-matrix is derived corresponding to separation of variables in the parabolic coordinates, and the analytic structure of the partial Jost functions and S-matrices are studied.

I. INTRODUCTION

The Stark effect for a hydrogen-like system in an external constant electric field is described by the Hamiltonian:

$$H = H_0 + V(x), \quad x \in R^3,$$
$$H_0 = -\Delta + (\mathscr{E},x), \quad V(x) = z/|x|. \tag{1.1}$$

Recently, the scattering theory for this Hamiltonian was developed in Refs. 1–6. In the framework of the scattering theory the Stark "energy levels" get the natural meaning of the resonances corresponding to the poles of S-matrix in the complex energy plane. To justify this interpretation, one has to prove two facts:

(i) the S-matrix for the Hamiltonian (1.1) is an analytic function of the energy E within a domain Ω including the real half-axis $E<0$;

(ii) in the limit $\mathscr{E} \to 0$ the poles of the S-matrix in Ω tend to the binding energies of the Hamiltonian $h = -\Delta + V$.

Previously[4] both (i) and (ii) were proved for the Stark Hamiltonians with short-range potentials V. For the Coulomb potential, the basic property of the S-matrix (i) has not been established until now. But the limit $\mathscr{E} \to 0$ is well studied (see Refs. 7–9 and references therein) via a different approach related to a nonselfadjoint spectral problem for the radial Schrödinger equation in parabolic coordinates. From the scattering theory point of view, one may interpret corresponding results as the solution of (ii) under the assumption that (i) holds true.

The main goal of the present work is to prove (i) in the case of the Coulomb potential.

We attack the problem by making use of the well-known procedure[10] of the separation of variables in the parabolic coordinates for the Schrödinger equation associated with the Hamiltonian (1.1). The proposed scheme involves a lot of technical details but follows, in principle, standard ideology of the scattering theory for the Schrödinger equation without external fields[11,12] that hereafter we shall call "the conventional scattering theory."

The paper is organized as follows. In the next section we describe preliminary results following our previous work.[6] In Sec. III we derive a partial decomposition of the S-matrix, which corresponds to the mentioned separation of variables, and introduce Jost functions. In Sec. IV the analytic structure of the Jost functions and partial S-matrices is studied.

II. PRELIMINARIES

The basic objects of scattering theory for the Hamiltonian (1.1) are the wave operators

$$U^{(\pm)} = s - \lim_{t \to \mp\infty} e^{iHt} e^{-iH_0 t}. \tag{2.1}$$

The existence and completeness of these operators were proved[1-3] for a large class of the potentials V that includes the Coulomb potential.

The kernels of the wave operators (2.1) are eigenfunctions of the absolutely continuous spectrum $\mathbf{R}^1$ of the Hamiltonian (1.1). These kernels are of the simplest structure in a special representation[4] that diagonalizes the operator H_0.

To describe this representation, we introduce the following notations:

$$\epsilon = |\mathscr{E}|, \quad x_3 = -(x,\hat{\mathscr{E}}), \quad x_\perp = x + \hat{\mathscr{E}} x_3,$$

so that ϵ is the electric field intensity, x_3 is the component of x antiparallel to the field, and $x_\perp \in \mathbf{R}^2$ is the projection of x onto the plane orthogonal to the field.

Consider now the unitary transformation $\mathscr{F}$ in $L_2(\mathbf{R}^3)$:

$$(\mathscr{F}f)(k) = \int f(x)\Psi_0(x,k)\,dx; \quad k = \{k_\perp, k_3\} \in \mathbf{R}^3,$$

with the kernel expressed through the Airy function:[13]

$$\Psi_0(x,k) = (\epsilon^{1/3}/2\pi)\mathrm{Ai}(\epsilon^{1/3}(x_3 - k_3)) \times \exp\{i(k_\perp, x_\perp)\}.$$

The function Ψ_0 is a wave function of the Hamiltonian H_0:

$$H_0\Psi_0(\cdot,k) = E(k)\Psi_0(\cdot,k),$$

corresponding to the energy

$$E(k) = k_\perp^2 + \epsilon k_3. \tag{2.2}$$

The transformation $\mathscr{F}$ yields mentioned representation for the Hamiltonian (1.1):

$$H \to \mathscr{F}H\mathscr{F}^* = E(k) + \mathscr{F}V\mathscr{F}^*, \tag{2.3}$$

in which the kernels of the wave operators (2.1) are given by[6]

[a)] Permanent address: Department of Mathematical and Computational Physics, Institute for Physics, Leningrad University, Leningrad SU-199034, USSR.

 0022-2488/90/112731-06$03.00

$$U^{(\pm)}(k,k') = \delta(k-k') - \frac{t(k,k';E(k')\pm i0)}{E(k)-E(k')\mp i0}, \tag{2.4}$$

where t stands for the T-matrix of the Hamiltonian (1.1).

The S operator related to the Stark effect is defined in a standard way,

$$S = U^{(-)*}U^{(+)}. \tag{2.5}$$

Making use of Eq. (2.4) leads one[6] to the following formula for the kernel of S:

$$S(k,k') = \delta(k-k') - 2\pi i\delta(E(k)-E(k')) \times t(k,k';E(k)+i0). \tag{2.6}$$

According to this representation, we can write S operator as a direct integral over the continuous spectrum of H:

$$S = \epsilon^{-1}\int_{-\infty}^{\infty} \oplus S(E)\,dE,$$

where $S(E)$ denotes the S-matrix at fixed energy E. This is an integral operator in $L_2(\mathbf{R}^2)$ with a kernel expressed through the on-shell T-matrix:

$$S(k_\perp,k_\perp';E) = \delta(k_\perp-k_\perp') - 2\pi i\epsilon^{-1}t(k,k';E+i0)|_{E(k)=E(k')=E}. \tag{2.7}$$

Equations (2.4)–(2.7) are similar to corresponding formulas of the conventional scattering theory[12] and differ from them only by the form of the "dispersion rule" (2.2) [in the conventional scattering theory $E(k)=k^2$]. However, this analogy is as yet formal until we give a physical interpretation of the S-matrix in our problem. To this end one has to establish which dynamical processes are described by the S-matrix (2.7).

The problem was solved in our previous work[6] by means of the investigation of asymptotics of the wave functions of the Hamiltonian (1.1), which are defined as a transformation of the wave operators associated with the representation (2.3):

$$\Psi(x,k) = \int \mathcal{F}(x,q)U^{(+)}(q,k)\,dq.$$

For purposes of the present paper we need only one particular result of Ref. 6. Namely, let us consider the trajectories of the classical motion in constant electric field. These are parabolas parametrized by two-dimensional vector $q_\perp$:

$$\beta(q_\perp) = [x: x_\perp = q_\perp t,\ x_3 = -\epsilon t^2/4 + (E-q_\perp^2)/\epsilon; t\in(-\infty,\infty)]. \tag{2.8}$$

As $x\to\infty$ along a parabola $\beta(q_\perp)$, the following representation holds true in the sense of distributions on the Hilbert space $L_2(\mathbf{R}^2)$ with respect to the argument $k_\perp$:

$$\Psi(x,k)|_{\substack{x\to\infty \\ x\in\beta(q_\perp)}} \sim -(i\epsilon/2\pi)[\delta(k_\perp+q_\perp)W_-(x,|k_\perp|) - S(q_\perp,k_\perp;E)W_+(x,|q_\perp|)], \tag{2.9}$$

where $W_\pm$ are given by linear combinations of the Airy functions:

$$W_\pm(x,\lambda) = -\pi\epsilon^{-1/6}\exp(\pm i\lambda|x_\perp|) \times[(E/\epsilon-x_3)^2-x_\perp^2]^{-1/4} \times G_\pm(-\epsilon^{-1/3}[(E-\lambda^2)/\epsilon-x_3]), \tag{2.10}$$

$$G_\pm(t) = \tfrac{1}{2}[\mathrm{Bi}(t)\pm i\,\mathrm{Ai}(t)]. \tag{2.11}$$

The functions $W_\pm$ are analogs of the incoming and outgoing spherical waves arising in conventional scattering theory; the first term in (2.9) describes an incoming wave proceeding along the parabola $\beta(k_\perp)$ whereas the second term is an outgoing wave moving along the parabola $\beta(q_\perp)$. Therefore, the S-matrix (2.7) determines the transition of scattered particle from the in-trajectory $\beta(k_\perp)$ to the out-trajectory $\beta(q_\perp)$.

The representation (2.9) means that for any smooth function $\varphi(k_\perp)\in L_2(\mathbf{R}^2)$,

$$\tilde{\Psi}(x;E) = \int \Psi(x,k)\varphi(k_\perp)dk_\perp|_{\beta(q_\perp)\ni x\to\infty} \sim \varphi(-q_\perp)W_-(x,|q_\perp|) - (S(E)\varphi)(q_\perp)W_+(x,|q_\perp|). \tag{2.12}$$

This equation describes asymptotic behavior of wave packets composed of the wave functions of the Hamiltonian (1.1) with the same energy.

III. PARTIAL DECOMPOSITION OF THE S-MATRIX AND JOST FUNCTIONS

In this section we get a partial representation of the S-matrix (2.7) analogous to the standard decomposition of S-matrix on the spherical harmonics in the conventional scattering theory.[12] It enables us to introduce in a consistent way the phase shifts and the Jost functions related to the Stark effect.

Our starting point is the well-known[10] procedure of the separation of variables in the Schrödinger equation for the Hamiltonian (1.1):

$$(-\Delta - \epsilon x_3 + z/|x| - E)\Psi = 0.$$

Upon writing this equation in parabolic coordinates

$$\xi = |x| + x_3\in[0,\infty),\quad \eta = |x| - x_3\in[0,\infty),\quad \phi = \mathrm{atan}(x_2/x_1)\in[0,2\pi]\quad (x_\perp = \{x_1,x_2\}), \tag{3.1}$$

we can construct solutions of the form

$$\Psi_{mn}(x,E) = \eta^{-1/2}H_{mn}(\eta,E)Y_{mn}(\xi,E)e^{im\phi};\quad m = 0,\pm1,...;\quad n = 0,1,..., \tag{3.2}$$

where Y_{mn} are eigenfunctions of an operator $\Lambda_m(E)$ given by the self-adjoint extension in $L_2(\mathbf{R}_+,\xi\,d\xi)$ of the differential operator:

$$\Lambda_m(E) = -\partial_\xi\,\xi\,\partial_\xi + \frac{m^2}{4\xi} - \frac{E\xi}{4} + \frac{\epsilon\xi^2}{8};$$

$$\Lambda_m(E)Y_{mn}(\cdot,E) = \lambda_{mn}(E)Y_{mn}(\cdot,E). \tag{3.3}$$

Eigenvalues λ_{mn} (so-called constants of separation) determine the equation for H_{mn}:

$$\left[-\partial_\eta^2 + \frac{m^2-1}{4\eta^2} - \frac{E}{4} - \frac{\epsilon}{8}\eta + \frac{z+\lambda_{mn}(E)}{\eta}\right]H_{mn}(\eta,E) = 0, \tag{3.4}$$

with the regular boundary condition at $\xi = 0$:

$$\lim_{\eta \to 0} \eta^{-(|m|+1)/2} H_{mn}(\eta, E) = 1. \tag{3.5}$$

The functions H_{mn} have a meaning of radial wave functions of scattering states of the Hamiltonian (1.1). It is easy to check that, as $\eta \to \infty$, they have the asymptotics

$$H_{mn}(\eta, E) \sim \text{const}\, \eta^{-1/4} \sin\left[\frac{2}{3}\left(\frac{\eta}{2}\right)^{3/2} \epsilon^{1/2} + \left(\frac{\eta}{2\epsilon}\right)^{1/2} E + \Delta_m + \delta_{mn}(E)\right], \tag{3.6}$$

$$\Delta_m = (\pi/4)(2m+1),$$

where the constants δ_{mn} represent the contribution of the Coulomb potential (if $z = 0$, then $\delta_{mn} = 0$ according to the results of Ref. 14).

Following the analogy with the conventional scattering theory, one may interpret the constants δ_{mn} as the phase shifts in our problem.[15] But then one faces the question: How are these phase shifts connected with the S-matrix (2.7)?

To get correspondence between these objects, let us consider the asymptotics of the function (3.2) as $x \to \infty$ along a parabola (2.8). In this case we have

$$\xi = 2q_\perp^2/\epsilon + o(1), \quad \eta = -2x_3 + 2q_\perp^2/\epsilon + o(1).$$

Substituting these equations into (3.2) and (3.6) enables us to write the asymptotics of Ψ_{mn} in terms of the incoming and outgoing waves (2.10):

$$\Psi_{mn}(x, E) \sim \text{const}\, \mathcal{Y}_{mn}(q_\perp, E)\{W_-(x, |q_\perp|) - (-1)^m \exp\{2i\delta_{mn}(E)\} W_+(x, |q_\perp|)\}, \tag{3.7}$$

where

$$\mathcal{Y}_{mn}(q_\perp) = Y_{mn}(2q_\perp^2/\epsilon, E) \exp(im\phi_\perp),$$

and $\phi_\perp$ is the polar angle (3.1) for $q_\perp$.

