Path integration over the n-dimensional Euclidean group

M. Böhm and G. Junker
Physikalisches Institut der Universität Würzburg, Am Hubland, 8700 Würzburg, Federal Republic of Germany

(Received 2 December 1987; accepted for publication 25 January 1989)

The path integral for the n-dimensional free particle is considered. According to the underlying symmetry, the short time propagator is expanded in zonal spherical functions of the Euclidean group $G = T^n \circledS \mathrm{SO}(n)$ with respect to the subgroup $H = \mathrm{SO}(n)$. The group theoretical approach to path integration, including the radial part, is explicitly demonstrated.

I. INTRODUCTION

Recently, the present authors[1] have developed a general scheme for path integration on a homogeneous space given by a group quotient $G/H, H \subset G$. For a trivial group $H = \{e\}$ the short time propagator has been expanded in group characters of G. In all other cases $H \neq \{e\}$ the group expansion has led to a decomposition of the short time propagator in zonal spherical functions. Up to now this technique has only been applied to the generalized polar coordinate path integral[1] and to the path integration on spaces with positive and negative curvature.[2] The purpose of the present paper is to include radial path integrals in this group theoretical approach. The free particle in n dimensions is considered where the Euclidean space E_n is viewed as the quotient G/H. Here G is the n-dimensional Euclidean group, which is a semidirect product of the translation group and the rotation group in n dimensions, $T^n \circledS \mathrm{SO}(n)$, and $H = \mathrm{SO}(n)$.

This paper is organized as follows. In the next section, the n-dimensional Euclidean group and its representations are discussed in some detail. The Fourier decomposition of functions $f(g)$ of $g \in G$, satisfying $f(h^{-1}gh) = f(g)$ for $h \in H$, is constructed explicitly. In Sec. III this decomposition is applied to expand the short time propagator in zonal spherical functions $D^k_{00}(g)$ of $G \supset H$. An integral representation of the free particle propagator is obtained, leading to the well-known result of Feynman.[3]

II. THE EUCLIDEAN GROUP IN n DIMENSIONS, $G = T^n \circledS \mathrm{SO}(n)$

The Euclidean group $G = T^n \circledS \mathrm{SO}(n)$ acts as a transformation group in the Euclidean space E_n of n dimensions via the map

$$g: \mathbf{a} \mapsto h\mathbf{a} + \mathbf{r}, \quad g \in G, \tag{2.1}$$

where h is an $n \times n$ matrix representation of the subgroup $H = \mathrm{SO}(n)$. The parameters of the group element $g = g(\mathbf{r},h)$ are the $n(n-1)/2$ Euler angles of h and the n coordinates of the translation vector $\mathbf{r}$ given (for convenience) in polar coordinates $(r,\varphi_1,\varphi_2,\ldots,\varphi_{n-1})$. The group composition law is

$$g(\mathbf{r}_1,h_1)g(\mathbf{r}_2,h_2) = g(\mathbf{r}_1 + h_1\mathbf{r}_2, h_1h_2). \tag{2.2}$$

A general group element may be decomposed into a translation and a rotation (see Ref. 4, p. 548)

$$g(\mathbf{r},h) = g(\mathbf{r},\mathbf{1})g(\mathbf{o},h) = g(\mathbf{o},h)g(h^{-1}\mathbf{r},\mathbf{1}), \tag{2.3}$$

where $\mathbf{1}$ stands for the $n \times n$ unit matrix and $\mathbf{o}$ is the n-dimensional null vector. Obviously any point $\mathbf{r}$ in E_n may be obtained via a translation of the origin $\mathbf{o}$,

$$g: \mathbf{o} \mapsto \mathbf{r}. \tag{2.4}$$

Accordingly we may restrict g in (2.4) to the form $g(\mathbf{r},\mathbf{1})$, as the origin is invariant under pure rotations $g(\mathbf{o},h)$. Moreover, any function $f(\mathbf{r})$ defined over E_n may be viewed as a function $f(g)$ on the group manifold of G. Especially if $f(\mathbf{r}) = f(r)$ depends only on the radial distance r, it is a function invariant under rotations $g(\mathbf{o},h)$. The zonal spherical functions $D^k_{00}(g)$ having this property are given by Bessel functions (see Ref. 4, p. 553)

$$D^k_{00}(g) = \Gamma(n/2)(2/kr)^{(n-2)/2} J_{(n-2)/2}(kr), \tag{2.5}$$

where r is the radial polar coordinate of the translation vector $\mathbf{r}$ in $g(\mathbf{r},h)$. The basis states of G are usually labeled by k, l, and M corresponding to the conserved energy ($E = \hbar^2 k^2/2m$), angular momentum, and its degeneracy, respectively.

For a translation by r along a fixed axis $\mathbf{a}$, e.g., the unit vector in x_n-direction, the associate zonal spherical function reads (see Ref. 4, p. 554)

$$\begin{aligned} D^k_{L0}(g(r\mathbf{a},\mathbf{1})) &= i^l \Gamma(n/2)\left[(2l+n-2)\frac{\Gamma(l+n-2)}{l!\,\Gamma(n-1)}\right]^{1/2} \\ &\quad \times \left(\frac{2}{kr}\right)^{(n-2)/2} J_{l+(n-2)/2}(kr), \end{aligned} \tag{2.6}$$

where L stands for the $(n-1)$-tuple $L = (l,0,\ldots,0)$, with $l = 0,1,2,\ldots$. Note that any $\mathbf{r}$ may be obtained from $r\mathbf{a}$ through a pure rotation $h \in \mathrm{SO}(n)$, $\mathbf{r} = h(r\mathbf{a})$. (See Ref. 1.)

As is known, a function $f(g)$ invariant under a rotation $g(\mathbf{o},h)$ may be expanded in zonal spherical functions:

$$f(g) = \int_0^\infty dk\, F(k) d_k D^k_{00}(g), \tag{2.7}$$

where

$$F(k) = \int_G dg\, f(g) D^{k*}_{00}(g). \tag{2.8}$$

In the above, dg is the invariant volume element of G given by

$$dg = d\mathbf{r}\, dh, \tag{2.9}$$

where dh is the normalized invariant Haar measure of

 0022-2488/89/061195-03$02.50

$H = \mathrm{SO}(n)$, $\int_H dh = 1$ and $d\mathbf{r}$ is the usual Euclidean measure. The "dimension" d_k is defined by[1]

$$\frac{\delta(k-k')}{d_k} = \int_G dg\, D^k_{00}(g) D^{k'}_{00}(g). \tag{2.10}$$

Using the explicit form (2.5) we find

$$\frac{\delta(k-k')}{d_k} = \Gamma\left(\frac{n}{2}\right) \frac{2^{n-1}\pi^{n/2}}{(kk')^{(n-2)/2}} \times \int_0^\infty dr\, r J_{(n-2)/2}(kr) J_{(n-2)/2}(k'r). \tag{2.11}$$

Comparison with the closure relation of Bessel's function[5]

$$\int_0^\infty dr\, r J_\nu(kr) J_\nu(k'r) = \frac{1}{k}\delta(k-k') \tag{2.12}$$

leads to the identification

$$d_k = k^{n-1}/[2^{n-1}\pi^{n/2}\Gamma(n/2)]. \tag{2.13}$$

For $n = 2$, d_k agrees with Barut and Raczka[6] who discuss the harmonic analysis of $T^2 \otimes \mathrm{SO}(2)$.

Finally we would like to mention that for $\tilde{f}(r) = f(g) r^{(n-1)/2}(2\pi)^{n/2}$ and $\tilde{F}(k) = F(k) k^{(n-1)/2}$ the transformation (2.7) leads to the Hankel transformation of order $\nu = (n-2)/2$:

$$\begin{aligned} \tilde{f}(r) &= \int_0^\infty dk\, \tilde{F}(k) J_\nu(kr)\sqrt{kr}, \\ \tilde{F}(k) &= \int_0^\infty dr\, \tilde{f}(r) J_\nu(kr)\sqrt{kr}. \end{aligned} \tag{2.14}$$

III. PATH INTEGRATION OVER *G*

As an application of the above group expansion we consider the Feynman propagator of an n-dimensional free particle given in the sliced time basis

$$K(\mathbf{r}_b,\mathbf{r}_a;T) = \lim_{N\to\infty} \int \prod_{j=1}^{N} \left[\left(\frac{m}{2\pi i\hbar\epsilon}\right)^{n/2} \exp\left\{\left(\frac{i}{\hbar}\right) S_j\right\}\right] \times \prod_{j=1}^{N-1} d\mathbf{r}_j, \tag{3.1}$$

where the short time action is given by

$$S_j = (m/2\epsilon)(\Delta\mathbf{r}_j)^2. \tag{3.2}$$

Here we have adopted the usual notation $\Delta\mathbf{r}_j = \mathbf{r}_j - \mathbf{r}_{j-1}$, $\mathbf{r}_{a=} \mathbf{r}_0$, $\mathbf{r}_b = \mathbf{r}_N$ and an isometric time slicing $\epsilon N = t_b - t_a = T$.