On the other hand, the function Ψ_{mn} can be represented in the form (2.12) being a wave packet with appropriate density $\varphi(q_\perp)$. Comparing general result (2.12) with (3.7) follows that in the case $\varphi(q_\perp) = \mathcal{Y}_{nm}(-q_\perp)$ and the functions $\mathcal{Y}_{mn}$ are eigenfunctions of the S-matrix,

$$S(E)\mathcal{Y}_{mn} = S_{mn}(E)\mathcal{Y}_{mn},$$

where

$$S_{mn}(E) = \exp\{2i\delta_{mn}(E)\}. \tag{3.8}$$

Upon setting the normalization condition

$$\int_{R^2} dq_\perp\, \mathcal{Y}_{mn}(q_\perp, E) \overline{\mathcal{Y}}_{m'n'}(q_\perp, E) = \delta_{mm'}\delta_{nn'},$$

we can write the spectral representation for the S-matrix,

$$S(q_\perp, k_\perp; E) = \sum_{m=-\infty}^{\infty} \sum_{n=0}^{\infty} S_{mn}(E) \mathcal{Y}_{mn}(q_\perp, E) \overline{\mathcal{Y}}_{nm}(k_\perp, E), \tag{3.9}$$

which is the partial decomposition we are looking for. It provides the connection between the S-matrix (2.7), the phase shifts from (3.6), and the eigenfunctions of the spectral problem (3.3) for the constants of separation.

To study analytical structure of the partial S-matrix (3.8), we define it now in terms of Jost functions related to Eq. (3.6). At first let us introduce solutions of (3.6) $f_{mn}^{(\pm)}$ satisfying the asymptotic boundary conditions

$$\lim_{\eta \to \infty} f_{mn}^{(\pm)}(\eta, E)\left[G_\pm(-\epsilon^{1/3}(E/\epsilon + \eta/2))\right]^{-1} = 1, \tag{3.10}$$

where $G_\pm$ are the Airy functions (2.11) satisfying Eq. (3.4) upon neglecting the terms $\sim \eta^{-2}$ and $\sim \eta^{-1}$. The existence of such solutions is proved in the next section. Obviously, the regular solution H_{mn} is a linear combination of $f_{mn}^{(\pm)}$ which we write in the form

$$H_{mn}(\eta, E) = e^{-i\Delta_m} F_{mn}^{(-)}(E) f_{mn}^{(-)}(\eta, E) + e^{i\Delta_m} F_{mn}^{(+)}(E) f_{mn}^{(+)}(\eta, E), \tag{3.11}$$

where Δ_m is the phase from (3.6).

The coefficients $F_{mn}^{(\pm)}$ we call the Jost functions in our problem. Comparing the asymptotics of (3.12) with (3.6) yields the representation for partial S-matrix (3.8) through the Jost functions

$$S_{mn}(E) = F_{mn}^{(+)}(E)/F_{mn}^{(-)}(E).$$

IV. ANALYTIC CONTINUATION OF THE S-MATRIX

The main result of this section reads as follows.

Theorem 1: The Jost functions $F_{mn}^{(\pm)}$ and the partial S-matrix $S_{mn}(E)$ are analytic in the complex E plane within the wedge

$$\Omega = \{E: |\arg(-E)| < 2\pi/3\}. \tag{4.1}$$

This theorem provides a rigorous interpretation of the Stark effect. Namely, the resonance states of a hydrogen-like system in the electric field may be defined now as poles of $S_{mn}(E)$ in Ω corresponding to the zeros of the Jost function $F_{mn}^{(-)}(E)$. Their real parts give the Stark splitting whereas the imaginary parts determine widths of the resonances.

The behavior of these resonances in the limit $\epsilon \downarrow 0$ is described by the following theorem.

Theorem 2: Let $E_{nm}^{(n_2)}$ be the zeros of $F_{mn}^{(-)}$ in Ω; $n_2 = 0, 1, \ldots$. Then

$$\operatorname{Re}\left[-E_{nm}^{(n_2)}\right]^{1/2} \underset{\epsilon \downarrow 0}{=} -\frac{z}{2N} - \frac{3N^2\Delta}{z^2}\epsilon + o(\epsilon^2), \tag{4.2a}$$

$$\operatorname{Im}\left[-E_{nm}^{(n_2)}\right]^{1/2} \underset{\epsilon \downarrow 0}{=} -\frac{z}{4N(n_2 + |m|)! n_2!}\left[\frac{|z|^3}{2N^3\epsilon}\right]^{2n_2 + |m| + 1} \times \exp\left\{-\frac{|z|^3}{12N^3}\epsilon^{-1} - 3\Delta\right\}(1 + o(1)), \tag{4.2b}$$

where

$$N = n + n_2 + |m| + 1, \quad \Delta = n - n_2.$$

In the case $z < 0$, the leading term of (4.2) gives the binding energy of the Coulomb system with the principal quantum number N; if $z > 0$, it reproduces the poles of the Coulomb S-matrix on the unphysical sheet of energy.

The formulas (4.2) are well known and were obtained long ago.[9] However, previously the constants $E_{nm}^{(n_2)}$ have been interpreted as generalized eigenvalues of a nonselfad-

joint spectral problem associated with Eq. (3.3) at complex E. Theorem 1 provides a correct basis for the calculations performed in Ref. 9. These can be repeated now in the framework of our scheme to get the same result but from a new point of view. Namely, due to (3.11) the zeros of $F_m^{(-)}$ satisfy the equation

$$W(H_{mn}(\eta,E), f_{mn}^{(+)}(\eta,E)) = 0. \tag{4.3}$$

We show below that H_{mn} and $f_{mn}^{(+)}$ are analytic functions of E in Ω, so that the Wronskian (4.3) is well defined. To evaluate these functions at $\epsilon \to 0$ one can follow the method of Ref. 9, which, being applied to Eq. (4.3), provides the formulas (4.2). We restrict ourselves with this remark and refer to Ref. 9 for more technical details.

The rest of this section is devoted to the proof of Theorem 1. From now on we set $\epsilon = 1$. It can be done without loss of generality, since the cases $\epsilon = 1$ and $\epsilon \neq 1$ are related by a scaling transformation of variables.

We start from the following fact.

Theorem 3: The constants of separation $\lambda_{mn}(E)$ are analytic functions in the wedge (4.1).

Using new variable $x = \sqrt{\xi}$, one can write Eq. (3.3) as

$$\left\{\partial_x^2 + \frac{1-4m^2}{4x^2} - Ex^2 - \frac{x^4}{2} + 4\lambda_{mn}(E)\right\} g_{mn}(x,E) = 0, \tag{4.4}$$

where $g_{mn}(x,E) = \sqrt{x}\ Y_{mn}(\xi)$. This equation coincides with the radial Schrödinger equation for the two-dimensional anharmonic oscillator.[16] It allows us to apply directly the technique developed for the one-dimensional anharmonic oscillator[17] providing the stated result.

Now let us introduce the equation

$$\left(-\partial_\eta^2 + \frac{m^2-1}{4\eta} - \frac{\alpha}{\eta} - \frac{E}{4} - \frac{1}{8}\eta\right) h_m = 0, \tag{4.5}$$

which coincides with (3.4) at $\alpha = z + \lambda_{mn}(E)$. We denote by $h_m(\eta,\alpha,E)$ and $f_m^{(\pm)}(\eta,\alpha,E)$ the solutions of (4.5) corresponding to the boundary conditions (3.5) and (3.10), respectively. Obviously

$$H_{mn}(\eta,E) = h_m(\eta, z + \lambda_{mn}(E), E),$$
$$f_{mn}^{(\pm)}(\eta,E) = f_m^{(\pm)}(\eta, z + \lambda_{mn}(E), E). \tag{4.6}$$

Theorem 4: The solutions $h_m(\eta,\alpha,E)$ and $f_m^{(\pm)}(\eta,\alpha,E)$ are analytic functions of α and E.

This theorem, Eq. (4.6), and Theorem 3 provide that solutions $H_{mn}(\eta,E)$ and $f_{mn}^{(\pm)}(\eta,E)$ are analytic functions of E within the sector (4.1). Then, in accordance with (3.11), the Jost functions $F_{mn}^{(\pm)}(E)$ are as well analytic which proves Theorem 1.

From the described results it is clearly seen that peculiarities of the S-matrix in the complex E-plane are generated by the points of nonanalyticity of the constants of separation $\lambda_{mn}(E)$ outside the wedge (4.1). These can be investigated on the basis of Eq. (4.4) by virtue of the technique developed for the anharmonic oscillator.[17] It can be straightforwardly generalized to take into account the centrifugal term in (4.4) which, of course, causes nothing especial new. As a result, the structure of the points of nonanalyticity of $\lambda_{mn}(E)$ is the same as that for the eigenvalues of the anharmonic oscillator.[17]

Thus it remains to prove Theorem 4. We begin with the more simple case of the regular solution h_m. Equation (4.5) for h_m can be converted into an integral equation

$$h_m(\eta,\alpha,E) = h_m^{(0)} + \int_0^\eta K_m(\eta,y)\, V(\alpha,E,y) \times h_m(y,\alpha,E)\, dy, \tag{4.7}$$

where

$$h_m^{(0)} = \eta^{(|m|+1)/2}, \quad V = -\frac{\alpha}{y} + \frac{E}{4} + \frac{y}{8}, \tag{4.8}$$

and K_m is the Green's function for the operator $[\partial_\eta^2 + (1-m^2)/4\eta]$ corresponding to the boundary condition (3.5):

$$K_m(\eta,y) = \begin{cases} m^{-1}\sqrt{\eta y}\,[(\eta/y)^{m/2} - (y/\eta)^{m/2}], & m \neq 0, \\ \sqrt{\eta y}\,\log(y/\eta), & m = 0. \end{cases}$$

As (4.7) is a Volterra equation, it can be solved by iteration:

$$h_m = \sum_{k=0}^{\infty} h_m^{(k)}, \tag{4.9}$$

where $h_m^{(0)}$ is defined in (4.8) and

$$h_m^{(k)}(\eta,\alpha,E) = \int_0^\eta K_m(\eta,y)\, V(\alpha,E,y)\, h_m^{(k-1)}(y,\alpha,E)\, dy. \tag{4.10}$$

Let $m \neq 0$. Inserting into (4.10) the bound

$$|K_m(\eta,y)| \leqslant (2/|m|)(\eta/y)^{|m|/2}\sqrt{\eta y}, \quad \eta \geqslant y,$$

we get

$$|h_m^{(k)}(\eta,\alpha,E)| \leqslant \eta^{(|m|+1)/2}\, \frac{[(2/|m|)Q(\eta,\alpha,E)]^k}{k!}, \tag{4.11}$$

where

$$Q(\eta,\alpha,E) = \int_0^\eta \left| V(\alpha,E,y) \right| y\, dy \leqslant |\alpha|\eta + \frac{|E|}{8}\eta^2 + \frac{1}{24}\eta^3.$$

The bound (4.11) guarantees uniform convergence of the series (4.9) for α and E in any finite region. Thus h_m is analytic for all α and E provided the individual terms $h_m^{(k)}$ are. The latter is obvious as due to the definition (4.10), $h_m^{(k)}$ are polynomials with respect to α and E. The case $m = 0$ can be treated in a similar manner.

Consider now the functions $f_m^{(\pm)}$. We give the proof only for the function $f_m^{(+)}$. For $f_m^{(-)}$ it can be done quite analogously.

First, we prove that the function $f_m^{(+)}(\eta,\alpha,E)$ can be continued as an analytic function for all $\alpha \in \mathbf{C}$ and E: $\operatorname{Im} E > 0$. To this end we write Eq. (4.5) in the form of a Volterra equation:

$$f_m^{(+)}(\eta,\alpha,E) = g^{(0)}(\eta,E) + \int_\eta^\infty L(E,\eta,y)\, V(y,\alpha) f_m^{(+)}(y,\alpha,E)\, dy, \tag{4.12}$$

where

$$g^{(0)}(\eta,E)=G_+\left(-E-\frac{\eta}{2}\right),\quad V(y,\alpha)=\frac{m^2-1}{4y^2}-\frac{\alpha}{y},\tag{4.13}$$

and L is the Green's function of the operator $(\partial_\eta{}^2+E/4+\eta/8)$ corresponding to the boundary condition (3.10):

$$L(E,\eta,y)=-4\pi i\left[G_+\left(-E-\frac{\eta}{2}\right)G_-\left(-E-\frac{y}{2}\right)-G_+\left(-E-\frac{y}{2}\right)G-\left(-E-\frac{\eta}{2}\right)\right].$$

Consider the iterative solution of (4.12):

$$f_m{}^{(+)}=\sum_{k=0}^{\infty}g^{(k)},\tag{4.14}$$

$$g^{(k)}(\eta,\alpha,E)=\int_\eta^\infty L(E,\eta,y)V(y,\alpha)g^{(k-1)}(y,\alpha,E)dy.\tag{4.15}$$

For complex E and real η, y the inhomogeneous term and the integral kernel of (4.12) satisfy the bounds

$$\left|G_\pm\left(-E-\frac{\eta}{2}\right)\right|\leqslant\frac{\text{const}}{(1+\eta)^{1/4}}\exp\left\{\mp\left(\frac{\eta}{2}\right)^{1/2}\text{Im}\,E\right\},$$

$$|L(E,\eta,y)|\leqslant\frac{\text{const}}{(1+\eta)^{1/4}(1+y)^{1/4}}\exp\left\{|\text{Im}\,E|\left[\left(\frac{y}{2}\right)^{1/2}-\left(\frac{\eta}{2}\right)^{1/2}\right]\right\},\quad y\geqslant\eta,\tag{4.16}$$

which follow from the asymptotics of the Airy functions.[13] Inserting (4.16) into (4.15) leads us to the bound

$$|g^{(k)}(\eta,\alpha,E)|\leqslant\frac{\exp\{-(\eta/2)^{1/2}|\text{Im}\,E|\}}{(1+\eta)^{1/4}}C^k\times\int_\eta^\infty\frac{dy_1\,V(y_1,\alpha)}{(1+y_1)^{1/2}}\prod_{i=1}^{k-2}\int_{y_i}^\infty dy_{i+1}\frac{V(y_{i+1},\alpha)}{(1+y_{i+1})^{1/2}}\times\int_{y_{k-1}}^\infty\frac{dy_k\,V(y_k,\alpha)}{(1+y_k)^{1/2}}\exp\left\{\left(\frac{y_k}{2}\right)^{1/2}(|\text{Im}\,E|-\text{Im}\,E)\right\}\tag{4.17}$$

with a constant C.

When $\text{Im}\,E>0$ the exponential factor in the last integral vanishes and we get

$$|g^{(k)}(\eta,\alpha,E)|\leqslant\frac{\exp\{-(\eta/2)^{1/2}\,\text{Im}\,E\}}{(1+\eta)^{1/4}}\times\frac{C^k}{k!}\left[\int_\eta^\infty\frac{|V(y,\alpha)|}{(1+y)^{1/2}}dy\right]^k,\quad\text{Im}\,E>0.\tag{4.18}$$

This bound guarantees uniform convergence of the iterative series (4.18) for η in any region $\eta\geqslant a>0$. Therefore, the function $f_m{}^{(+)}$ is analytic for any $\alpha\in\mathbb{C}$, $\text{Im}\,E>0$ and η bounded away from zero, as are the individual terms of (4.14). At the point $\eta=0$ the integral in (4.18) diverges at its lower limit. To include the point $\eta=0$, one should add the centrifugal term of V to the unperturbed Hamiltonian which allows us to improve the bounds (4.16) by factors taking into account the behavior of corrected inhomogeneous term and Green's function near the origin. It can be done exactly as in the conventional scattering theory.[11]

Thus it remains to be shown that $f_m{}^{(+)}$ is analytic in the lower half-plane of E. In this case the scheme above breaks down due to the increasing exponent in the integrand of (4.17). We shall treat the case $\text{Im}\,E<0$ by means of a dilatation technique based on the fact that $f_m{}^{(+)}$ is an analytic function of η.

First we prove that for complex η the analytic continuation of $f_m{}^{(+)}$ has the asymptotics of the same form (3.10) as for real η. Let $\text{Im}\,E>0$ and $\eta=re^{i\theta}$. Generally speaking we have

$$f_m{}^{(+)}(\eta,\alpha,E)\underset{r\to\infty}{\sim}a(\theta)G_+(-E-\eta/2)+b(\theta)G_-(-E-\eta/2),$$

with $a(0)=1$, $b(0)=0$. To show that $a(\theta)\equiv1$, $b(\theta)\equiv0$ we make use of the Montel theorem[11] which, in our case, says that $a(\theta)\equiv1$ provided that

$$\lim_{r\to\infty}\frac{f_m{}^{(+)}(\eta,\alpha,E)}{G_+(-E-\eta/2)}=\text{const}\,a(\theta),\quad\eta=e^{i\theta}r.\tag{4.19}$$

To check the latter let us consider the Wronskian

$$G_-\left(-E-\frac{\eta}{2}\right)\partial_r f_m{}^{(+)}(\eta,\alpha,E)-f_m{}^{(+)}(\eta,\alpha,E)\partial_r G_-\left(-E-\frac{\eta}{2}\right)\underset{r\to\infty}{\sim}\frac{ie^{i\theta}}{2\pi}a(\theta).$$

This asymptotic equation has a solution of the form

$$f_m{}^{(+)}(\eta,\alpha,E)\underset{r\to\infty}{=}(ie^{i\theta}/2\pi)a(\theta)G_-(-E-\eta/2)\times\int_{r_0}^r d\xi\,[G_-(-E-\xi e^{i\theta})]^{-2}(1+o(1)).$$

The asymptotics of the integral in this formula can be evaluated by making use of the substitution $\xi=xr$ and corresponding asymptotics of G_- involved in the integrand. As a result one gets (4.19).

Next we prove that $f_m{}^{(+)}(re^{i\theta},\alpha,E)$ is an analytic function on the whole E-plane for $\theta\in(0,2\pi/3)$. Consider Eq. (4.5) along a fixed ray $\arg\eta=\theta$. It can be converted into the integral equation

$$f_m{}^{(+)}(re^{i\theta},\alpha,E)=g^{(0)}(re^{i\theta},E)+\int_r^\infty dy\,L(E,re^{i\theta},ye^{i\theta})\times e^{2i\theta}V(ye^{i\theta},\alpha)f_m{}^{(+)}(ye^{i\theta},\alpha,E),\tag{4.20}$$

with the same notations as in (4.12). Using the bounds

$$\left|G_\pm\left(-E-\frac{re^{i\theta}}{2}\right)\right|\leqslant\frac{A}{(1+r)^{1/4}}\exp\left\{\mp\frac{2}{3}\left(\frac{r}{2}\right)^{3/2}\sin\frac{3}{2}\theta\mp\left(\frac{r}{2}\right)^{1/2}|E|\sin\left(\frac{\theta}{2}+\arg E\right)\right\},\tag{4.21}$$

we find that for $\sin 3\theta/2 > 0$ the terms of the iterative series for (4.20) are bounded as follows:

$$|g^{(k)}(re^{i\theta},\alpha,E)| \leqslant \frac{c^k}{(1+r)}\frac{Q(r)}{k!} \times \exp\left\{-\frac{2}{3}\left(\frac{r}{2}\right)^{3/2}\sin\frac{3\theta}{2} - \left(\frac{r}{2}\right)^{1/2}|E|\sin\left(\frac{\theta}{2}+\arg E\right)\right\},$$

where

$$Q(r) = \int_r^\infty \frac{|V(e^{i\theta}y,\alpha)|}{(1+y)^{1/2}}\,dy.$$

Thus the iterative solution converges uniformly within any wedge $\theta\in[\delta, 2\pi/3-\delta]$ $(\delta>0)$ to an analytic function satisfying the bound

$$|f_m^{(+)}(re^{i\theta},\alpha,E)| \leqslant (1+r)^{-1/4}\exp\{CQ(r)\} \times \exp\left\{-\frac{2}{3}\left(\frac{r}{2}\right)^{3/2}\sin\frac{3\theta}{2} - \left(\frac{r}{2}\right)^{1/2}|E|\sin\left(\frac{\theta}{2}+\arg E\right)\right\}. \tag{4.22}$$

Now we can continue $f_m^{(+)}(\eta,\alpha,E)$ with real η into the whole E-plane as follows: First consider Eq. (4.9); due to the bounds (4.16), (4.21), (4.22) the integral in (4.9) can have its contour bent onto a ray

$$\gamma_\theta = \{\xi = \eta + te^{i\theta}, t\in[0,\infty)\}$$

with any $\theta\in(0, 2\pi/3)$; so that

$$f_m^{(+)}(\eta,\alpha,E) = G_+(-E-\eta/2) + \int_{\gamma_\theta} L(E,\eta,\xi)V(\xi,\alpha)f_m^{(+)}(\xi,\alpha,E)\,d\xi. \tag{4.23}$$

Due to (4.16), (4.21), and (4.22) the last integral converges uniformly at the upper limit and defines an analytic function of E in the whole E-plane. So, we have two functions, defined by Eqs. (4.12) and (4.23), which obviously coincide in the region of overlap $\operatorname{Im} E > 0$. Therefore, (4.23) provides continuation of $f_m^{(+)}(\eta,\alpha,E)$ with real η into the lower half-plane of E.

ACKNOWLEDGMENT

One of us (A. A. K.) is grateful to S. Putterman for the hospitality at UCLA.

[1] J. E. Avron and I. W. Herbst, Comm. Math. Phys. **52**, 239 (1977).
[2] I. W. Herbst, Math. Z. **155**, 55 (1977).
[3] K. Yadjima, J. Fac. Sci. Univ. Tokyo Sec. IA **26**, 377 (1979).
[4] K. Yadjima, J. Fac. Sci. Univ. Tokyo Sec. IA **28**, 1 (1981).
[5] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1978), Vol. IV.
[6] A. A. Kvitsinsky and V. V. Kostrykin, Theor. Math. Phys. **75**, 619 (1988).
[7] "Proceedings of the International Workshop on Perturbation Theory at Large Order," Int. J. Quantum Chem. **21** (1) (1982).
[8] S. Graffi and V. Grecchi, Comm. Math. Phys. **62**, 83 (1978); **79**, 91 (1981).
[9] S. Yu. Slavyanov, in *Topics in Mathematical Physics*, edited by M. S. Birman (Leningrad Univ., Leningrad, USSR, 1970), Vol. IV, p. 125.
[10] L. D. Landau and E. M. Lifshitz, *Quantum Mechanics* (Pergamon, Oxford, 1977).
[11] V. de Alfaro and T. Regge, *Potential Scattering* (Wiley, New York, 1965).
[12] S. P. Merkuriev and L. D. Faddeev, *Quantum Scattering Theory for Few-Particle Systems* (Nauka, Moscow, 1985) [in Russian].
[13] M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1964).
[14] D. I. Abramov, Vestnik LGU **2**, 15 (1986) [in Russian].
[15] G. F. Drukarev, Zh. Eksper. Theor. Fiz. **75**, 473 (1978).
[16] E. C. Titchmarch, *Eigenfunctions Expansions Associated with Second Order Differential Equations* (Clarendon, Oxford, 1958).
[17] B. Simon, Ann. Phys. **58**, 76 (1970).

Inverse spectral transform for the Ishimori equation: I. Initial value problem

B. G. Konopelchenko[a)]
C. E. N. Saclay, Service de Physique Theorique, Laboratoire de l'Institut de Recherche Fondamentale du Commissariat à l'Energie Atomique, 91191 Gif-sur-Yvette Cedex, France

B. T. Matkarimov
Novosibirsk State University, Novosibirsk-90, USSR

(Received 29 December 1988; accepted for publication 25 April 1990)

The initial value problem for the $(2+1)$-dimensional integrable generalization of the $(1+1)$-dimensional continuous Heisenberg ferromagnet model (Ishimori-I,II equations) is solved by the $\bar{\partial}$ and nonlocal Riemann–Hilbert problems (NRHP) method.

I. INTRODUCTION

A number of nonlinear partial differential equations and systems are integrable by the inverse spectral transform method (see e.g., Refs. 1–5). Some of the integrable equations are of great interest for physical applications.

The continuous Heisenberg ferromagnet model described by the equation $\mathbf{S}_t(x,t) = \mathbf{S}\times\mathbf{S}_{xx}$ is one of the most important nonlinear model in $1+1$ (one spatial and one temporal) dimensions of this type[6,7] (see also Ref. 8).

The $(2+1)$-dimensional (two spatial and one temporal dimensions) integrable generalization of this model has been found by Ishimori in Ref. 9. It is described by the system

$$\mathbf{S}_t(x,y,t) + \mathbf{S}\times(\mathbf{S}_{xx} + \alpha^2\mathbf{S}_{yy}) + \phi_x\mathbf{S}_y + \phi_y\mathbf{S}_x = 0,$$
$$\phi_{xx} - \alpha^2\phi_{yy} + 2\alpha^2\mathbf{S}(\mathbf{S}_x\times\mathbf{S}_y) = 0, \tag{1.1}$$

where $\mathbf{S}(x,y,t)\mathbf{S}(x,y,t) = 1$, $\phi(x,y,t)$ is a scalar function,

$$\alpha^2 = \pm 1, \quad \mathbf{S}_t = \frac{\partial\mathbf{S}}{\partial t}, \quad \mathbf{S}_x = \frac{\partial\mathbf{S}}{\partial x}, \quad \mathbf{S}_y = \frac{\partial\mathbf{S}}{\partial y}.$$

The Ishimori equation (1.1) has the commutativity operator representation

$$[L_1, L_2] = 0, \tag{1.2}$$

where the operators L_1 and L_2 are of the form[9]

$$L_1 = \alpha\partial_y + P\partial_x, \tag{1.3}$$
$$L_2 = \partial_t - 2iP\partial_x^2 - (iP_x + i\alpha P_y P + \alpha^3 P\phi_x - \phi_y)\partial_x, \tag{1.4}$$

where $P = \mathbf{S}(x,y,t)\boldsymbol{\sigma}$ and $\boldsymbol{\sigma} = (\sigma_1,\sigma_2,\sigma_3)$ are the Pauli matrices and

$$\partial_t = \frac{\partial}{\partial t}, \quad \partial_x = \frac{\partial}{\partial x}, \quad \partial_y = \frac{\partial}{\partial y}.$$

The model (1.1) is the first and very interesting integrable classical model that describes the nonlinear spin system on the plane (x,y). A very important feature of this model is the existence of the topological charge

$$N = \frac{1}{4\pi}\int dx\,dy\,\mathbf{S}(\mathbf{S}_x\times\mathbf{S}_y),$$

which characterizes the solutions of Eq. (1.1).[9] In Ref. 9 the Hirota bilinearization method has been applied to Eq. (1.1). This has allowed us to construct explicitly the vortex type nonsingular solutions of the Ishimori equation (1.1) with $\alpha = \iota$ and with arbitrary topological charge N. The model (1.1) describes, in particular, the time dynamics of the system of vortices on the plane (x,y) (Ref. 9). Within the framework of the direct linearizing transform method the Ishimori equation (1.1) has been derived and treated in Ref. 10.