Let us consider the group element $g_j = g(\mathbf{r}_j,\mathbf{1})$. Obviously the origin is mapped onto $\mathbf{r}_j$ via the translation g_j. The combination,

$$g_{j-1}^{-1} g_j = g^{-1}(\mathbf{r}_{j-1},\mathbf{1}) g(\mathbf{r}_j,\mathbf{1}) = g(\mathbf{r}_j - \mathbf{r}_{j-1},\mathbf{1}), \tag{3.3}$$

is just the translation mapping $\mathbf{o}$ onto $\Delta\mathbf{r}_j$. Therefore the short time propagator in (3.1) may be considered as a function $f(g_{j-1}^{-1} g_j)$ on G which depends only on the parameter $r = |\Delta\mathbf{r}_j|$ of the group element (3.3). Hence the Fourier decomposition (2.7) may be applied to $\exp\{(i/\hbar)S_j\}$, where the coefficient (2.8) is given by ($z = m/2i\hbar\epsilon$),

$$F(k) = 2\pi^{n/2}(2/k)^{(n-2)/2} \times \int_0^\infty dr\, r^{n/2} e^{-zr^2} J_{(n-2)/2}(kr). \tag{3.4}$$

In (3.4) the integration over the subgroup H and the group parameters $(\varphi_1,\ldots,\varphi_{n-1})$ of the translation vector $\mathbf{r}$ has been performed. Using the integral formula[7]

$$\begin{aligned} &\int_0^\infty dx\, x^{\nu+1} e^{-\alpha x^2} J_\nu(\beta x) \\ &\quad = \beta^\nu (2\alpha)^{-\nu-1} \exp\{-\beta^2/4\alpha\}, \\ &\quad \mathrm{Re}\,\alpha > 0, \beta > 0, \mathrm{Re}\,\nu > -1, \end{aligned} \tag{3.5}$$

we find

$$F(k) = (\pi/z)^{n/2} \exp\{-k^2/4z\}. \tag{3.6}$$

To be more explicit we have derived the decomposition

$$\left(\frac{m}{2\pi i\hbar\epsilon}\right)^{n/2} \exp\left\{\frac{im}{\hbar 2\epsilon}|\Delta\mathbf{r}_j|^2\right\} = \int_0^\infty dk \exp\left\{-\frac{i\hbar k^2\epsilon}{2m}\right\} d_k D^k_{00}(g_{j-1}^{-1} g_j). \tag{3.7}$$

With the aid of the orthogonality relation,

$$\int d\mathbf{r}_j\, D^k_{00}(g_{j-1}^{-1} g_j) D^{k'*}_{00}(g_j^{-1} g_{j+1}) = \frac{\delta(k-k')}{d_k} D^k_{00}(g_{j-1}^{-1} g_{j+1}), \tag{3.8}$$

the path integration can be performed leading to the following integral representation of the free particle propagator:

$$K(\mathbf{r}_b,\mathbf{r}_a;T) = \int_0^\infty dk \exp\left\{-\frac{i}{\hbar}\frac{\hbar^2 k^2}{2m} T\right\} d_k D^k_{00}(g_a^{-1} g_b). \tag{3.9}$$

The energy spectrum may be identified to be $E_k = \hbar^2 k^2/2m$. In order to obtain the normalized wave functions we make use of the group property

$$D^k_{00}(g_a^{-1} g_b) = \sum_L D^k_{L0}(g_b) D^{k*}_{L0}(g_a). \tag{3.10}$$

As $\mathbf{r} = h(r\mathbf{a})$, it follows from (2.3) that $g = g(\mathbf{r},\mathbf{1}) = g(\mathbf{o},h) g(r\mathbf{a},\mathbf{1}) g(\mathbf{o},h^{-1})$ and the associate spherical functions

$$D^k_{L0}(g(\mathbf{o},h) g(r\mathbf{a},\mathbf{1}) g(\mathbf{o},h^{-1})) = D^k_{L0}(g(\mathbf{o},h) g(r\mathbf{a},\mathbf{1}))$$

decompose into

$$D^k_{L0}(g) = \sum_{L'} D^k_{LL'}(g(\mathbf{o},h)) D^k_{L'0}(g(r\mathbf{a},\mathbf{1})). \tag{3.11}$$

Note that the sum vanishes unless L' is of the form $L' = (l,0,\ldots,0)$ (see Ref. 4, p. 555) and $D^k_{LL'}(g(\mathbf{o},h))$ reduces to the associate spherical functions of $\mathrm{SO}(n)$, $D^k_{LL'}(g(\mathbf{o},h)) = d^l_{M0}(h)$, given in Ref. 1. Collecting everything, the propagator (3.9) is rewritten as

$$K(\mathbf{r}_b,\mathbf{r}_a;T) = \int_0^\infty dk \exp\left\{-\left(\frac{i}{\hbar}\right) E_k T\right\} \times \sum_{l,M} \Psi_{klM}(\mathbf{r}_b) \Psi^*_{klM}(\mathbf{r}_a) \tag{3.12}$$

with

$$\Psi_{KlM}(\mathbf{r}) = \sqrt{d_k}\, D^{k}_{L',0}(g(r\mathbf{a},\mathbf{1}))d^{l}_{M0}(h)$$
$$= i^{l}(k/r^{n-2})^{1/2} J_{l+(n-2)/2}(kr)$$
$$\times\sqrt{\Gamma(n/2)/2\pi^{n/2}}\, Y_{lM}(\mathbf{e}). \tag{3.13}$$

In the last step, we have used Eq. (2.6). $Y_{lM}(\mathbf{e})$ are the hyperspherical harmonics in n dimensions.[1] The integers m_i of the set $M=(m_1,\ldots,m_{n-2})$ are related by $l\geqslant m_1\geqslant m_2\geqslant\cdots\geqslant m_{n-3}\geqslant|m_{n-2}|\geqslant 0$. With Eq. (2.12) the normalization

$$\int d\mathbf{r}\, \Psi_{klM}(\mathbf{r})\Psi^{*}_{K'l'M'}(\mathbf{r}) = \delta(k-k')\delta_{ll'}\delta_{MM'} \tag{3.14}$$

is shown immediately.

Performing the integration in (3.12) by using formula #6.6332 of Ref. 7, we obtain

$$K(\mathbf{r}_b,\mathbf{r}_a;T)$$
$$= \frac{m}{i\hbar T}(r_a r_b)^{(2-n)/2}\exp\left\{\frac{im}{2\hbar T}(r_b^2+r_a^2)\right\}$$
$$\times\sum_{l=0}^{\infty} I_{l+(n-2)/2}\left(\frac{mr_a r_b}{i\hbar T}\right)$$
$$\times\sum_{M}\frac{\Gamma(n/2)}{2\pi^{n/2}}\, Y_{lM}(\mathbf{e}_b)Y^{*}_{lM}(\mathbf{e}_a). \tag{3.15}$$

For $n=3$, (3.15) reduces to the result of Peak and Inomata.[8]

Finally we would like to mention that the k integration can be directly performed in Eq. (3.9) via (3.5), leading to the original result of Feynman,[3]

$$K(\mathbf{r}_b,\mathbf{r}_a;T) = \left(\frac{m}{2\pi i\hbar T}\right)^{n/2}\exp\left\{\frac{im}{2\hbar T}|\mathbf{r}_b-\mathbf{r}_a|^2\right\}. \tag{3.16}$$

IV. DISCUSSION

In the present work we have applied the expansion in zonal spherical functions, developed in Ref. 1, to the path integration over the Euclidean group in n dimensions. The technique has been explicitly demonstrated for the n-dimensional free particle. Our result for the free particle coincides with that obtained via the Gaussian path integration, as expected. However, in the present approach the application of group theoretical methods has been extended to include the radial path integration. Until now only the angular path integration over rotation groups had been considered. Now we may conclude that the complete path integral treatment can be incorporated in the formalism of Ref. 1. Here the Euclidean group has been considered. However, the same technique may be applied, for example, to the path integration over the pseudo-Euclidean group $T^n \circledS \mathrm{SO}(n-1,1)$ in n dimensions.

ACKNOWLEDGMENTS

The authors are grateful to the referee for his comments.

One of us (G. J.) would like to thank the Studienstiftung des Deutschen Volkes for support.

[1] M. Böhm and G. Junker, J. Math. Phys. **28**, 1978 (1987).

[2] M. Böhm and G. Junker, "Group theoretical approach to path integration on spheres," in *Path Summation: Achievements and Goals*, edited by S. Lundqvist, A. Ranfagni, V. Sa-yakanit, and L. S. Schulman (World Scientific, Singapore, 1988).

[3] R. P. Feynman, Rev. Mod. Phys. **20**, 367 (1948); R. P. Feynman and H. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965).

[4] N. J. Vilenkin, *Special Functions and the Theory of Group Representations* (Am. Math. Soc., Providence, RI, 1968).

[5] G. Arfken, *Mathematical Methods for Physicists* (Academic, New York, 1970), 2nd ed., p. 495.

[6] A. O. Barut and R. Raczka, *Theory of Group Representation and Applications* (Polish Scientific, Warsaw, 1980), 2nd ed., p. 435.

[7] I. S. Gradsteyn and I. M. Rhyzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1969), p. 717f, formulas #6.6314 and #6.6332.

[8] D. Peak and A. Inomata, J. Math. Phys. **10**, 1422 (1969).

The Riemann–Hilbert transformation for an approach to a representation of the Virasoro group

Wei Li[a)]
Department of Mathematics and Computer Science, Clarkson University, Potsdam, New York 13676

Bo-yu Hou
Center of the Theoretical Physics, CCAST, and Institute of Modern Physics, Northwest University, Xian, China

(Received 18 May 1988; accepted for publication 25 January 1989)

In this paper, it is the intent to apply the Riemann–Hilbert transformation developed by Hauser and Ernst [J. Math. Phys. **21**, 1126, 1418 (1980)] in providing a new representation of the Virasoro group. It is found that the Geroch group that acts on the solution space of the Einstein field equations is extended to the semidirect product of the Virasoro and Kac–Moody groups; also, the relationship between the infinitesimal transformation given previously [B. Y. Hou and W. Li, Lett. Math. Phys. **13**, 1 (1987); J. Phys. A **20**, L897 (1987); W. Li, Phys. Lett. A **129**, 301 (1988)] and the infinitesimal Riemann–Hilbert transformation is pointed out. Finally, it is shown that the well-known Neugebauer–Backlund transformation can be derived from the Riemann–Hilbert transformation.

I. INTRODUCTION

Several years ago Hauser and Ernst[1] first pointed out that the Riemann–Hilbert transformation is associated with the infinite-dimensional loop group: They were able to give the explicit action of the Geroch group,[2] which is shown to be isomorphic to an affine Kac–Moody group,[3] in the two-dimensional Einstein field equations, the Ernst equation. The work of Hauser and Ernst developed the result of Kinnersley and Chitre,[4] who gave an infinite set of generators of the Geroch group. The Hauser–Ernst method can also offer an effective and powerful technique for generating a new solution of the Ernst equation from the known solution. Later, Ueno and Nakamura[5] applied the Hauser–Ernst method to the principal chiral model and the self-dual Yang–Mills fields and established similar representations of the Kac–Moody groups in these systems. Moreover, Ueno and Nakamura pointed out the link between the Kac–Moody algebra found by Dolan[6] and the so-called hidden symmetry transformations.[7] Thus far much investigation and application has been made toward understanding the Hauser–Ernst method as related to the Riemann–Hilbert transformation for some integrable systems.[8]

In this paper we shall find a way to extend the Hauser–Ernst method to the more general case. We shall indicate that the Riemann–Hilbert transformation for the Hauser–Ernst approach can be related with the Virasoro group as well as the Kac–Moody group and has a richer structure than previously expected. The Riemann–Hilbert transformation gives rise to the construction of the semidirect product of the Kac–Moody and Virasoro groups.

In a series of recent papers,[9,10] the present authors have succeeded in constructing an infinite set of infinitesimal transformations for the Ernst equation. Our transformations are different from those given by Kinnersley and Chitre.[4] Careful calculation shows that these new transformations constitute a representation of the Virasoro algebra which has no central extension and no highest weight, so that the representation is nonunitary. As a result, a new symmetry, like the Kac–Moody symmetry, is confirmed to exist in the solution space of the Ernst equation and the Geroch group is thus extended by the Virasoro and Kac–Moody groups.

We are motivated to find the exponentiation of our infinitesimal transformations and give the representation of the enlarged Geroch group. The problem can be solved by giving an integral equation which will be proved by means of the Riemann–Hilbert transformation in Sec. III; it is given by

$$\frac{1}{2\pi i}\int_{C_{0,t}} \frac{F(s)u(s)F_0(v(s))^{-1}}{s(s-t)}\,ds = 0, \tag{1.1}$$

where C represents a circle surrounding the origin and t in the complex s plane; $u(s)$ and $v(s)$ are, respectively, a 2×2 matrix function and a scalar function of s; and $F(s)$ and $F_0(s)$ satisfy the Hauser–Ernst linearization equations. The details of the restrictions to the quantities in (1.1) will be given in Sec. III.

We see that if $v(s) = s$, the integral equation (1.1) is identical with that initially given by Hauser and Ernst[1] and is used to provide the representation of the Kac–Moody group. If $u(t) = I$ (where I is a unit matrix), the successive transform of $F(t)$ for Eq. (1.1) is offered by

$$v_2 = v_0 \circ v_1 \quad \text{or} \quad v_2(t) = v_0(v_1(t)). \tag{1.2}$$

We know that according to the representation theory of an infinite-dimensional group,[11] Eq. (1.2) is the composition law of the Virasoro group, for which elements consisting of the set of functions $v(t)$ satisfy the conditions given in Sec. III. It is apparent that our work generalizes the application of the Riemann–Hilbert transformation and presents a new

[a)] Permanent address: Institute of Modern Physics, Northwest University, Xian, China.

 0022-2488/89/061198-07$02.50

approach for generating solutions of the Ernst equation.

The structure of this paper is organized as follows. In Sec. II we recall the formulation developed by Hauser and Ernst[1] for the Ernst equation which will be used in the following discussions. In Sec. III, we shall describe how to generalize the Hauser and Ernst approach to the Riemann–Hilbert transformation. Then we shall exploit the Riemann–Hilbert transformation to prove the integral equation (1.1). In Sec. IV, we derive the infinitesimal transformations given in (1.1) and identify the infinitesimal Riemann–Hilbert transformations with those of the Virasoro algebra. Finally, we shall discuss the group structure of the Riemann–Hilbert transformation and apply it to rederive some known Backlund transformations such as the Neugebauer[12] and Maison–Cosgrove transformations.[13]

II. NOTATIONS AND CONVENTIONS

In order to describe our objective more clearly, it is helpful to introduce a few of the notations and conventions that shall often be used in this paper.

We first start with the metric of the space-time, which admits two commuting Killing vectors under the line element

$$ds^2 = g_{ij}\,dx^i\,dx^j + g_{+-}\,dx_+\,dx_-, \tag{2.1}$$

where g_{ij} $(i,j=1,2)$ and g_{+-} are functions of x_+ and defined by

$$x_+ = \tfrac{1}{2}(x^3 + \Lambda x^4), \quad x_- = \tfrac{1}{2}(x^3 - \Lambda x^4). \tag{2.2}$$

Here we have two cases to be distinguished for the space-time: If one of the Killing vectors is timelike, i.e., the space-time possesses the stationary and axially symmetric fields, we take the value of Λ as $-i$; if both Killing vectors are spacelike, i.e., the space-time has the cylindrically symmetric fields or the gravitational plane-wave fields, we take Λ as 1. The treatments in the following discussion are very similar for these two cases; thus we no longer underline their differences.

The reduction of the vacuum Einstein field equations leads to

$$\partial_+(\alpha^{-1}g\Omega\,\partial_- g) + \partial_-(\alpha^{-1}g\Omega\,\partial_+ g) = 0, \tag{2.3}$$

where

$$\partial_+ = \frac{\partial}{\partial x_+}, \quad \partial_- = \frac{\partial}{\partial x_-}, \qquad \Omega = \begin{pmatrix} 0 & i \\ -i & 0 \end{pmatrix}, \tag{2.4}$$

and g is the 2×2 symmetric real matrix whose elements g_{ij} are the metric components in the line element (2.1) satisfying

$$\det g = (\Lambda\alpha)^2. \tag{2.5}$$

We will not consider the remainder of the vacuum Einstein field equations governing the metric component g_{+-} in this paper.

We then introduce the matrix Ernst potential E, which is a 2×2 matrix field and may be defined as a solution of

$$\begin{aligned} 2(\beta + \Lambda^3\alpha)\partial_+ E &= (E + E^\dagger)\Omega\,\partial_+ E, \\ 2(\beta - \Lambda^3\alpha)\partial_- E &= (E + E^\dagger)\Omega\,\partial_- E \end{aligned} \tag{2.6}$$

such that

$$g = \tfrac{1}{2}(E + E^\dagger) - \beta\Omega, \tag{2.7}$$

$$2\beta = \operatorname{tr}(E\Omega), \tag{2.8}$$

where the dagger stands for the Hermitian conjugation. It is not difficult to show that Eq. (2.3) is equivalent to Eqs. (2.6).

According to the definition of the fields α and β from Eq. (2.3), it is apparent that α and β are solutions of the wave equation in two dimensions and have the relation

$$\partial_+\beta = +\Lambda^3\,\partial_+\alpha, \quad \partial_-\beta = -\Lambda^3\,\partial_-\alpha. \tag{2.9}$$

By defining

$$\eta_+ = \beta + \Lambda^3\alpha, \quad \eta_- = \beta - \Lambda^3\alpha, \tag{2.10}$$

Eq. (2.9) implies that

$$\partial_+\eta_- = \partial_-\eta_+ = 0. \tag{2.11}$$

Following Hauser and Ernst's treatment for the linearization of Eq. (2.6), we define a set of 2×2 matrix functions F of x_+ and x_- and a complex parameter t such that for a given E, $F(t) = F(x_+, x_-; t)$ is any solution of the linearization equations

$$\begin{aligned} \partial_+ F(t) &= [t/(1 - 2t\eta_+)]\,\partial_+ E\,\Omega F(t), \\ \partial_- F(t) &= [t/(1 - 2t\eta_-)]\,\partial_- E\,\Omega F(t). \end{aligned} \tag{2.12}$$

In the sense of Frobenius the integrable condition of Eqs. (2.12) has to be identical to that of Eq. (2.6). We know that the linearization equations do not define $F(t)$ uniquely; thus we need to suppress some subsidiary conditions consistent with Eqs. (2.12) such that $F(t)$ is holomorphic in a neighborhood $t = 0$ and $F(t)$ satisfies

$$F(0) = I, \tag{2.13}$$

$$\dot{F}(0) = E\Omega, \tag{2.14}$$

$$\det F(t) = \lambda^{-1}(t), \tag{2.15}$$

$$F(t)^\dagger\Omega A(t) F(t) = \Omega, \tag{2.16}$$

where

$$\dot{F}(t) = \frac{\partial F(t)}{\partial t}, \qquad F(t)^\dagger = \text{Hermitian conjugate of } F(t^*) \tag{2.17}$$

and

$$\begin{aligned} \lambda(t) &= [(1 - 2t\eta_+)(1 - 2t\eta_-)]^{1/2}, \\ A(t) &= I - t(E + E^\dagger)\Omega. \end{aligned} \tag{2.18}$$

[Also, I denotes the 2×2 unit matrix and the asterisk denotes the complex conjugation.]