In the present paper, we study the Ishimori equation (1.1) by the inverse spectral transform method. The solution of the initial value problem for the model (1.1) is given within the class of solutions $\mathbf{S}(x,y,t)$, which tend fast enough to $\mathbf{S}_\infty = (0,0,-1)$ at $x^2+y^2\to\infty$. Both the Ishimori-I (Ish-I) equation [Eq. (1.1) with $\alpha = i$] and the Ishimori-II (Ish-II) equation [Eq. (1.1) with $\alpha = 1$] are considered. For the Ish-I equation the inverse problem equations are connected with the $\bar{\partial}$ equation while for the Ish-II equation the inverse problem equations are generated by the nonlocal Riemann–Hilbert problem (NRHP).

The technique we use has already been applied for the integration of the several $(2+1)$-dimensional nonlinear equations, for instance, the Kadomtsev–Petviashvili equation,[11–13] the Davey–Stewartson equation,[14–17] the Nizhnik–Veselov–Novikov equation,[18,19] and some other equations (see e.g., Refs. 20 and 21). The Ishimori equation is one more nonlinear equation solvable by the $\bar{\partial}$ problem and nonlocal Riemann–Hilbert problem method.

Note that the function $\phi(x,y,t)$ in the system (1.1) is defined nonuniquely but up to the boundary term of the form $\varphi_1(y+\sigma x,t) + \varphi_2(y-\sigma x,t)$, where φ_1 and φ_2 are arbitrary functions decreasing at the infinities. Different choices of the functions φ_1 and φ_2 give rise to different operators L_2 and to different time evolutions of the initial data $\mathbf{S}(x,y,0)$ for Eq. (1.1).

In the present paper we will consider when all these boundary terms vanish ($\varphi_1 = \varphi_2 = 0$). Note also that the Ish-I equation has been discussed briefly in Ref. 22.

II. THE ISHIMORI-I EQUATION

First, we will discuss the initial value problem for the Ish-I equation (1.1) ($\alpha = i$). We will assume that $\mathbf{S}(x,y,t)$

[a)] Permanent address: Institute of Nuclear Physics, Novosibirsk-90, 630090, USSR.

 0022-2488/90/112737-10$03.00

$\to (0,0,-1)$ at $x^2+y^2\to\infty$ fast enough.

In the case $\alpha = i$, the linear problem $L_1\psi = 0$, where the operator L_1 given by (1.3), can be rewritten in the form

$$\begin{pmatrix} \partial_{\bar z} & 0 \\ 0 & \partial_z \end{pmatrix}\psi + \frac{1}{2}(P+\sigma_3)(\partial_z - \partial_{\bar z})\psi = 0, \tag{2.1}$$

where

$$z = \tfrac{1}{2}(y+ix), \quad \bar z = \tfrac{1}{2}(y-ix).$$

The main steps of our construction are typical for the $\bar\partial$ and nonlocal Riemann–Hilbert problems method (shortly $\bar\partial$-NRHP method).[11-21] At first, we should introduce the spectral parameter λ into the problem. This can be done in different ways. We introduce the spectral parameter λ into the problem by the transition in Eq. (2.1) from the function ψ to the function $\chi(z,\bar z,\lambda)$:

$$\chi(z,\bar z,\lambda) \equiv \psi\cdot\begin{pmatrix} e^{-iz/\lambda} & 0 \\ 0 & e^{i\bar z/\lambda} \end{pmatrix}. \tag{2.2}$$

The function $\chi(z,\bar z,\lambda)$ obeys the equation

$$\begin{pmatrix} \partial_{\bar z} & 0 \\ 0 & \partial_z \end{pmatrix}\chi - \frac{i}{2\lambda}[\sigma_3,\chi] + \frac{1}{2}Q\left(\sigma_z - \sigma_{\bar z} + \frac{i}{\lambda}\right)\chi = 0, \tag{2.3}$$

where $Q \equiv P + \sigma_3$.

The spectral problem (2.3) is our starting problem. We should first solve the inverse problem for Eq. (2.3). Following to the standard $\bar\partial$-NRHP method, we will consider the solution of Eq. (2.3) bounded for all λ (possibly with the finite number of singular points) and normalized by the condition $\chi \to_{\lambda\to\infty} 1$. Such solutions of Eq. (2.3) obey also the integral equation

$$\chi(z,\bar z,\lambda) = 1 - \tfrac{1}{2}(G(\cdot,\cdot,\lambda)Q(\partial' - \bar\partial' + i/\lambda)\chi(\cdot,\cdot,\lambda))(z,\bar z), \tag{2.4}$$

where $G(z,\bar z,z',\bar z')$ is the bounded Green's function of the operator

$$L_0 = \begin{pmatrix} \partial_{\bar z} & 0 \\ 0 & \partial_z \end{pmatrix} - (i/2\lambda)[\sigma_3,\cdot]$$

and $\partial' = \partial_{z'}$, $\bar\partial' = \partial_{\bar z'}$.

Emphasize that just the introduction of the spectral parameter λ via (2.2) with $1/\lambda$, instead of a usual λ, leads to the function χ, which admits the canonical normalization (i.e., $\chi \to_{\lambda\to\infty} 1$).

The Green's function $G(z,\bar z,z',\bar z')$ can be easily calculated if one notes that, similar to the Davey–Stewartson equation case,[17] the operator L_0 is representable in the factorized form

$$L_0 = E_\lambda^{-1}\mathscr{D}E_\lambda, \tag{2.5}$$

where

$$\mathscr{D} = \begin{pmatrix} \partial_{\bar z} & 0 \\ 0 & \partial_z \end{pmatrix}$$

and the operator E_λ acts as follows:

$$E_\lambda\phi(z,\bar z) = \begin{pmatrix} \phi_{11}(z,\bar z), & e^{(i/\lambda)\bar z + (i/\bar\lambda)z}\phi_{12}(z,\bar z) \\ e^{-(i/\lambda)z - (i/\bar\lambda)\bar z}\phi_{21}(z,\bar z), & \phi_{22}(z,\bar z) \end{pmatrix}, \tag{2.6}$$

where $\phi(z,\bar z)$ is an arbitrary 2×2 matrix.

The formula (2.5) immediately gives

$$\hat G \doteq L_0^{-1} = E_\lambda^{-1}\mathscr{D}^{-1}E_\lambda. \tag{2.7}$$

Hence, the operator $\hat G$ acts as follows:

$$(\hat G\phi)(z,\bar z) = \frac{1}{2\pi i}\int dz'\wedge d\bar z' \times\begin{pmatrix} \dfrac{\phi_{11}(z',\bar z')}{z'-z}, & \dfrac{\phi_{12}(z',\bar z')}{z'-z}e^{(i/\lambda)(\bar z - \bar z') + (i/\bar\lambda)(z-z')} \\ \dfrac{\phi_{21}(z',\bar z')}{\bar z' - \bar z}e^{-(i/\lambda)(z-z') - (i/\bar\lambda)(\bar z - \bar z')}, & \dfrac{\phi_{22}(z',\bar z')}{\bar z' - \bar z} \end{pmatrix}, \tag{2.8}$$

where $\phi(z,\bar z')$ is an arbitrary 2×2 matrix. The kernel $G(z-z',\bar z-\bar z')$ of the operator $\hat G$ is the Green's function G we are interested in.

It is easy to see from (2.8) that the Green's function G is nowhere analytic in λ. As a result, the solution of Eq. (2.4) is nowhere analytic too.

Following the $\bar\partial$ method, we must now construct a corresponding $\bar\partial$ equation for χ. Differentiating Eq. (2.4) with respect to $\bar\lambda$ and taking account the explicit form of the Green's function (2.8), we obtain

$$\frac{\partial\chi(z,\bar z;\lambda,\bar\lambda)}{\partial\bar\lambda} = \begin{pmatrix} 0, & F_1(\lambda,\bar\lambda)e^{(i/\lambda)\bar z + (i/\bar\lambda)z} \\ F_2(\lambda,\bar\lambda)e^{-(i/\lambda)z - (i/\bar\lambda)\bar z}, & 0 \end{pmatrix} - \frac{1}{2}\left(GQ\left(\partial' - \bar\partial' + \frac{i}{\lambda}\right)\frac{\partial\chi(\cdot,\cdot)}{\partial\bar\lambda}\right)(z,\bar z), \tag{2.9}$$

where

$$F_1(\lambda,\bar\lambda) = \frac{1}{4\pi\bar\lambda^2}\int dz\wedge d\hat z\, e^{-(i/\lambda)\bar z - (i/\bar\lambda)z} \times\sum_{k=1}^{2} Q_{1k}(z,\hat z)\left(\bar\partial' - \partial' - \frac{i}{\lambda}\right)\chi_{k2}(z,\bar z,\lambda,\bar\lambda),$$

$$F_2(\lambda,\bar\lambda) = -\frac{1}{4\pi\bar\lambda^2}\int dz\wedge d\bar z\, e^{(i/\lambda)z + (i/\bar\lambda)\bar z} \times\sum_{k=1}^{2} Q_{2k}(z,\bar z)\left(\bar\partial' - \partial' - \frac{i}{\lambda}\right)\chi_{k1}(z,\bar z,\lambda,\bar\lambda). \tag{2.10}$$

The terms proportional to $\delta(\lambda)$ that could appear in the rhs of (2.9) are equal to zero due to the special matrix structure of Q and the vanishing at $\lambda = 0$ of the integrals that contain the highly oscillating exponents similar to (2.10).

Then we introduce another solution $N(z,\bar z,\lambda,\bar\lambda)$ of Eq. (2.3) that also obeys the integral equation:

$$N(z,\bar z,\lambda,\bar\lambda) = \Sigma_\lambda(z,\bar z) - \tfrac{1}{2}(GQ(\partial' - \bar\partial' + (i/\lambda))N(\cdot,\cdot,\lambda,\bar\lambda))(z,\bar z), \tag{2.11}$$

where

$$\Sigma_\lambda(z,\bar z) = \begin{pmatrix} 0 & e^{+(i/\lambda)z + (i/\bar\lambda)\bar z} \\ e^{-(i/\lambda)\bar z - (i/\bar\lambda)z} & 0 \end{pmatrix}. \tag{2.12}$$

Comparing Eqs. (2.9) and (2.11) and assuming that the homogeneous equation (2.11) has no nontrivial solutions, one obtains

$$\frac{\partial\chi(z,\bar z;\lambda,\bar\lambda)}{\partial\bar\lambda}=N(z,\bar z,\lambda,\bar\lambda)\begin{pmatrix}F_2 & 0\\ 0 & F_1\end{pmatrix}. \tag{2.13}$$

Now it is necessary to establish the relation between functions N and χ. Using the integral equations (2.4) and (2.11) and the identity

$$\begin{aligned}&G(z,z';\lambda,\bar\lambda)Q(z',\bar z')(\partial'-\bar\partial'+i/\lambda)\phi(z',\bar z')\Sigma_\lambda(z',\bar z')\\&\quad=G(z,z';\bar\lambda,\lambda)Q(z',\bar z')(\partial'-\bar\partial'+i/\bar\lambda)\\&\qquad\times\phi(z',\bar z')\Sigma_\lambda(z,\bar z).\end{aligned} \tag{2.14}$$

We find

$$N(z,\bar z;\lambda,\bar\lambda)=\chi(z,\bar z;\bar\lambda,\lambda)\Sigma_\lambda(z,\bar z). \tag{2.15}$$

Substituting the expression (2.15) into (2.13), we finally arrive at the linear $\bar\partial$ equation:

$$\frac{\partial\chi(z,\bar z;\lambda,\bar\lambda)}{\partial\bar\lambda}=\chi(z,\bar z;\bar\lambda,\lambda)F(\lambda,\bar\lambda,z,\bar z), \tag{2.16}$$

where

$$F(\lambda,\bar\lambda,z,\bar z)=\begin{pmatrix}0, & F_1(\lambda,\bar\lambda)e^{(i/\lambda)\bar z+(i/\bar\lambda)z}\\ F_2(\lambda,\bar\lambda)e^{-(i/\lambda)z-(i/\bar\lambda)\bar z}, & 0\end{pmatrix} \tag{2.17}$$

and F_1 and F_2 are given by (2.10).