On the other hand, Eqs. (2.6) can be written in the form

$$A(t)[t/(1 - 2t\eta_\pm)]\partial_\pm E = t\,\partial_\pm E. \tag{2.19}$$

We then operate $A(t)$ on the linearization equations (2.12) and use Eq. (2.19) to obtain

$$A(t)\partial_\pm F(t) = t\,\partial_\pm E\,\Omega F(t), \tag{2.20}$$

which will be used in the following discussion.

Except for these restrictions on $F(t)$ there still exists the general gauge transformation of the function $F(t)$, i.e., for given any solution $F(t)$,

$$F'(t) = F(t)u(t) \tag{2.21}$$

is a solution of the linearization equations (2.12) also, where $u(t)$ is any 2×2 matrix function of t only and $u(t)$ is holomorphic in a neighborhood $t=0$ satisfying

$$u(0)=I,\qquad u(t)^{\dagger}\Omega u(t)=\Omega,\quad \det u(t)=1. \tag{2.22}$$

Since we know that the analytic properties of the function $F(t)$ play an important role in our discussion we thus hope to restrict the gauge to one for which the set of t-plane singularities of the function $F(t)$ is minimized. Hauser and Ernst[1,14] proved that for fixed (x_+, x_-) the function $F(t)$ has t-plane singularities at $t=1/2\eta_+$ and $t=1/2\eta_-$ regardless of the choice of gauge: For this reason we can always choose the gauge such that for fixed (x_+, x_-) $F(t)$ is a holomorphic function of t on the whole t plane except for the points $t=1/2\eta_+$ and $t=1/2\eta_-$.

III. PROOF OF THE INTEGRABLE EQUATION

Before describing the proof of the integrable equation (1.1) given in Sec. I, we would like to briefly recall the formulation of the Riemann–Hilbert transformation, which is an essential tool to our discussion.

First, let us select a circle C surrounding the origin in the complex t plane, in which the interior and exterior regions of C are, respectively, denoted by C_+ and C_-. Then there exist a pair of functions $X_+(t)$ and $X_-(t)$ of x_+, x_- and a parameter t such that $X_+(t)$ and $X_-(t)$ are holomorphic in C_+ and C_-, respectively, and both are continuous on C. If a given function $G(t)$ on C, called the kernel, is analytic and connects $X_+(t)$ with $X_-(t)$ by a relation

$$X_-(t)=X_+(t)G(t)\ \text{ on }\ C, \tag{3.1}$$

the solutions $X_+(t)$ and $X_-(t)$ to Eq. (3.1) are unique: This is the so-called Riemann–Hilbert transformation. It is well known that the Riemann–Hilbert transformation can be used in connection with the generation of a new solution to some nonlinear equations from prior solutions once the kernel $G(t)$ is explicitly given.

The key to the problem is how to find an explicit expression of $G(t)$ and solve $X_+(t)$ and $X_-(t)$ in the different regions. Hauser and Ernst[1,14] proposed that for a given solution $F_0(t)$, which is assumed to be holomorphic on the whole t plane except at $t_+=1/2\eta_+$ and $t_-=1/2\eta_-$ and lying in C_-, the kernel $G(t)$ can be constructed by the form

$$G(t)=F_0(t)u(t)F_0^{-1}(t), \tag{3.2}$$

where $u(t)$ is defined as a 2×2 matrix function of t, independent of x_+ and x_-, such that $u(t)$ is holomorphic in $C+C_-$ and satisfies the conditions

$$u(t)^{\dagger}\Omega u(t)=\Omega,\quad \det u(t)=1. \tag{3.3}$$

Then Hauser and Ernst were able to verify that with the aid of the solutions $X_+(t)$ and $X_-(t)$ to the Riemann–Hilbert transformation, with the boundary condition

$$X_+(0)=I, \tag{3.4}$$

one can construct a new solution $F(t)$ to Eqs. (2.12)–(2.17) by defining

$$\begin{aligned}F(t)&=X_+(t)F_0(t)\ \text{ in }\ C_+,\\ &=X_-(t)F_0(t)u(t)^{-1}\ \text{ in }\ C_-,\end{aligned} \tag{3.5}$$

with the new matrix Ernst potential E:

$$E=E_0+\dot{X}_+(0). \tag{3.6}$$

Further investigation shows that the Hauser–Ernst approach to the Riemann–Hilbert transformation establishes a simpler representation of the Geroch group, a group whose elements consist of the set of 2×2 matrix functions $u(t)$ subject to the conditions (3.3) and whose composition law corresponds to the exact form of the Kac–Moody group. However, for our purpose we would like to extend the Hauser–Ernst method to the more general cases in which the Virasoro group, of which an infinite set of infinitesimal generators were found to act on the solution space of the Ernst field equations, will be described in addition to the Kac–Moody group.

Now let us first define a scalar function $v(t)$ such that $v(t)$ is independent of x_+ and x_- and is holomorphic on $C+C_-$ except at infinity, where $v(t)$ tends to linear divergence and such that $v(t)$ is a linear function or has singularities in C_+. We further state the restriction that for fixed x_+ and x_-, t_+ and t_- are single-value solutions to

$$(1-2v(t)\eta_+)(1-2v(t)\eta_-)=0, \tag{3.7}$$

which lie in C_-; $\dot{v}(t_\pm)\neq0$. We introduce the new notations η'_+ and η'_- such that

$$\begin{aligned}t_+&=1/2\eta'_+=v^{-1}(1/2\eta_+),\\ t_-&=1/2\eta'_-=v^{-1}(1/2\eta_-).\end{aligned} \tag{3.8}$$

From Eq. (2.11), we can easily prove that η'_+ and η'_- satisfy

$$\partial_+\eta'_-=\partial_-\eta'_+=0. \tag{3.9}$$

In fact, Eqs. (3.8) can be interpreted as the transforms of variables between different coordinate systems. Thus we define that under the transformations (3.8) the given solution $F_0(t)$ to Eqs. (2.12)–(2.16) is changed into $F'_0(t)$, which satisfies the linearization equations and corresponding subsidiary conditions by replacing η_+ and η_- with η'_+ and η'_-. Thus we have E'_0.

Hence we prefer to select the kernel $G(t)$ as the form

$$G(t)=F'_0(t)u(t)F_0(v(t))^{-1}\ \text{ on }\ C, \tag{3.10}$$

where $u(t)$ is the same as given above. Thus in our case the Riemann–Hilbert transformation can be written in the form

$$\begin{aligned}X_-&=X_+(t)G(t)\\ &=X_+(t)F'_0(t)u(t)F_0(v(t))^{-1}\ \text{ on }\ C,\end{aligned} \tag{3.11}$$

with the boundary condition

$$X_+(0)=I. \tag{3.12}$$

Similar to Hauser and Ernst,[1,14] we shall exploit $X_+(t)$ and $X_-(t)$ in order to construct the new function $F(t)$ by

$$\begin{aligned}F(t)&=X_+(t)F'_0(t),\ \text{ in }\ C_+,\\ &=X_-(t)F_0(v(t))u^{-1}(t)\ \text{ in }\ C_-.\end{aligned} \tag{3.13}$$

We shall show that $F(t)$ constructed in Eqs. (3.13) is a solution to the following:

$$\partial_+ F(t) = [t/(1-2t\eta'_+)]\partial_+ E\,\Omega F(t),$$
$$\partial_- F(t) = [t/(1-2t\eta'_-)]\partial_- E\,\Omega F(t), \tag{3.14}$$

$$A(t)\partial_+ F(t) = \partial_+ E\,\Omega F(t),$$
$$A(t)\partial_- F(t) = \partial_- E\,\Omega F(t), \tag{3.15}$$

$$F(0) = I, \tag{3.16}$$

$$\dot{F}(0) = E\Omega, \tag{3.17}$$

$$\det F(t) = \lambda'(t)^{-1}, \tag{3.18}$$

$$F(t)^\dagger \Omega A(t) F(t) = \Omega, \tag{3.19}$$

where

$$E = E'_0 + \dot{X}'_+(0)\Omega, \tag{3.20}$$

$$A(t) = I - t(E + E^\dagger)\Omega, \tag{3.21}$$

$$\lambda'(t) = [(1-2t\eta'_+)(1-2t\eta'_-)]^{1/2}. \tag{3.22}$$

To prove Eqs. (3.14), we operate Eqs. (3.13) in differentiation with respect to x_+ and obtain

$$\begin{aligned}\partial_+ F(t)\,F(t)^{-1} &= \partial_+ X_+(t) X_+(t)^{-1} + [t/(1-2t\eta'_+)] X_+(t)\partial_+ E'_0\,\Omega\, X_+^{-1}(t) \text{ in } C_+,\\ &= \partial_+ X_-(t) X_-(t)^{-1} + \{v(t)/[1-2v(t)\eta_+]\} X_-(t)\partial_+ E_0\,\Omega\, X_-^{-1}(t) \text{ in } C_-.\end{aligned} \tag{3.23}$$

Since $X_+(t)$ and $X_-(t)$ are, respectively, holomorphic in C_+ and C_- and $v(t)$ is holomorphic in C_- we observe Eqs. (3.23) such that there exists only one simple singularity at $t = 1/2\eta'_+$ on the whole t plane. Thus Eqs. (3.23) imply that

$$\partial_+ F(t)\,F(t)^{-1} = P + [t/(1-2t\eta'_+)]Q, \tag{3.24}$$

where P and Q independent of t are undetermined. From Eqs. (3.12) and (3.23) we obtain

$$P = 0, \quad Q = \partial_+(\dot{X}_+(0) + E'_0\Omega). \tag{3.25}$$

If we set

$$Q = \partial_+ E, \tag{3.26}$$

we obtain Eqs. (3.14) for x_+, together with Eq. (3.20); in a similar way, we can verify a component of Eqs. (3.14) for x_-.

Before we prove Eqs. (3.15), we show

$$\begin{aligned}&X_\pm^{-1}(t)^\dagger \Omega A'_0(t) X_\pm^{-1}(t)\\ &\quad = X_\mp^{-1}(t)^\dagger \Omega A_0(v(t)) X_\mp^{-1}(t)\\ &\quad = \Omega A(t)\end{aligned} \tag{3.27}$$

and

$$\begin{aligned}&\Omega A(t)\partial_\pm X_+(t) X_\pm^{-1}(t)\\ &\quad + t X_\pm^{-1}(t)^\dagger \Omega\,\partial_\pm E'_0\,\Omega X_\pm^{-1}(t)^\dagger\\ &\quad = \Omega A(t)\partial_\pm X_-(t) X_\mp^{-1}(t)\\ &\qquad + v(t) X_\mp^{-1}(t)^\dagger \Omega\,\partial_\pm E_0\,\Omega X_\mp^{-1}(t)^\dagger\\ &\quad = t\Omega\,\partial_\pm E.\end{aligned} \tag{3.28}$$

The first equality of Eq. (3.27) is derived from Eqs. (2.16) and (3.13); it can be expressed as a linear function of t since $v(t)$ is linear in t as well as $A(v(t))$ when t tends to infinity. After determining the coefficients of this linear function, we confirm the second equality of Eq. (3.27). Under similar consideration and using Eq. (2.20), it is not difficult to prove Eq. (3.28).

Therefore, it follows from Eqs. (2.16), (3.27), and (3.28) that

$$\begin{aligned}A(t)\partial_\pm F(t) &= A(t)\partial_\pm F'_0(t)\\ &\quad + A(t)X_+(t)\partial_\pm F'_0(t)\\ &= A(t)\partial_\pm F'_0(t) + \Omega X_+(t)\Omega\\ &\quad \times A'_0(t)\,\partial_\pm F'_0(t)\\ &= A(t)\partial_\pm F'_0(t) + t\Omega X_+(t)\Omega\\ &\quad \times \partial_\pm E'_0\Omega F'_0(t)\\ &= t\Omega\,\partial_\pm E\,\Omega X_+(t)\,F'_0(t)\\ &= t\Omega\,\partial_\pm E\,\Omega F(t).\end{aligned} \tag{3.29}$$

Thus we complete the proof of Eqs. (3.15).

To prove Eq. (3.18) from Eqs. (3.14), we observe that

$$\det X_+(t) = \det X_-(t)\,[\lambda(v(t))/\lambda'(t)]. \tag{3.30}$$

Since the singularities at t_+ and t_- on the rhs of Eq. (3.30) can be eliminated, the functions on both sides entice functions for all t. Under the restriction of boundary condition (3.12) we conclude that the function is equal to 1. As a consequence of Eqs. (2.15), (3.13), and (3.30), we can obtain Eq. (3.18).

By using Eqs. (2.16), (3.3), and (3.27), we observe that

$$\begin{aligned}&F(t)^\dagger \Omega A(t) F(t)\\ &\quad = u^{-1}(t)^\dagger F'_0(t)^\dagger X_+(t)^\dagger \Omega A(t) X_+(t)\,F'_0(t) u^{-1}(t)\\ &\quad = u^{-1}(t)^\dagger F'_0(t)^\dagger \Omega\, A'_0(t)\,F'_0(t) u^{-1}(t)\\ &\quad = u^{-1}(t)^\dagger \Omega u^{-1}(t)\\ &\quad = \Omega.\end{aligned} \tag{3.31}$$

Now let us prove that the integral equation (1.1) given in Sec. I is identical to the representation of the Riemann–Hilbert transformation, i.e., both are equivalent. To do this, we note that since $X_-(t)$ is analytic in $C + C_-$ (including $t = \infty$), we have

$$\frac{1}{2\pi i}\int_C \frac{X_-(s)}{s(s-t)}\,ds = 0 \quad (t \text{ in } C_+), \tag{3.32}$$

where we used

$$\frac{1}{2\pi i}\int_C \frac{1}{s(s-t)}\,ds = 0 \quad (t \text{ in } C_+). \tag{3.33}$$

In terms of the Cauchy theorem, we substitute Eq. (3.13) into Eq. (3.32) to obtain

$$\frac{1}{2\pi i}\int_{C_{0,t}} \frac{1}{s(s-t)} F(s)u(s)\, F_0^{-1}(v(s))ds = 0 \tag{3.34}$$

subject to the conditions

$$F(0) = I, \quad \dot{F}(0) = E\Omega. \tag{3.35}$$

Therefore, we complete our proof of the integral equation (1.1).

Finally, we should prove that E given in Eq. (3.20) satisfies the Ernst field equations and that its trace is equal to β'; we can obtain the new metric g from this new Ernst potential.

From Eqs. (3.14) and (3.15), we have

$$A(t)\partial_{\pm} F(t) = (1 - 2t\,\eta'_{\pm})\partial_{\pm} F(t). \tag{3.36}$$

Differentiating (3.36) twice with respect to t and setting $t = 0$, we see that E satisfies Eqs. (2.6) with $\eta'_{\pm}$.

We then take the trace to Eq. (3.20) to obtain

$$\mathrm{tr}(E\Omega) = \mathrm{tr}(E'_0\Omega) = 2\beta' \tag{3.37}$$

because

$$\begin{aligned}\mathrm{tr}\,\dot{X}_+(0) &= \mathrm{tr}(\dot{X}_+(0)\, X_+^{-1}(0)) \\ &= (\det X_+(0))^{-1}\frac{\partial}{\partial t}(\det X_+(0)) = 0,\end{aligned} \tag{3.38}$$

where we used $X_+(0) = I$. Equation (3.37) is equivalent to

$$E - E^T = 2\beta'\Omega, \tag{3.39}$$

where the superscript T stands for the transport operator.

According to the definition of g,

$$g = \tfrac{1}{2}(E + E^{\dagger}) - \beta'\Omega, \tag{3.40}$$

it is obvious that g is Hermitian. If we show that g is real, then it must be real symmetric. From Eq. (3.39), it follows that

$$\begin{aligned} g &= \tfrac{1}{2}(E - E^T) + \tfrac{1}{2}(E^T + E^{\dagger}) - \beta'\Omega \\ &= \tfrac{1}{2}(E^T + E^{\dagger}) \\ &= g^*. \end{aligned} \tag{3.41}$$

It is not difficult to deduce

$$\det g = (\Lambda\alpha')^2, \tag{3.42}$$

which follows from

$$A(t)^T\Omega A(t) = \det A(t)\Omega = \lambda'^2(t)\Omega \tag{3.43}$$

and Eqs. (3.19) and (3.40).