In order to complete Eq. (2.16) one should also add the information about the singular points of the function χ. We will assume that the homogeneous equation (2.4) has a finite number of simple points $\lambda_1,\dots,\lambda_n$. This implies that the solution of Eq. (2.4) have a form

$$\chi(z,\bar z;\lambda,\bar\lambda)=\sum_i\frac{\phi_i(z,\bar z)}{\lambda-\lambda_i}+\tilde\chi(z,\bar z;\lambda,\bar\lambda), \tag{2.18}$$

where ϕ_i are the solutions of the homogeneous equation (2.4) and $\tilde\chi$ is a function bounded in λ. A precise structure of the singular part S of the function χ can be determined by the use of the following two properties of Eq. (2.4). First, each column of the 2×2 matrix ϕ_i obeys the homogeneous equation (2.4) separately. Therefore, the columns

$$\begin{pmatrix}\phi_{i11}\\ \phi_{i21}\end{pmatrix}\quad\text{and}\quad\begin{pmatrix}\phi_{i12}\\ \phi_{i22}\end{pmatrix}$$

can be the solution of the homogeneous equation (2.4) in different points λ_i and μ_k. Second, it follows from the identity (2.14) that if the matrix

$$\begin{pmatrix}\phi_{11} & 0\\ \phi_{21} & 0\end{pmatrix}$$

is the solution of the homogeneous equation (2.4) at the point λ_i then the matrix

$$\begin{pmatrix}0 & \phi_{11}\\ 0 & \phi_{21}\end{pmatrix}\exp\left(\frac{i\bar z}{\bar\lambda}+\frac{iz}{\lambda}\right)$$

is the solution of the homogeneous equation (2.4) at the point $\bar\lambda_i$. As a consequence of these two facts the singular part S of the function χ is in the form

$$S_{\alpha_i}=\sum_{k=1}^{n_i}\frac{\phi_{(k)\alpha_i}}{\lambda-\lambda_k}+\sum_{e=1}^{n_2}\frac{\phi_{(e)\alpha_2}}{\lambda-\bar\mu_e}\exp\left(-\frac{i\bar z}{\mu_e}-\frac{iz}{\bar\mu_e}\right),$$

$$S_{22}=\sum_{e=1}^{h_2}\frac{\phi_{(e)\alpha_2}}{\lambda-\mu_e}+\sum_{k=1}^{n_1}\frac{\phi_{(k)\alpha_1}}{\lambda-\bar\lambda_k}\exp\left(\frac{i\bar z}{\bar\lambda_k}+\frac{iz}{\lambda_k}\right)\quad(\alpha=1,2). \tag{2.19}$$

Linear $\bar\partial$ equation plays a fundamental role in the inverse spectral transform method. As for other $(2+1)$-dimensional integrable equations[11–21] the $\bar\partial$ equation (2.16) generates the equation of the inverse problem for the spectral problem (2.3). Indeed, taking into account (2.18) and using the generalized Cauchy formula (see e.g., Refs. 20 and 21),

$$\chi(\lambda)=\frac{1}{2\pi i}\int\int_c d\lambda'\wedge d\bar\lambda'\,\frac{\partial\chi(\lambda')}{\partial\bar\lambda'}(\lambda'-\lambda)^{-1}+\frac{1}{2\pi i}\int_{c_\infty}d\lambda'\,\frac{\chi(\lambda')}{\lambda'-\lambda}, \tag{2.20}$$

where C is the entire complex plane, we obtain from (2.16) the following integral equation:

$$\chi(z,\bar z;\lambda,\bar\lambda)=1+S(z,\bar z;\lambda,\bar\lambda)+\frac{1}{2\pi i}\int d\lambda'\wedge d\bar\lambda'\,\frac{\chi(z,\bar z;\bar\lambda',\lambda')F(\lambda',\bar\lambda')}{\lambda'-\lambda}, \tag{2.21}$$

where S is given by (2.19).

The two-dimensional singular integral equation (2.21) is the basic equation of the inverse problem for the spectral problem (2.3). In order to extract the complete set of the inverse problem equations from (2.21) one should use also the relation

$$\lim_{\lambda\to\lambda_i}\left(\chi-\frac{\phi_i}{\lambda-\lambda_i}\right)=\phi_i\begin{pmatrix}-\frac{iz}{\lambda_i^2}+\gamma_{1i} & 0\\ 0 & \frac{i\bar z}{\lambda_i^2}+\gamma_{2i}\end{pmatrix}, \tag{2.22}$$

where γ_i are some constants. This relation can be proved similar to the KP-II and DS-II equation cases[13,15,16] (see Appendix A).

Taking into account (2.22) and proceeding in (2.21) to the limits $\lambda\to\lambda_i$; $\lambda\to\mu_k$, one gets the system of equations

$$\begin{aligned}&\delta_{\alpha1}-\phi_{(i)\alpha1}\left(-\frac{iz}{\lambda_i^2}+\gamma_{1i}\right)+\sum_{k\neq i}^{n_1}\frac{\phi_{(k)\alpha1}}{\lambda_i-\lambda_k}\\&\quad+\sum_{l=1}^{n_2}\frac{\phi_{(e)\alpha2}}{\lambda_i-\bar\mu_e}\exp\left(-\frac{i\bar z}{\mu_e}-\frac{iz}{\bar\mu_e}\right)\\&\quad+\frac{1}{2\pi i}\int d\lambda'\wedge d\bar\lambda'\,\frac{\chi(z,\bar z;\bar\lambda',\lambda')F(\lambda',\bar\lambda';z,\bar z)}{\lambda'-\lambda_i}=0\\&\qquad(\alpha=1,2;i=1,\dots,n_1),\end{aligned}$$

$$\begin{aligned}&\delta_{\alpha2}-\phi_{(j)\alpha2}\left(\frac{i\bar z}{\mu_j^2}+\gamma_{2j}\right)+\sum_{l\neq j}^{n_2}\frac{\phi_{(e)\alpha2}}{\mu_j-\mu_e}\\&\quad+\sum_{k=1}^{n_1}\frac{\phi_{(k)\alpha1}}{\mu_j-\bar\lambda_k}\exp\left(\frac{i\bar z}{\bar\lambda_k}+\frac{iz}{\lambda_k}\right)\end{aligned}$$

$$+\frac{1}{2\pi i}\int d\lambda'\wedge d\bar{\lambda}'\,\frac{\chi(z,\bar{z};\bar{\lambda}',\lambda')F(\lambda',\bar{\lambda}';z,\bar{z})}{\lambda'-\mu_j}=0$$

$$(\alpha=1,2;j=1,\ldots,n_2). \tag{2.23}$$

At last, directly from the problem (2.3) one has $[\sigma_3,\chi_0]=Q\chi_0$ where $\chi_0=\chi(z,\bar{z};\lambda,\bar{\lambda})|_{\lambda=0}$. This gives $P=-\chi_0\sigma_3\chi_0^{-1}$ or equivalently

$$\mathbf{S}(x,y,t)=-\tfrac{1}{2}\operatorname{tr}(\sigma\chi_0\sigma_3\chi_0^{-1}). \tag{2.24}$$

Equations (2.21), (2.23), and (2.24) form the complete set of equations that solve the inverse problem for the spectral problem (2.3). The set

$$\mathcal{F}(\lambda,\bar{\lambda})=\{F_1(\lambda,\bar{\lambda}),F_2(\lambda,\bar{\lambda}),\lambda_i,$$
$$\gamma_{1i}(i=1,\ldots,n_1),\mu_k,\gamma_{2k}(k=1,\ldots,n_2)\}$$

is the inverse problem data for the problem (2.3). Given $\mathcal{F}(\lambda,\bar{\lambda})$, one can calculate functions χ,ϕ_i with the use of the integral equations (2.21) and (2.23). Finally, we reconstruct the spin variable $\mathbf{S}(x,y,t)$ by the formula (2.24) where

$$\chi_0=1+S|_{\lambda=0}+\frac{1}{2\pi i}\int d\lambda'\wedge d\bar{\lambda}'\,\frac{\chi(z,\bar{z};\bar{\lambda}',\lambda')F(\lambda',\bar{\lambda}')}{\lambda'-0}. \tag{2.25}$$

Note that Eqs. (2.21) and (2.23) are solvable at least for the small data $F(\lambda,\bar{\lambda})$.

In the general case, the inverse problem equations give the complex-valued spin vector $\mathbf{S}(x,y,t)$. In order to have a physically sensible real-valued spin vector $\mathbf{S}(x,y,t)$, one must impose the additional constraint on the inverse problem data.

For real-valued $\mathbf{S}(x,y,t)$ one has $\bar{p}=-\sigma_2 p\sigma_2$ and therefore $\bar{Q}=-\sigma_2 Q\sigma_2$. Comparison of Eq. (2.3) and its complex conjugated in this case gives the constraint

$$\bar{\chi}(z,\bar{z};\lambda,\bar{\lambda})=\sigma_2\chi(z,\bar{z};\bar{\lambda},\lambda)\sigma_2. \tag{2.26}$$

This condition leads to the following constraint for the inverse data for the real-valued spin variable $\mathbf{S}$:

$$\bar{F}_1(\lambda,\bar{\lambda})=F_2(\bar{\lambda},\lambda),\quad \phi_{(k)\alpha\beta}=0. \tag{2.27}$$

One can straightforwardly check that the inverse problem data constrained by the condition (2.27), in fact, generates the real-valued $\mathbf{S}(x,y,t)$.

Now we are able to solve the initial value problem for the Ish-I equation. For this purpose one must establish the evolution of the inverse problem data $\mathcal{F}(\lambda,\bar{\lambda})$ in time t. Using the equation $L_2\psi=0$, where the operator L_2 is given by (1.4) with $\alpha=i$, one, in a standard manner, gets

$$\begin{aligned}
&\frac{dF_1(\lambda,\bar{\lambda},t)}{dt}=\frac{i}{2}\left(\frac{1}{\lambda^2}+\frac{1}{\bar{\lambda}^2}\right)F_1(\lambda,\bar{\lambda},t),\\
&\frac{dF_2(\lambda,\bar{\lambda},t)}{dt}=-\frac{i}{2}\left(\frac{1}{\lambda^2}+\frac{1}{\bar{\lambda}^2}\right)F_2(\lambda,\bar{\lambda},t),\\
&\frac{d\lambda_k}{dt}=\frac{d\mu_k}{dt}=0,\\
&\frac{d\gamma_{1i}}{dt}=-\frac{i}{\lambda_i^3},\quad \frac{d\gamma_{2k}}{dt}=\frac{i}{\mu_k^3}.
\end{aligned} \tag{2.28}$$

Therefore,

$$\begin{aligned}
&F_1(\lambda,\bar{\lambda},t)=F_1(\lambda,\bar{\lambda},0)\exp\frac{i}{2}\left(\frac{1}{\lambda^2}+\frac{1}{\bar{\lambda}^2}\right)t,\\
&F_2(\lambda,\bar{\lambda},t)=F_2(\lambda,\bar{\lambda},0)\exp\left[-\frac{i}{2}\left(\frac{1}{\lambda^2}+\frac{1}{\bar{\lambda}^2}\right)t\right],\\
&\gamma_{1i}(t)=\gamma_{1i0}+\frac{i}{\lambda_i^3}t,\quad \gamma_{2i}(t)=\gamma_{2i0}-\frac{i}{\mu_i^3}t,
\end{aligned} \tag{2.29}$$

where γ_{1i0} and γ_{2i0} are arbitrary constants.

The use of the formula (2.29) allows us to solve the initial value problem for Eq. (1.1) by the inverse spectral transform method standard procedure.

$$\mathbf{S}(x,y,0)\overset{(2.10)}{\to}\mathcal{F}(\lambda,\bar{\lambda},0)\overset{(2.29)}{\to}\mathcal{F}(\lambda,\bar{\lambda},t)\overset{(2.21),(2.23),(2.24)}{\to}\mathbf{S}(x,y,t). \tag{2.30}$$

Emphasize that the procedure described gives the solution of the initial value problem for the spin variables $\mathbf{S}$ which tends to the asymptotic value $(0,0,-1)$ sufficiently fast as $x^2+y^2\to\infty$. Note also that the evolution law (2.29) preserve in time the reality conditions (2.27).