IV. DERIVATION OF THE INFINITESIMAL SYMMETRY TRANSFORMATIONS

In this section we shall discuss the relationship between the Reimann–Hilbert transformation and the infinitesimal symmetry transformations given by us previously.[9] We proposed[9] the infinitesimal symmetry transformations

$$\begin{aligned}\delta F(t) = &-[t/(t-t')]\{t\dot{F}(t)F^{-1}(t) \\ &- t'\dot{F}(t')F^{-1}(T')\}F(t)\end{aligned} \tag{4.1}$$

to the linearization equations (2.12) and

$$\delta E = -t'\dot{F}(t')F^{-1}(t')\Omega \tag{4.2}$$

to the Ernst field equation, where for the convenience of discussion infinitesimal constants are not written out. We showed[9] that these transformations constitute the Virasoro algebra by expanding the powers of the parameter t'. On the other hand, as indicated above, the Riemann–Hilbert transformation for $u(t) = I$ corresponds to the representation of the Virasoro group; thus we expect to derive the transformations (4.1) and (4.2) from

$$\frac{1}{2\pi i}\int_{C_{0,t}} \frac{F(s)\, F_0^{-1}(v(s))}{s(s-t)}\, ds = 0, \tag{4.3}$$

with the boundary conditions

$$F(0) = I, \quad \dot{F}(t) = E\Omega. \tag{4.4}$$

Let us consider the infinitesimal case for Eq. (4.4). Under the infinitesimal transform, we set

$$\delta t = v(t) - t \tag{4.5}$$

and

$$\delta F_0(t) = F(t) - F_0(t). \tag{4.6}$$

Substituting Eqs. (4.4) and (4.5) into Eq. (4.3) we have

$$\begin{aligned}\frac{1}{2\pi i}\int_{C_{0,t}} \frac{1}{s(s-t)}&(F_0(s) + \delta F_0(s))(F_0^{-1}(s) \\ &- \dot{F}_0^{-1}(s)\delta s)\, ds = 0,\end{aligned} \tag{4.7}$$

where we omitted the higher orders of $(\delta s)^2$. Since $F_0(t)$ is holomorphic in C_+ and $F_0(0) = I$, we can integrate

$$\frac{1}{2\pi i}\int_{C_{0,t}} \frac{\delta F_0(s)\, F_0^{-1}(s)}{s(s-t)}\, ds = \delta F_0(t)\, F_0^{-1}(t). \tag{4.8}$$

Hence Eq. (4.6) can be written in the form

$$\delta F_0(t)\, F_0^{-1}(t) = -\frac{1}{2\pi i}\int_{C_{0,t}} \frac{\dot{F}_0(s)\, F_0^{-1}(s)}{s(s-t)}\,\delta s\, ds. \tag{4.9}$$

For convenience, we no longer write the subscript of $F_0(t)$.

According to the definition of $v(t)$ such that $v(t)$ has the singularities of or is a linear function in C_+, without loss of generality, we can select

$$\delta t = \delta^{(k)}t = -t^{-k+1} \quad (k \geqslant 0). \tag{4.10}$$

Substituting Eq. (4.10) into Eq. (4.9), we have

$$\begin{aligned}\delta^{(k)}&F(t)\, F^{-1}(t) \\ &= -\frac{1}{2\pi i}\int_{C_{0,t}} \frac{s^{-k+1}\dot{F}(s)\, F^{-1}(s)}{s(s-t)}\, ds \quad (k \geqslant 0).\end{aligned} \tag{4.11}$$

Equation (4.11) is the infinitesimal Riemann–Hilbert transformation for the Virasoro symmetry.[9,10]

To derive Eq. (4.1), we obtain

$$\delta F(t) = \sum_{k=0}^{\infty} \delta^{(k)}F(t)t'^k, \tag{4.12}$$

where t' lies in C_+. Since we always have $|t'/s| \leqslant 1$, by using

$$\frac{1}{s-t} = \sum_{k=0}^{\infty} \frac{t'^k}{s^{k+1}} \tag{4.13}$$

we thus obtain

$$\delta F(t)\, F^{-1}(t)$$
$$= -\frac{1}{2\pi i}\int_{C_{0,t}} \sum_{k=0}^{\infty} \frac{t'^k s^{-k+1}}{s(s-t)} \dot{F}(s)\, F^{-1}(s)\, ds$$
$$= -\frac{1}{2\pi i}\int_{C_{0,t}} \frac{s}{(s-t)(s-t')} \dot{F}(s)\, F^{-1}(s)\, ds$$
$$= -\frac{t}{t-t'}\{t\dot{F}(t)F^{-1}(t) - t'\dot{F}(t')F^{-1}(t')\}. \quad (4.14)$$

Equation (4.14) implies Eq. (4.2) as a result of Eq. (4.3).

Moreover, we need to give the transform of $\eta_{\pm}$ under the infinitesimal Riemann–Hilbert transformation. In terms of Eqs. (3.8) and (4.10), the transform yields

$$\delta^{(k)}\eta_{\pm} = \eta'_{\pm} - \eta_{\pm}$$
$$= 1/2v^{-1}(t_{\pm}) - 1/2t_{\pm}$$
$$= (1/2t^2_{\pm})\, \delta^{(k)} t_{\pm}$$
$$= -\eta_{\pm}(2\eta_{\pm})^k \quad (k \geqslant 0) \quad (4.15)$$

Then we obtain

$$\delta\eta_{\pm} = \sum_{k=0} \delta^{(k)}\eta_{\pm} t'^k$$
$$= -\eta_{\pm} \sum_{k=0} (2\eta_{\pm})^k t'^k$$
$$= -\frac{\eta_{\pm}}{1-2t'\eta_{\pm}}, \quad (4.16)$$

or equivalently,

$$\delta\alpha = -\alpha/\lambda^2, \quad (4.17)$$
$$\delta\beta = -[\beta(1-2t\beta) + 2\Lambda^2 t\alpha^2]/\lambda^2, \quad (4.18)$$

which are the same as derived from the infinitesimal transformation (4.2) directly.

In order to investigate the structures of the infinitesimal Riemann–Hilbert transformation, we consider the integral equation (1.1) in the case of $v(t) = t$. Parallel to the above treatment, it is known that Kinnesley–Chitre transformations are given by

$$\gamma_a^{(k)} F(t)\, F^{-1}(t)$$
$$= -\frac{1}{2\pi i}\int_{C_{0,t}} \frac{s^{-k} F(s) T_a F^{-1}(s)}{s(s-t)} ds \quad (k \geqslant 0) \quad (4.19)$$

if we take

$$\gamma_a^{(k)} T_a = u(t) - T_a = -T_a t^{-k} \quad (k \geqslant 0), \quad (4.20)$$

where T_a $(a = 1, 2, 3)$ are generators of the Lie algebra SL $(2,R)$ for which the structure constant is denoted by C^c_{ab}. Following the calculations in Ref. 9, we can obtain the following commutations:

$$[\delta^{(k)}, \delta^{(l)}]E = (k-1)\delta^{(k+l)}E, \quad (4.21)$$
$$[\delta^{(k)}, \gamma_b^{(l)}]E = -l\gamma_b^{(k+l)} E, \quad (4.22)$$
$$[\gamma_a^{(k)}, \gamma_b^{(l)}]E = C^c_{ab}\, \gamma_c^{(k+l)} E \quad (4.23)$$

for all $k, l \geqslant 0$. Equations (4.21)–(4.23) reveal the fact that the infinitesimal Riemann–Hilbert transformations span the structure of the semidirect product of the Kac–Moody and Virasoro algebras.

According to the previous discussion, it should be emphasized that the infinite-dimensional Lie algebra only has a positive part because k and l are not allowed to be negative. In another paper,[10] we considered the transform $t \to 1/t$ for the linearization equations: Using solutions to these new linearization equations, we proceeded to find another type of infinitesimal transformations constituting the negative part of the infinite-dimensional Lie algebra. We combined these two parts to form the full algebra, which still lacks the central term. In a similar way, in the present paper we can also show that the other type of infinitesimal transformations originate from the infinitesimal Riemann–Hilbert transformation. Therefore, we can drop out the restriction of the positive k and l in Eqs. (4.21)–(4.23).

V. DISCUSSION AND APPLICATION

To examine the structure of the infinite-dimensional group for the new form of the Riemann–Hilbert transformation, let us take the following cases into account.

(i) If $u(t) = I$, Eqs. (3.13) will give rise to

$$F(t) = X_+(t)\, F'_0(t) \text{ in } C_+,$$
$$= X_-(t) F_0(v(t)) \text{ in } C_-. \quad (5.1)$$

Now we take the transformations

$$F_1(t) = X^0_+(t)\, F'_0(t) \text{ in } C_+,$$
$$= X^0_-(t)\, F_0(v_0(t)) \text{ in } C_- \quad (5.2)$$

and

$$F_2(t) = X^1_+(t)\, F'_1(t) \text{ in } C_+,$$
$$= X^1_-(t) F_1(v_1(t)) \text{ in } C_-. \quad (5.3)$$

Then we define the new transformation

$$F_2(t) = X^2_+(t)\, F''_0(t) \text{ in } C_+,$$
$$= X^2_-(t) F_0(v_2(t)) \text{ in } C_-, \quad (5.4)$$

where $F''_0(t)$ denotes the transform of $F_0(t)$ by replacing $1/2\eta_+$ in $v_2^{-1}(1/2\eta_+)$. Using Eqs. (5.2) and (5.3), the successive transform of $F_2(t)$ can be expressed by

$$v_2(t) = v_0(v_1(t)), \quad (5.5)$$

where $v_0(t)$, $v_1(t)$, and $v_2(t)$, respectively, transform $F_0(t)$ into $F_1(t)$, $F_1(t)$ into $F_2(t)$, and $F_0(t)$ into $F_2(t)$ and where $X^2(t)$ will be determined by the forms $X^0(t)$ and $X^1(t)$ in terms of the inside or outside of circle C. As explained above, this formulation will provide us with a representation of the Virasoro group.

(ii) If $u(t) = I$, Eqs. (3.13) will be reduced to the original Riemann–Hilbert transformation proposed by Hauser and Ernst,[1] i.e,

$$F(t) = X_+(t) F_0(t) \text{ in } C_+,$$
$$= X_-(t) F_0(t) u(t)^{-1} \text{ in } C_-. \quad (5.6)$$

Similar to procedure (i), we can express a successive transform of $F(t)$ as the form

$$u_2(t) = u_0(t) u_1(t), \quad (5.7)$$

where $u_0(t)$, $u_1(t)$, and $u_2(t)$, respectively, transform $F_0(t)$ into $F_1(t)$, $F_1(t)$ into $F_2(t)$, and $F_0(t)$ into $F_2(t)$. This corresponds to a representation of the Kac–Moody group.

(iii) In general, neither $v(t) \neq t$ nor $u(t) \neq I$, the Riemann–Hilbert transformation given in Sec. III, admits the representation of the semidirect product of the Kac–Moody and Virasoro groups. This expression of the Riemann–Hilbert transformation is very important and useful because it becomes possible to prove a Geroch conjecture stating that any given stationary axisymmetric vacuum space-time can be generated from Minkowski space by an infinite set of the symmetry transformations.[14]

In order to see how to apply our method to the derivation of some useful Backlund transformations, let us give a simpler example. Under our consideration, we set

$$u(t) = b \tag{5.8}$$

and

$$v(t) = \tau t + \mu, \tag{5.9}$$

where b is an element of SL $(2, R)$ independent of t and τ and μ are parameters. Thus the corresponding Riemann–Hilbert transformation can be written in the form

$$F(t) b F_0^{-1}(\tau t + \mu) = X_-(t). \tag{5.10}$$

Since the lhs of Eq. (5.10) is analytic on the whole t plane, we can always set

$$X_-(t) = B, \tag{5.11}$$

where B is independent of t. From $F(0) = I$, it follows that

$$B = b F_0^{-1}(\mu). \tag{5.12}$$

Thus we finally obtain that

$$F(t) = b F_0^{-1}(\mu) F_0(\tau t + \mu) b^{-1}, \tag{5.13}$$

which is the formulation of the Neugebauer Backlund transformation found by Cosgrove.[13] If we set $b = I$, the transformation (5.13) is reduced to the combination of the Maison–Cosgrove transformation.[13] We would like to emphasize that from the Hauser–Ernst formalism[1] of the Riemann–Hilbert transformation one cannot derive the Neugebauer Backlund transformation because it was proved[13] that the Neugebauer Backlund transformation lies outside of transformations of the Kac–Moody group. However, since we enlarge the Geroch group by the Virasoro group, we can state that the Neugebauer Backlund transformation is still included in the Geroch symmetry transformations.

It is interesting to compare the present formalism of the Riemann–Hilbert transformation with the Belinskii-Zakharov transformation.[15] We notice that in the Belinskii–Zakharov method some scalar functions are put into a solution to the Belinskii–Zakharov linearization equations to obtain a new solution. This treatment is similar to the present paper. However, the new η'_+ and η'_- in our case are still solutions of the wave equation (2.11); this is not true for the Belinskii–Zakharov method. This is the reason why we, unlike Belinskii and Zakharov, need not redefine the determinant of the new metric to satisfy the wave equation. On the other hand, it is shown[16] that the Belinskii–Zakharov formalism does not correspond to the Hauser–Ernst formalism[1] completely. Thus we conclude that the Belinskii–Zakharov transformation must be associated with the Virasoro symmetry transformation. We need to further investigate these relationships in the future.

Finally, we point out that the approach to the Riemann–Hilbert transformation introduced in this paper can also be applied in other nonlinear systems such as the two-dimensional Heisenberg model and the nonlinear Schrödinger equation.[17] We are hopeful that we can use this new transformation to generate some new solution of the Einstein field equations, which is of much interest in physics, from the known solution.

ACKNOWLEDGMENTS

One of the authors (WL) should like to thank Professor F. Ernst for his support and encouragement.

This research was supported in part by Grant No. PHY-8605958 from the National Science Foundation.

[1] I. Hauser and F. J. Ernst, J. Math. Phys. **21**, 1126; **21**, 1418 (1980).
[2] R. Geroch, J. Math. Phys. **12**, 918 (1971); **13**, 394 (1972).
[3] Y. S. Wu and M. L. Ge, J. Math. Phys. **24**, 1187 (1983).
[4] W. Kinnersly and D. M. Chitre, J. Math. Phys. **18**, 1538 (1977); **19**, 1926 (1978).
[5] K. Ueno and Y. Nakamura, Phys. Lett. B **109**, 273; B **117**, 208 (1982).
[6] L. Dolan, Phys. Rev. Lett. **47**, 1371 (1981).
[7] B. Y. Hou, M. L. Ge, and Y. S. Wu, Phys. Rev. D **24**, 2238 (1982); C. Devchand and D. B. Fairlia, Nucl. Phys. B **194**, 1086 (1982); Y. S. Wu, Commun. Math. Phys. **90**, 461 (1983).
[8] H. Eichenherr, Phys. Lett. B **115**, 385 (1982); H. Bohr, M. L. Ge, and T. V. Volovich, Nucl. Phys. B **260**, 701 (1985).
[9] B. Y. Hou and W. Li, Lett. Math. Phys. **13**, 1 (1987); J. Phys. A **20**, L897 (1987).
[10] W. Li, Phys. Lett. A **129**, 301 (1988).
[11] P. Goddard and D. Olive, J. Mod. Phys. A **1**, 303 (1986).
[12] G. Neugebauer, J. Phys. A **12**, L67 (1979).
[13] D. Maison, J. Math. Phys. **20**, 871 (1979); C. M. Cosgrove, *ibid.* **21**, 2417 (1980).
[14] I. Hauser and F. J. Ernst, J. Math. Phys. **22**, 1051 (1981).
[15] V. A. Belinskii and V. E. Zakharov, Sov. Phys. JETP **48**, 985 (1978).
[16] C. M. Cosgrove, J. Math. Phys. **23**, 615 (1982).
[17] W. Li, Phys. Lett. B **214**, 79 (1988).

Representation theory of the symplectic groups. I

M. D. Gould
Department of Mathematics, University of Queensland, St. Lucia, Queensland 4067, Australia

(Received 6 June 1988; accepted for publication 21 December 1988)

Certain aspects of symplectic group representation theory are investigated. In particular, it is shown that every irreducible representation of $Sp(n)$ admits a relatively large class of states, referred to herein as canonical states, which possess properties analogous to the Gelfand–Tsetlin states appearing in the theory of the orthogonal and unitary groups. The properties of canonical states are investigated and some matrix element formulas are derived.

I. INTRODUCTION

The theory of Lie groups has now been established as an invaluable tool in physical applications. In particular, physicists are familiar with the well-known treatment of the quantum theory of angular momenta, where the group of interest is $SU(2)$. The theory of angular momenta is of fundamental importance in atomic and nuclear physics, where it has been applied to the calculation of wavefunctions, energy levels, and transition probabilities.

Since the development of the angular momenta theory, principally by Racah and Wigner, it has become apparent that higher order Lie groups play an important role in physics. For example, the unitary group $U(n)$ plays a fundamental role in the second quantized formulation of the quantum many-body problem and was made the cornerstone of Moshinsky's work on the nuclear shell model.[1] The orthogonal and symplectic groups also play an important role in physics, particularly in connection with parastatistics,[2] wave equations, and for the classification of states in atomic and nuclear physics.[3] Lie groups also figure prominently in the classification of symmetries in elementary particle physics, which have been extensively studied since the early 1960's. More recently, the group theoretical methods of Moshinsky on the nuclear many-body problem have been extended[4–6] to the many electron problems of atomic physics and quantum chemistry, allowing large-scale configuration interaction calculations to be performed[7] which would be interactable by other methods.

Therefore, it is not surprising that a great deal of interest has been generated in extending the Racah–Wigner theory of angular momenta to all the classical Lie groups. The first major step in this direction was made by Gelfand and Tsetlin (GT),[8] who constructed a full set of basis vectors for the Lie groups $U(n)$ and $O(n)$: The matrix elements of the (elementary) group generators in the GT basis were also obtained. This work was subsequently extended by Baird and Biedenharn,[9] who obtained the matrix elements of all $U(n)$ generators in the GT basis. Moreover, the structure of the matrix elements (as a product of a reduced matrix element and a Wigner coefficient) was determined and hence the fundamental (i.e., vector) Wigner coefficients for $U(n)$ were given for the first time. The evaluation of all multiplicity-free Wigner coefficients for $U(n)$ was subsequently given by Biedenharn and Louck[10] and Baird and Giovannini.[11] Recently, an alternative algebraic approach to this problem was presented by Gould[12] and extended to the orthogonal groups[13] in order to yield a pattern calculus for the Lie groups $U(n)$ and $O(n)$.

Although considerable progress has been made in the representation theory of the orthogonal and unitary groups, the symplectic groups, the remaining family of classical Lie groups, have received comparatively little attention. This is probably due to the fact that, unlike the Lie groups $U(n)$ and $O(n)$, there does not exist a canonical orthogonal basis for the irreducible representations of $Sp(n)$. The crucial property that makes the GT scheme work for $U(n)$ is that in the reduction of an irreducible representation of $U(n)$ into irreducible representations of $U(n-1)$ all irreducible representations occur with unit multiplicity.[9,14] This property is also shared by the orthogonal groups and one would ideally like to obtain a similar solution for the symplectic groups. Unfortunately, however, in the reduction of an irreducible representation of $Sp(n)$ into irreducible representations of its subgroup $Sp(n-2)\times Sp(2)$ multiplicities generally occur[15] and extra invariants are required to completely specify the basis states (the symplectic group state labeling problem). In such a case there still remains the problem of obtaining the eigenvalues of these missing labeling invariants, which are known to be irrational in general, so that the action of the group generators in such a basis is likely to be complicated. In this respect the symplectic group state labeling problem may be regarded as the prototype of all state labeling problems in Lie group theory and applications.