As usual (see e.g., Refs. 12–20), one can find the solution of the Ish-I equation (1.1) that corresponds to the case $F(\lambda,\bar{\lambda},t)\equiv 0$ in an explicit form. Indeed, in this case the system (2.23) is the linear algebraic system that can be easily solved with respect to $\phi_{(i)\alpha\beta}$. Substituting these $\phi_{(i)\alpha\beta}$ into (2.21), we obtain the solutions of Eq. (1.1) given by the formula (2.24) where $\chi_0=1+S|_{\lambda=0}$. The simplest from those solution corresponds to the case $n_1=n_2=1$ and it is of the form

$$\begin{aligned}
&S_+\equiv S_1+iS_2=2dc/(ad-bc),\\
&S_-\equiv S_1-iS_2=2ab/(ad-bc),\\
&S_3=(ad+bc)/(bc-ad),
\end{aligned} \tag{2.31}$$

where

$$\begin{aligned}
a=\chi_{011}&=1-\frac{1}{\pi}\left[\frac{1}{\lambda}\left(-\frac{i\bar{z}}{\mu^2}-\frac{it}{\mu^3}+\gamma_2\right)\right.\\
&\left.+\frac{\exp[i(1/\bar{\lambda}-1/\mu)\bar{z}+iz(1/\lambda-1/\mu)]}{\bar{\mu}(\mu-\bar{\lambda})}\right],\\
b=\chi_{012}&=-\frac{1}{\pi}\left[\frac{1}{\mu(\mu-\bar{\lambda})}\right.\\
&\left.+\frac{1}{\bar{\lambda}}\left(-\frac{i\bar{z}}{\mu^2}-\frac{it}{\mu^3}+\gamma_2\right)\right]e^{i\bar{z}/\bar{\lambda}+iz/\lambda},\\
c=\chi_{021}&=\frac{1}{\pi}\left[\frac{1}{\lambda(\bar{\mu}-\lambda)}\right.\\
&\left.+\frac{1}{\mu}\left(-\frac{iz}{\lambda^2}-\frac{it}{\lambda^3}-\gamma_1\right)\right]e^{-i\bar{z}/\mu-iz/\bar{\mu}},\\
d=\chi_{022}&=1+\frac{1}{\pi}\left[\frac{1}{\mu}\left(-\frac{iz}{\lambda^2}-\frac{it}{\lambda^3}-\gamma_1\right)\right.\\
&\left.+\frac{\exp[i\bar{z}(1/\bar{\lambda}-1/\mu)+iz(1/\lambda-1/\bar{\mu})]}{\bar{\lambda}(\bar{\mu}-\lambda)}\right],
\end{aligned}$$

and

$$\pi=\left(\frac{iz}{\lambda^2}+\frac{it}{\lambda^3}+\gamma_1\right)\left(-\frac{i\bar{z}}{\mu^2}-\frac{it}{\mu^3}+\gamma_2\right)+\frac{\exp[i\bar{z}(1/\bar{\lambda}-1/\mu)+iz(1/\lambda-1/\mu)]}{|\mu-\bar{\lambda}|^2}.$$

We emphasize that formulas (2.31) give the solution of the Ish-I equation (1.1) for generic complex $S(x,y,t)$. For real S such solutions are absent, as follows from the constraints (2.27). So the solutions of the Ish-I equation (1.1) constructed here are of different type compared with those found by the Hirota method in Ref. 9. Note, in conclusion, that after completing this paper the more rigorous consideration of the inverse problem for the Ish-I equation has been given in Refs. 23 and 21.

III. THE ISHIMORI-II EQUATION

At the case $\alpha=1$ the linear problem $L_1\psi=0$ is the hyperbolic linear system. In the characteristic variables $\xi=\frac{1}{2}(y+x)$, $\eta=\frac{1}{2}(y-x)$ this problem is of the form

$$\begin{pmatrix}\partial_\eta & 0\\ 0 & \partial_\xi\end{pmatrix}\psi+\frac{1}{2}Q(\partial_\xi-\partial_\eta)\psi=0, \tag{3.1}$$

where $Q(x,y,t)=P(x,y,t)+\sigma_3$.

The spectral parameter λ is introduced into the problem similar to the Ish-I case, namely, by the transition to the function χ:

$$\chi(\xi,\eta,\lambda)\doteq\psi(\xi,\eta)\begin{pmatrix}e^{-i\xi/\lambda}, & 0\\ 0, & e^{i\eta/\lambda}\end{pmatrix}. \tag{3.2}$$

The function χ obeys the equation

$$\begin{pmatrix}\partial_\eta & 0\\ 0 & \partial_\xi\end{pmatrix}\chi-\frac{i}{2\lambda}[\sigma_3,\chi]+\frac{1}{2}Q\left(\partial_\xi-\partial_\eta+\frac{i}{\lambda}\right)\chi=0. \tag{3.3}$$

We will calculate the Green's function for the operator

$$L_0=\begin{pmatrix}\partial_\eta & 0\\ 0 & \partial_\xi\end{pmatrix}-\frac{i}{2\lambda}[\sigma_3,\cdot]$$

by the same method as that has been used in the previous section. Analogously to (2.5) the operator L_0 is representable in the factorized form

$$L_0=E_\lambda^{-1}\mathscr{D}E_\lambda, \tag{3.4}$$

where

$$\mathscr{D}=\begin{pmatrix}\partial_\eta & 0\\ 0 & \partial_\xi\end{pmatrix}$$

and the operator E_λ acts by the rule

$$E_\lambda\phi=\begin{pmatrix}\phi_{11}(\xi,\eta) & e^{-i\eta/\lambda-i\xi/\bar{\lambda}}\phi_{12}(\xi,\eta)\\ e^{i\xi/\lambda+i\eta/\bar{\lambda}}\phi_{21}(\xi/\eta) & \phi_{22}(\xi,\eta)\end{pmatrix}, \tag{3.5}$$

where ϕ is an arbitrary 2×2 matrix. The operator L_0 is not bounded as the function on λ in contrast to the operator (2.5). However, this disadvantage is more than compensated by the much simpler structure of the operator

$$\mathscr{D}^{-1}:\quad \mathscr{D}^{-1}=\begin{pmatrix}\partial_\eta^{-1} & 0\\ 0 & \partial_\xi^{-1}\end{pmatrix}.$$

As a result, the inverse operator $\hat{G}=L_0^{-1}$ is given by the simple formula

$$\hat{G}=E_\lambda^{-1}\mathscr{D}^{-1}E_\lambda, \tag{3.6}$$

i.e.,

$$(\hat{G}\phi)(\xi,\eta)\doteq\begin{pmatrix}\partial_\eta^{-1}(\phi_{11}(\xi',\eta')), & \partial_\eta^{-1}(e^{i(\eta-\eta')/\lambda}\phi_{12}(\xi',\eta'))\\ \partial_\xi^{-1}(e^{-i(\xi-\xi')/\lambda}\phi_{21}(\xi',\eta')), & \partial_\xi^{-1}(\phi_{22}(\xi',\eta'))\end{pmatrix}. \tag{3.7}$$

The kernel $G(\xi-\xi',\eta-\eta',\lambda)$ of the operator $\hat{G}$ is the usual Green's function for the operator L_0.

The main feature of the Green's function (3.7), in comparison with (2.8), consists of the absence of $\bar{\lambda}$ dependence. Thus the Green's function (3.7) is an analytic function on the entire complex plane of λ.

Another feature of the Green's function (3.7) is that it is defined nonuniquely. The freedom in the definition of this Green's function is connected with the possibility to choose the different concrete realizations of the formal operators ∂_ξ^{-1} and ∂_η^{-1}. This freedom can be used for the construction of the bounded Green's functions. Indeed, choosing

$$\partial_\eta^{-1}f\doteq\int_{+\infty}^{\eta}d\eta'\,f(\xi,\eta')$$

and

$$\partial_\xi^{-1}f\doteq\int_{-\infty}^{\xi}d\xi'\,f(\xi',\eta),$$

we define the Green's function:

$$(\hat{G}^+(\cdot,\lambda)\phi(\cdot))(\xi,\eta)\doteq\begin{pmatrix}\int_{+\infty}^{\eta}d\eta'\,\phi_{11}(\xi,\eta'), & \int_{+\infty}^{\eta}d\eta'\,e^{i(\eta-\eta')/\lambda}\phi_{12}(\xi,\eta')\\ \int_{-\infty}^{\xi}d\xi'\,e^{-i(\xi-\xi')/\lambda}\phi_{21}(\xi',\eta), & \int_{-\infty}^{\xi}d\xi'\,\phi_{22}(\xi',\eta)\end{pmatrix}. \tag{3.8}$$

The choice

$$\partial_\xi^{-1}f\doteq\int_{+\infty}^{\xi}d\xi'\,f(\xi',\eta)\ \text{and}\ \partial_\eta^{-1}f\doteq\int_{-\infty}^{\eta}d\eta'\,f(\xi,\eta')$$

gives the Green's function

$$(\hat{G}^{-}(\cdot,\lambda)\phi(\cdot))(\xi,\eta)\doteq\begin{pmatrix}\int_{-\infty}^{\eta} d\eta'\,\phi_{11}(\xi,\eta'), & \int_{-\infty}^{\eta} d\eta'\,e^{i(\eta-\eta')/\lambda}\phi_{12}(\xi,\eta')\\ \int_{+\infty}^{\xi} d\xi'\,e^{-i(\xi-\xi')/\lambda}\phi_{21}(\xi',\eta), & \int_{+\infty}^{\xi} d\xi'\,\phi_{22}(\xi',\eta)\end{pmatrix}. \tag{3.9}$$

It is easy to see that the Green's function $G^{+}(\lambda)$ is bounded at the upper half-plane $\operatorname{Im}\lambda>0$, while the Green's function $G^{-}(\lambda)$ is bounded at the lower half-plane.

Now let us introduce the solution χ^{+} and χ^{-} of the problem (3.3) which simultaneously are the solutions of the integral equation

$$\chi^{\pm}(\xi,\eta,\lambda)=1-\{G^{\pm}(\cdot,\lambda)\tfrac{1}{2}Q(\partial'-\tilde{\partial}'+i/\lambda)\chi^{\pm}(\cdot,\lambda)\}(\xi,\eta), \tag{3.10}$$

where G^{+} and G^{-} are given by the formulas (3.8) and (3.9);

$$Q=P+\sigma_3,\quad \partial'=\partial_{\xi'},\quad \tilde{\partial}'=\partial_{\eta'}.$$

As far as the Green's functions G^{+} and G^{-}, the solution χ^{+} and χ^{-} are analytic and bounded functions in the upper and lower half-planes, respectively. Further, since $G^{+}-G^{-}\neq 0$ at $\operatorname{Im}\lambda=0$, then $\chi^{+}-\chi^{-}\neq 0$ at $\operatorname{Im}\lambda=0$ too thus, one can define the function

$$\chi\doteq\begin{cases}\chi^{+}, & \operatorname{Im}\lambda>0,\\ \chi^{-}, & \operatorname{Im}\lambda<0,\end{cases}$$

which is analytic and bounded on the entire complex plane and has a jump across the real axis. So, we arrive at the standard Riemann–Hilbert problem. We will also assume that the homogeneous equation (3.10) has no nontrivial solutions.

At this stage, according to the standard procedure[11–20] one must find out the relation between the functions χ^{+} and χ^{-} on the real axis. This relation can be derived by the same method as for the Davey–Stewartson equation.[15–17]

First, we note that Eqs. (3.10) straightforwardly give

$$(\chi^{+}-\chi^{-})(\xi,\eta,\lambda)=\Gamma(\xi,\eta,\lambda)-[\tilde{G}(\cdot,\lambda)\tfrac{1}{2}Q(\cdot)\times(\partial'-\tilde{\partial}'+i/\lambda)(\chi^{+}-\chi^{-})](\xi,\eta), \tag{3.11}$$

where

$$(\tilde{G}(\cdot,\lambda)\phi)(\xi,\eta)=\begin{pmatrix}\int_{-\infty}^{\eta} d\eta'\,\phi_{11}(\xi,\eta'), & \int_{-\infty}^{\eta} d\eta'\,e^{i(\eta-\eta')/\lambda}\phi_{12}(\xi,\eta')\\ \int_{-\infty}^{\xi} d\xi'\,e^{-i(\xi-\xi')/\lambda}\phi_{21}(\xi',\eta), & \int_{-\infty}^{\xi} d\xi'\,\phi_{22}(\xi',\eta)\end{pmatrix} \tag{3.12}$$

and

$$\Gamma(\xi,\eta,\lambda)=\begin{pmatrix}-\int_{-\infty}^{+\infty} d\eta'\frac{1}{2}\left[Q(\xi,\eta')\left(\partial_{\xi}-\partial_{\eta'}+\frac{i}{\lambda}\right)\chi^{+}(\xi,\eta')\right]_{11}, & -\int_{-\infty}^{+\infty} d\eta'\,e^{i(\eta-\eta')/\lambda}\frac{1}{2}\left[Q(\xi,\eta')\left(\partial_{\xi}-\partial_{\eta'}+\frac{i}{\lambda}\right)\chi^{+}\right]_{12}\\ \int_{-\infty}^{+\infty} d\xi'\,e^{-i(\xi-\xi')/\lambda}\frac{1}{2}\left[Q(\xi',\eta)\left(\partial_{\xi'}-\partial_{\eta}+\frac{i}{\lambda}\right)\chi^{-}(\xi',\eta)\right]_{21}, & \int_{-\infty}^{+\infty} d\xi'\frac{1}{2}\left[Q(\xi',\eta)\left(\partial_{\xi'}-\partial_{\eta}+\frac{i}{\lambda}\right)\chi^{-}(\xi',\eta)\right]_{22}\end{pmatrix}. \tag{3.13}$$

Then we introduce the 2×2 matrix $f(l,k)$ defined by the relation

$$f(l,k)-\int_{-\infty}^{+\infty} dk'\,T^{-}(l,k')f(k',k)=T^{-}(l,k)-T^{+}(l,k), \tag{3.14}$$

where

$$T_{21}^{-}(l,k)=\frac{1}{2\pi}\int\int d\xi\,d\eta\,e^{i\eta/l+i\xi/k}\times\frac{1}{2}\left[Q(\xi,\eta)\left(\partial_{\xi}-\partial_{\eta}+\frac{i}{k}\right)\chi^{-}(\xi,\eta,k)\right]_{21}, \tag{3.15}$$

$$T_{11}^{-}=T_{12}^{-}=T_{22}^{-}=0,$$

and

$$T_{12}^{+}(l,k)=\frac{1}{2\pi}\int\int d\xi\,d\eta\,e^{-i\xi/l-i\eta/k}\times\frac{1}{2}\left[Q(\xi,\eta)\left(\partial_{\xi}-\partial_{\eta}+\frac{i}{k}\right)\chi^{+}(\xi,\eta,k)\right]_{12},$$

$$T_{11}^{+}=T_{22}^{+}=T_{21}^{+}=0. \tag{3.16}$$

The integral equation (3.14), in fact, are easily solved and one gets

$$f(l,k)=\begin{pmatrix}0, & -T_{12}^{+}(l,k)\\ T_{21}^{-}(l,k), & -\int_{-\infty}^{+\infty} dk'\,T_{21}^{-}(l,k')T_{12}^{+}(k',k)\end{pmatrix}. \tag{3.17}$$

Further, the obvious identity

$$\left(\partial_\xi - \partial_\eta + \frac{i}{k}\right)\chi^-(\xi,\eta,k)\Sigma_l(\xi,\eta)f(l,k)\Sigma_k^{-1}(\xi,\eta)$$
$$= \left[\left(\partial_\xi - \partial_\eta + \frac{i}{l}\right)\chi^-(\xi,\eta,k)\right]$$
$$\times\Sigma_l(\xi,\eta)f(l,k)\Sigma_k^{-1}(\xi,\eta) \tag{3.18}$$

holds, where

$$\Sigma_\lambda(\xi,\eta) \doteq \begin{pmatrix} e^{i\xi/\lambda} & 0 \\ 0 & e^{-i\eta/\lambda} \end{pmatrix}. \tag{3.19}$$

At last, using (3.18), one can straightforwardly show that the quantity

$$\int_{-\infty}^{+\infty} dk\,\chi^-(\xi,\eta,k)\Sigma_k(\xi,\eta)f(k,\lambda)\Sigma_\lambda^{-1}(\xi,\eta)$$

obeys the same equation (3.11) as $(\chi^+ - \chi^-)(\xi,\eta,\lambda)$. In virtue of the absence of the nontrivial solutions for the homogeneous equation (3.11) this gives (see Appendix B)

$$\chi^+(\xi,\eta,\lambda) - \chi^-(\xi,\eta,\lambda)$$
$$= \int_{-\infty}^{+\infty} dl\,\chi^-(\xi,\eta,l)\,\Sigma_l(\xi,\eta)f(l,\lambda)\Sigma_\lambda^{-1}(\xi,\eta),$$
$$\operatorname{Im}\lambda = 0. \tag{3.20}$$

Thus the jump $\chi^+ - \chi^-$ at $\operatorname{Im}\lambda = 0$ is expressed linearly and nonlocally via χ^-. So, we have the regular nonlocal Riemann–Hilbert problem.

With the use of the standard formulas that solve the standard Riemann–Hilbert problem (see, e.g., Refs. 11, 13, and 20) we obtain from (3.20) the following integral equation:

$$\chi^-(\xi,\eta,\lambda) - \frac{1}{2\pi i}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dl\,dk$$
$$\times\frac{\chi^-(\xi,\eta,l)\Sigma_l(\xi,\eta)f(l,k)\Sigma_k^{-2}(\xi,\eta)}{k-\lambda+i0} = 1. \tag{3.21}$$

Further, directly from (3.3) one has

$$P(\xi,\eta) = \mathbf{S}(\xi,\eta)\sigma = -g\sigma_3 g^{-1}, \tag{3.22}$$

where $g(\xi,\eta) = \chi(\xi,\eta,\lambda=0)$.

From Eq. (3.21) it follows that

$$g(\xi,\eta) = 1 + \frac{1}{2\pi i}\int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty} dl\,dk$$
$$\times\frac{\chi^-(\xi,\eta,l)\Sigma_l(\xi,\eta)f(l,k)\Sigma_k^{-1}(\xi,\eta)}{k}. \tag{3.23}$$

The integral equation (3.21) and formulas (3.22) and (3.23) are the complete set of the inverse problem equations for the spectral problem (3.3). The functions $T_{21}^-(l,k)$ and $T_{12}^+(l,k)$ are the inverse problem data.

To solve the initial value problem for the Ish-II equation one should, as usual, find the time evolution of the inverse problem data. The evolution laws of the functions T_{12}^+ and T_{21}^- can be found in a standard manner from the second auxiliary linear problem $L_2\psi = 0$ where the operator L_2 is given by (1.4) with $\alpha = 1$. One obtains

$$T_{12}^+(k,\lambda,t) = T_{12}^+(k,\lambda,0)e^{(i/2)(1/k^2+1/\lambda^2)t},$$
$$T_{21}^-(k,\lambda,t) = T_{21}^-(k,\lambda,0)e^{-(i/2)(1/k^2+1/\lambda^2)t}. \tag{3.24}$$

For the 2×2 matrix $f(l,k)$ the time evolution looks like

$$f(l,k,t) = e^{(i\sigma_3/l^2)t}f(l,k,0)e^{-(i\sigma_3/k^2)t}. \tag{3.25}$$

The inverse problem equations (3.21)–(3.23) and the evolution law (3.24) allow us to solve, in principle, the initial value problem for the Ish-II equation by the standard procedure (2.30).

The formulas (3.21)–(3.23) give the possibility to construct exact solutions of the Ish-II equation. In particular, one can construct exact solutions with the functional parameter that are typical for equations connected with the nonlocal Riemann–Hilbert problem.[24] The solution of such type correspond to the factorized functions T_{12}^+ and T_{21}^-:

$$T_{12}^+(l,k,t) = \sum_{n=1}^{N} f_n^+(l)g_n^+(k)e^{(i/2)(1/l^2+1/k^2)t},$$
$$T_{21}^-(l,k,t) = \sum_{n=1}^{N} f_n^-(l)g_n^-(k)e^{-(i/2)(1/l^2+1/k^2)t}, \tag{3.26}$$

where $f^\pm(l)$, $g^\pm(k)$ are arbitrary functions.

For the inverse problem data of such type one has

$$\Sigma_l(\xi,\eta)f(l,k)\Sigma_k^{-1}(\xi,\eta)$$
$$= \sum_{n=1}^{N}\begin{pmatrix} f_n^+(l)e^{i\xi/l+it2l^2}, & 0 \\ 0, & f_n^-(l)e^{-i\eta/l-it/2l^2}\end{pmatrix}$$
$$\times\begin{pmatrix} 0, & -g_n^+(k)e^{i\eta/k+it/2k^2} \\ g_n^-(k)e^{-i\xi/k-it/2k^2}, & \rho_n(k)e^{(i\eta/k+it/2k^2}\end{pmatrix}, \tag{3.27}$$

where

$$\rho_n(k) \doteq -\sum_{m=1}^{N} g_m^+(k)\int_{-\infty}^{+\infty} dk'\,f_m^+(k')g_n^-(k'). \tag{3.28}$$

Using (3.27), it is not difficult to convert the integral equation (3.21) into the algebraic system

$$\chi_k - \sum_{n=1}^{N}\chi_n A_{nk} = h_k, \tag{3.29}$$

where

$$\chi_k(\xi,\eta) \doteq \int_{-\infty}^{+\infty} dl\,\chi^-(\xi,\eta,l)\begin{pmatrix} f_k^+(l)e^{i\xi/l+it2l^2}, & 0 \\ 0, & f_k^-(l)e^{-i\eta/l-it/2l^2}\end{pmatrix},$$
$$h_k(\xi,\eta) \doteq \begin{pmatrix} \int_{-\infty}^{+\infty} dl\,f_k^+(l)e^{i\xi/l+it/2l^2}, & 0 \\ 0, & \int_{-\infty}^{+\infty} dl\,f_k^-(l)e^{-i\eta/l-it/2l^2}\end{pmatrix}, \tag{3.30}$$

and

$$A_{nk}(\xi,\eta) \doteq \frac{1}{2\pi i}\int\int\frac{d\lambda\,dl}{l-\lambda}$$
$$\times\begin{pmatrix} 0, & -g_n^+(l)e^{i\eta/l+i\eta/2l^2} \\ g_n^-(l)e^{-i\xi/l-it/2l^2}, & \rho_n(l)e^{i\eta/l+it/2e^2}\end{pmatrix}$$
$$\times\begin{pmatrix} f_k^+(\lambda)e^{i\xi/\lambda+it/2\lambda^2}, & 0 \\ 0, & f_k^-(\lambda)e^{-i\eta/\lambda-it/2\lambda^2}\end{pmatrix}. \tag{3.31}$$

Solving the system (3.29) with h_k and A_{nk} given by (3.30) and (3.31) with respect to χ_k, we find $\chi(\xi,\eta,\lambda=0)$:

$$g(\xi,\eta)\doteq\chi(\xi,\eta,\lambda=0)$$
$$=1+\sum_{n=1}^{N}\chi_n(\xi,\eta)\frac{1}{2\pi i}\int\frac{dl}{l}$$
$$\times\begin{pmatrix}0, & -g_n^+(l)e^{i\eta/l+it/2l^2}\\ g_n^-(l)e^{-i\xi/l-it/2l^2}, & \rho_n(l)l^{\cdots+it/2l^2}\end{pmatrix}\cdot\quad(3.32)$$

Finally, the formula (3.22) gives us the exact solution of the Ish-II equation that contains $4N$ arbitrary functions $f_n^+(\lambda), g_n^\pm(\lambda), n=1,\ldots,N$ of λ one variable.

The simplest solution of this type is ($n=1$):

$$S(x,y,t)=-\tfrac{1}{2}\operatorname{tr}(\bar{\sigma}g\sigma_3g^{-1}),\quad(3.33)$$

where

$$g(x,y,t)=1+h(1-A)^{-1}f\quad(3.34)$$

and

$$h(\xi,\eta,t)=\begin{pmatrix}\int_{-\infty}^{+\infty}dlf^+(l)e^{i\xi/l+it/2l^2}, & 0\\ 0, & \int_{-\infty}^{+\infty}dlf^-(l)e^{-i\eta/l-it/2l^2}\end{pmatrix},$$

$$f(\xi,\eta,t)=\begin{pmatrix}0, & -\int_{-\infty}^{+\infty}\frac{dk}{2\pi ik}g^+(k)e^{i\eta/k+it/2k^2}\\ \int_{-\infty}^{+\infty}\frac{dk}{2\pi ik}g^-(k)e^{-i\xi/k-it/2k^2}, & -\int_{-\infty}^{+\infty}\frac{dk}{2\pi ik}g^+(k)\int_{-\infty}^{+\infty}dk'f^+(k')g^-(k')l^{i\eta/k+it/2k^2}\end{pmatrix},$$

$$A=\begin{pmatrix}0, & A_{12}\\ A_{21}, & A_{22}\end{pmatrix},$$

where

$$A_{12}(\eta,t)=\frac{-1}{2\pi i}\int\int\frac{d\lambda\,dl}{l-\lambda}g^+(l)f^-(\lambda)$$
$$\times\exp\left[i\eta\left(\frac{1}{l}-\frac{1}{\lambda}\right)+\frac{it}{2}\left(\frac{1}{l^2}-\frac{l}{\lambda^2}\right)\right],$$
$$A_{21}(\xi,t)=\frac{1}{2\pi i}\int\int\frac{d\lambda\,dl}{l-\lambda}g^-(l)f^+(\lambda)$$
$$\times\exp\left[i\xi\left(\frac{1}{\lambda}-\frac{1}{l}\right)+\frac{it}{2}\left(\frac{1}{\lambda^2}-\frac{1}{l^2}\right)\right],$$
$$A_{22}(\eta,t)=\frac{-1}{2\pi i}\int\int\frac{d\lambda\,dl}{l-\lambda}f^-(\lambda)g^+(l)$$
$$\times\int_{-\infty}^{+\infty}dk'f^+(k')g^-(k')$$
$$\cdot\exp\left[i\eta\left(\frac{1}{l}-\frac{1}{\lambda}\right)+\frac{it}{2}\left(\frac{1}{l^2}-\frac{1}{\lambda^2}\right)\right]$$

and $f^\pm(\lambda), g^\pm(l)$ are arbitrary functions.

IV. CONCLUSION

The spectral parameter λ can be introduced into the problem (2.1) in a more usual way, namely, by the transition to the function

$$\tilde{\chi}(z,\bar{z},\lambda)=\psi(z,\bar{z})\begin{pmatrix}e^{-i\lambda z}, & 0\\ 0, & e^{i\lambda\bar{z}}\end{pmatrix}.\quad(4.1)$$

The function $\tilde{\chi}$ obeys the equation

$$\begin{pmatrix}\partial_{\bar{z}} & 0\\ 0 & \partial_z\end{pmatrix}\tilde{\chi}-\frac{i\lambda}{2}[\sigma_3,\tilde{\chi}]$$
$$+\tfrac{1}{2}(p+\sigma_3)(\partial_z-\partial_{\bar{z}}+i\lambda)\tilde{\chi}=0.\quad(4.2)$$

Equation (4.2) is, of course, nothing but Eq. (2.3) with the change $\lambda\to1/\lambda$. But now the solutions of Eq. (4.2) cannot be canonically normalized. Indeed, it follows from (4.2) that at $\lambda\to\infty$ one has

$$\tilde{\chi}(z,\bar{z},\lambda)=g(z,\bar{z})+\lambda^{-1}\chi_1+\cdots,\quad(4.3)$$

where

$$P(z,\bar{z},t)=-g\sigma_3g^{-1}.\quad(4.4)$$

The spectral problem (4.2) is, obviously, equivalent to the integral equation of the type (2.4) with the Green's function $\tilde{G}$ which is given by (2.8) with the change $\lambda\to1/\lambda$. The corresponding $\bar{\partial}$ equation is of the form (2.16) with the substitution $\lambda\to1/\lambda$ in the rhs of (2.17).

The problem will appear when one tries to proceed from the linear $\bar{\partial}$ equation (2.16) to the linear integral equation, using the formula (2.20). The corresponding integral equation will be of the form (2.21) where one should substitute the unit by $g(z,\bar{z},t)$. The appearance of the function $g(z,\bar{z},t)$, which depends on the potential ($P(z,\bar{z},t)$, instead of the unit, produces the solvability problem.

This obstacle can be easily removed by the transition in the corresponding $\bar{\partial}$ equation from the function $\tilde{\chi}$ to a new function $\mu(z,\bar{z},\lambda)\doteq g^{-1}\tilde{\chi}(z,\bar{z},\lambda)$. This function has a canonical normalization $\mu\underset{\lambda\to\infty}{\longrightarrow}1$ and the corresponding integral equation for μ is exactly Eq. (2.21). The inverse problem equations are of the form (2.21) and (2.23) while the reconstruction formula (4.4) is where

$$g(z,\bar{z},t)=(\mu(z,\bar{z},t;\lambda=0))^{-1}.\quad(4.5)$$

In a similar manner one can consider the Ish-II equation also.

So, the two ways of introduction of the spectral parameters [i.e., (2.2) and (4.1)] lead to the equivalent results. The first way (2.2) is more preferable from the point of view of the absence of the additional step of the transition to the intermediate function μ. The second way (4.1) is more suit-

able for the discussion of the gauge equivalence of the Ishimori and Davey–Stewartson equations.[10,22,23]

This paper is the first one devoted to the Ishimori equation. In the second part we will discuss the exact solutions of the Ishimori equation that can be constructed by the use of the nonlocal $\bar{\partial}$ problem and the recursion and algebraic properties of the Ishimori equation.

ACKNOWLEDGMENTS

The authors are thankful to Professor R. Beals for sending the preprints of Refs. 21 and 23. One of the authors (B.G.K.) is very grateful to Service de Physique Theorique and, especially, to Professor H. Cornille for the kind hospitality.

APPENDIX A

Here, we will derive relation (2.22), i.e.,

$$\lim_{\lambda\to\lambda_i}\left(\chi-\frac{\phi_i}{\lambda-\lambda_i}\right)=\phi_i\begin{pmatrix}\frac{-iz}{\lambda_i^2}+\gamma_{1i} & 0\\ 0 & \frac{i\bar{z}}{\lambda_i^2}+\gamma_{2i}\end{pmatrix}, \tag{A1}$$

where γ_i are some constants and ϕ_i are normalized by the condition

$$\lim_{z\to\infty}\phi_i\begin{pmatrix}-\frac{iz}{\lambda_i^2} & 0\\ 0 & \frac{i\bar{z}}{\lambda_i^2}\end{pmatrix}=1.$$

To do this we first introduce the function $\tilde{\chi}$:

$$\tilde{\chi}(z,\bar{z};\lambda,\bar{\lambda})\doteq\chi(z,\bar{z};\lambda,\bar{\lambda})-\phi_i(z,\bar{z})/(\lambda-\lambda_i). \tag{A2}$$

In virtue of Eq. (2.4), the function $\tilde{\chi}$ obeys the integral equation

$$\left[1+\frac{1}{2}G(\cdot,\cdot,\lambda,\bar{\lambda})Q\left(\partial'-\bar{\partial}'+\frac{i}{\lambda}\right)\right]\tilde{\chi}(\cdot,\cdot,\lambda,\bar{\lambda})$$
$$=1-\frac{1}{\lambda-\lambda_i}\left[1+\frac{1}{2}G(\cdot,\cdot,\lambda,\bar{\lambda})\times Q\left(\partial'-\bar{\partial}'+\frac{i}{\lambda}\right)\right]\phi_i, \tag{A3}$$

where the operator G is given by (2.8). Proceeding to the $\lim\lambda\to\lambda_i$ in (A3), one obtains the equation

$$\left[1+\frac{1}{2}G(\cdot,\cdot,\lambda_i,\bar{\lambda}_i)Q\left(\partial'-\bar{\partial}'+\frac{i}{\lambda_i}\right)\right]\tilde{\tilde{\chi}}(\cdot,\cdot,\lambda_i)$$
$$=1-\frac{1}{2}\left(\frac{\partial}{\partial\lambda}G(\cdot,\cdot,\lambda,\bar{\lambda})\times Q\left(\partial'-\bar{\partial}'+\frac{i}{\lambda}\right)\phi_i\right)\Big|_{\lambda=\lambda_i}$$
$$=1+\frac{1}{2\lambda_i^2}G(\cdot,\cdot,\lambda_i,\bar{\lambda}_i)Q\phi_i-\left(\frac{\partial G(\cdot,\cdot,\lambda,\bar{\lambda})}{\partial\lambda}\right)_{\lambda=\lambda_i}Q\left(\partial'-\bar{\partial}'+\frac{i}{\lambda_i}\right)\phi_i, \tag{A4}$$

where $\tilde{\tilde{\chi}}(z,\bar{z},\lambda_i)=\lim_{\lambda\to\lambda_i}\tilde{\chi}(z,\bar{z},\lambda,\bar{\lambda})$.

Using the formula (2.8) and performing some transformations in (A4), one gets the following equation:

$$\left[1+\frac{1}{2}G(\cdot,\cdot,\lambda_i,\bar{\lambda}_i)Q\left(\partial'-\bar{\partial}'+\frac{i}{\lambda_i}\right)\right]\times\left(\tilde{\tilde{\chi}}(\cdot,\cdot,\lambda_i)-\phi_i\begin{pmatrix}-\frac{iz}{\lambda_i^2} & 0\\ 0 & \frac{i\bar{z}}{\lambda_i^2}\end{pmatrix}\right)$$
$$=1-\frac{1}{4\pi\lambda_i^2}\sigma_3\int\int dz\wedge d\bar{z}\times\left[Q(z,\bar{z})\left(\partial_z-\partial_{\bar{z}}+\frac{i}{\lambda_i}\right)\phi_i(z,\bar{z})\right]_{\text{diag}}. \tag{A5}$$

Now, taking into account the fact that the homogeneous integral equation (2.4) has the nontrivial solution ϕ_i, one concludes from (A5) via the Fredholm alternative's theorem that

$$\tilde{\tilde{\chi}}(z,\bar{z},\lambda_i)=\phi_i\begin{pmatrix}\gamma_{1i} & 0\\ 0 & \gamma_{2i}\end{pmatrix}, \tag{A6}$$

where γ_{1i} and γ_{2i} are some constants and

$$1-\frac{1}{4\pi\lambda_i^2}\sigma_3\int\int dz\wedge d\bar{z}\left[Q(z,\bar{z})\left(\partial_z-\partial_{\bar{z}}+\frac{i}{\lambda_i}\right)\times\phi_i(z,\bar{z})\right]_{\text{diag}}=0. \tag{A7}$$

The relation (A6) is nothing but (A1).

APPENDIX B

In this appendix we present some steps of calculations that gives rise to the relations (3.14) and (3.20).

First, one has the integral equation (3.11) for the jump $\chi^+-\chi^-$. Let us assume that the jump $\chi^+-\chi^-$ is of the form (3.20) where $f(l,\lambda)$ is some function. Then let us substitute this expression (3.20) for $\chi^+-\chi^-$ into Eq. (3.11). One gets

$$\int_{-\infty}^{+\infty}dl\,\chi^-(\xi,\eta,l)\Sigma_l(\xi,\eta)f(l,k)\Sigma_k^{-1}(\xi,\eta)$$
$$=\begin{pmatrix}\int_{-\infty}^{+\infty}d\eta'(Q(\xi,\eta')\mathscr{D}_k\chi^+(k))_{11}e^{i\xi/k}; & \int_{-\infty}^{+\infty}d\eta'(Q(\xi,\eta')\mathscr{D}_k\chi^+(k))_{12}e^{-i\eta'/k}\\ \int_{-\infty}^{+\infty}d\xi'(Q(\xi',\eta)\mathscr{D}_k\chi^-(k))_{21}e^{i\xi'/k}; & \int_{-\infty}^{+\infty}d\xi'(Q(\xi',\eta)\mathscr{D}_k\chi^-(k))_{22}e^{-i\eta/k}\end{pmatrix}\frac{\Sigma_k^{-1}(\xi,\eta)}{2}$$
$$-\frac{1}{2}\int_{-\infty}^{+\infty}dl\,GQ(\cdot)\mathscr{D}_k\chi^-(l)\Sigma_l(\xi,\eta)f(l,k)\Sigma_k^{-1}(\xi,\eta). \tag{B1}$$

Then, using the identity (3.18) and Eqs. (3.10) one can transform (B1) into the following form:

$$\int_{-\infty}^{+\infty} dl\, \chi^-(l)\Sigma_l f(l,k)\Sigma_k^{-1}$$
$$= \int_{-\infty}^{+\infty} dl\, \chi^-(l)\Sigma_l (T^+(l,k) - T^-(l,k))\Sigma_k^{-1}$$
$$+ \int_{-\infty}^{+\infty} dl\, \chi^-(l)\Sigma_l$$
$$\times \int_{-\infty}^{+\infty} dk'\, T^+(l,k') f(k',k)\Sigma_k^{-1}, \qquad \text{(B2)}$$

where $T^+(l,k)$ and $T^-(l,k)$ are given by (3.15) and (3.16). Equation (B2) obviously implies Eq. (3.14).

So, the jump $(\chi^+ - \chi^-)(\lambda)$ is really given by (3.20) where the function $f(l,\lambda)$ is defined by the integral equation (3.14).

[1]V. E. Zakharov, S. V. Manakov, S. P. Novikov, and L. P. Pitaevski, *Theory of Solitons. The Inverse Problem Method* (Nauka, Moscow, 1980); English translation (Plenum, New York, 1984).
[2]M. J. Ablowitz and H. Segur, *Solitons and Inverse Scattering Transform* (SIAM, Philadelphia, 1981).
[3]F. Calogero and A. Degasperis, *Spectral Transform and Solitons* (North-Holland, Amsterdam, 1982), Vol. I.
[4]R. K. Dodd, J. C. Eilbeck, J. D. Gibbon, and H. C. Morris, *Solitons and Nonlinear Waves* (Academic, New York, 1982).
[5]A. C. Newell, *Solitons in Mathematics and Physics* (SIAM, Philadelphia, 1985).
[6]M. Lakshmanan, Phys. Lett. A **61**, 53 (1977).
[7]L. A. Takhtajan, Phys. Lett. A **64**, 235 (1977).
[8]L. A. Takhtajan and L. D. Faddeev, *Hamiltonian Approach in Soliton Theory* (Nauka, Moscow, 1986).
[9]Y. Ishimori, Prog. Theor. Phys. **72**, 33 (1984).
[10]F. W. Nijhoff, H. W. Capel, and G. L. Wiersma, Lecture Notes in Physics, Vol. 239, 263 (1985).
[11]S. V. Manakov, Physica D **3**, 420 (1981).
[12]M. J. Ablowitz, D. Bar Yaacov, and A. S. Fokas, Stud. Appl. Math. **69**, 135 (1983).
[13]A. S. Fokas and M. J. Ablowitz, Stud. Appl. Math. **69**, 211 (1983).
[14]A. S. Fokas, Phys. Rev. Lett. **51**, 3 (1983).
[15]A. S. Fokas and M. J. Ablowitz, Phys. Rev. Lett. **51**, 7 (1983).
[16]A. S. Fokas and M. J. Ablowitz, J. Math. Phys. **25**, 2505 (1984).
[17]R. Beals and R. R. Coifman, Proc. Sym. Pure Math. **43**, 45 (1985).
[18]P. G. Grinevich and S. V. Manakov, Functional Anal. Appl. **20**(2), 14 (1986).
[19]M. Boiti, J. J. P. Leon, M. Manna, and F. Pempinelli, Inverse Problems **3**, 25 (1987).
[20]A. S. Fokas and M. J. Ablowitz, Lecture Notes in Physics, Vol. 189, 137 (1983).
[21]R. Beals and R. R. Coifman, "Linear spectral problems, nonlinear equations, and the $\bar{\partial}$-method," preprint Yale Univ., 1988; Inverse Problems **5**, 89 (1989).
[22]B. G. Konopelchenko and B. T. Matkarimov, "On the inverse scattering transform for the Ishimori equation," preprint Lecce University, July 1988; Phys. Lett. A **135**, 183 (1989).
[23]R. Beals and R. R. Coifman, "The spectral problem for the Davey-Stewartson and Ishimori hierarchies," talk given at Workshop on Nonlinear Evolution Equations, Como, Italy, July 1988.
[24]V. E. Zakharov and S. V. Manakov, Functional Anal. Appl. **19**(2), 11 (1985).