Recently, Gould and Kalnins (GK)[16] obtained a new projection-based solution to the $Sp(n)$ state labeling problem which yields a nonorthogonal GT-type basis for the irreducible representations. Although this solution offers several nice features, particularly the simple determination of the action of the group generators, the method involves the calculation of the relevant overlap coefficients, a problem which is currently unsolved. Nevertheless, the GK solution affords a useful tool, particularly for providing checks on our formalism, and will be applied throughout this paper.

This is the first paper in a series of three in which we investigate certain aspects of symplectic group representation theory. We shall not discuss the labeling problem in this series of papers; instead, we focus attention on a relatively large class of states, herein referred to as canonical states, which possess properties analogous to the GT states appearing in the representation theory of the orthogonal and unitary groups. The space of canonical states of an irreducible representation includes all distinct weights of the representation (at *least* once) and, in particular, contains all states with

 0022-2488/89/061205-14$02.50

weight Weyl group conjugate to the highest weight. The reduced matrix elements, fundamental Wigner coefficients, and matrix elements of *all* Sp(n) generators are determined between arbitrary canonical states, thus completing the first step in a general (multiplicity-free) pattern calculus for the symplectic groups. Following the approach of Refs. 12 and 13 to the orthogonal and unitary groups, it is evident that our formalism may be extended to obtain all multiplicity-free (canonical) Wigner coefficients for Sp(n).

In this paper we introduce canonical basis states and investigate some of their basic properties. In particular, we determine some matrix element formulas for a certain class of canonical states, referred to herein as S dominant, based on the representation theory of the unitary group U(n). In the final two papers of the series we will determine the matrix elements of all Sp(n) generators between arbitrary canonical states. The evaluation of the corresponding reduced matrix elements and Wigner coefficients are also given and their generalized Weyl group symmetries are determined.

Other developments in connection with the symplectic groups have been made by Lohe and Hurst,[17] who have advocated the use of modified boson operators as a method of constructing basis states for the irreducible representations of Sp(n), in analogy with the boson polynomials used[7] in the theory of U(n). Explicit matrix element formulas in certain degenerate representations of Sp(n) have recently been obtained by Klymik[18] and Wong and Yeh.[19] The method of raising and lowering operators to construct bases for the irreducible representations of Sp(n) has been advocated by Michelsson[20] and Bincer.[21] The symplectic groups also figure prominently in Cartan's classification of homogeneous spaces, which afford certain degenerate representations of Sp(n), as studied by Pajas and Raczka.[22] A full set of missing labeling invariants for the symplectic group was recently constructed by Bincer.[23] Finally, the Sp(n)$\downarrow$Sp($n-2$) $\times$Sp(2) branching rules were recently investigated by Cerkaski[24] based on the previous work of Zhelobenko.[15]

II. PRELIMINARIES

We begin by introducing the symplectic group as a subgroup of the unitary group. The n^2 generators a_{ij} ($1 \leqslant i,j \leqslant n$) of the Lie group U($n$) satisfy the commutation relations

$$[a_{ij},a_{kl}] = \delta_{kj}a_{il} - \delta_{il}a_{kj}$$

and are, moreover, required to satisfy the Hermiticity condition

$$a_{ij}^{\dagger} = a_{ji}$$

on finite-dimensional (i.e., unitary) representations of the group. To define the symplectic subgroup Sp(n) we introduce an antisymmetric metric $g_{ij} = -g_{ji}$ in terms of which our Sp(n) generators $\alpha_{ij} = \alpha_{ji}$ are expressible (where summation over p is implied),

$$\alpha_{ij} = g_{ip}a_{pj} + g_{jp}a_{pi}, \tag{1}$$

and satisfy the commutation relations

$$[\alpha_{ij},\alpha_{kl}] = g_{kj}\alpha_{il} - g_{il}\alpha_{kj} + g_{lj}\alpha_{ik} - g_{ik}\alpha_{lj}. \tag{2}$$

We also require the existence of a corresponding contravariant metric g^{ij} satisfying (where the summation convention over repeated indices is implied)

$$g^{ij}g_{jk} = \delta_k^i,$$

which imposes the usual limitation to even values of n.

Without loss of generality we choose the symplectic group metric g_{ij} to be given by

$$g_{ij} = \begin{cases} \delta_{j,i+1}, & i \text{ odd}, \\ -\delta_{j,i-1}, & i \text{ even}, \end{cases} \qquad 1 \leqslant i,j \leqslant n, \tag{3}$$

so that $g^{ij} = -g_{ij}$. We then introduce the operators

$$\alpha_j^i = g^{ik}\alpha_{kj}, \tag{4}$$

which, in view of Eq. (2), satisfy the commutation relations

$$[\alpha_j^i,\alpha_l^k] = \delta_j^k\alpha_l^i - \delta_l^i\alpha_j^k + g^{ik}\alpha_{jl} - g_{jl}\alpha^{ki}, \tag{5}$$

where we define

$$\alpha^{ki} = g^{ij}\alpha_j^k.$$

On finite-dimensional (i.e., unitary) representations of the group the Sp(n) generators (4) satisfy the Hermiticity condition

$$(\alpha_j^i)^{\dagger} = \alpha_i^j \tag{6}$$

and the symmetry property (see Appendix A)

$$\alpha_j^i = -(-1)^{i+j}\alpha_{\bar{i}}^{\bar{j}}, \tag{7}$$

where $\bar{i}$ (the opposite index to i) is defined by

$$\bar{i} = \begin{cases} i-1, & i \text{ even}, \\ i+1, & i \text{ odd}. \end{cases}$$

As a Cartan subalgebra for the Lie algebra of Sp($n = 2h+2$), we take the vector space spanned by the diagonal generators

$$\alpha_i^i, \quad 1 \leqslant i \leqslant n.$$

In view of Eq. (7) only $h+1$ of these operators are linearly independent, so we only need consider the Cartan generators

$$h_i = \alpha_{2i-1}^{2i-1} = -\alpha_{2i}^{2i}, \quad 1 \leqslant i \leqslant h+1$$

whose eigenvalues provide a unique labeling for the system of weights. In view of Eq. (6), we note that the Cartan generators h_i are to be represented by Hermitian matrices on finite-dimensional irreducible representations.

From the commutation relations (5) we deduce

$$[h_i,\alpha_l^k] = (\delta_{2i-1}^k - \delta_{2i-1}^l + \delta_{2i}^l - \delta_{2i}^k)\alpha_l^k,$$

from which it follows that the generators (4) are automatically in Cartan form. If we introduce the fundamental weights

$$\Delta_r = (0,0,\ldots,1,0,\ldots,0), \quad 1 \leqslant r \leqslant h+1$$

consisting of 1 in the rth position and zeros elsewhere, it follows immediately from the above that the roots for the symplectic group Lie algebra are given by the weights

$$\pm(\Delta_i + \Delta_j), \quad i \leqslant j \text{ and } \pm(\Delta_i - \Delta_j), \quad i < j.$$

We take as a system of positive roots the weights

$$\Delta_i + \Delta_j (i \leqslant j), \quad \Delta_i - \Delta_j \quad (i < j).$$

The corresponding generators are given by

$$\alpha_{2j}^{2i-1}(i\leqslant j), \quad \alpha_{2j-1}^{2i-1} \quad (i<j) \tag{8}$$

respectively, which constitute the set of raising generators. We draw particular attention to the elementary raising generators

$$\alpha_n^{n-1}, \quad \alpha_{2i+1}^{2i-1} \quad (1\leqslant i\leqslant h). \tag{9}$$

Every symplectic group raising generator (8) may be obtained by repeated commutation with generators of the form (9). By taking the Hermitian conjugate of Eqs. (8) and (9) we obtain the corresponding set of lowering operators.

We shall be concerned in this paper with the subgroup imbedding

$$\mathrm{Sp}(n) \supset \mathrm{Sp}(n-2)\times\mathrm{Sp}(2),$$

where our $\mathrm{Sp}(n-2)$ generators are given by α_j^i $(1\leqslant i,j\leqslant n-2)$ and our $\mathrm{Sp}(2)$ generators are given by

$$\gamma_\nu^\mu = \alpha_\nu^{\dot\mu}, \quad 1\leqslant\mu,\nu\leqslant 2, \tag{10}$$

where we have adopted the index convention

$$\dot\mu = n-2+\mu. \tag{11}$$

Throughout this paper we denote the weights for $\mathrm{Sp}(n)$ by the Greek letter λ and the weights of the subgroup $\mathrm{Sp}(n-2)$ by the subscripted Greek letter λ_0. We let W (resp. W_0) denote the Weyl group of $\mathrm{Sp}(n)$ [resp. $\mathrm{Sp}(n-2)$]: Recall that

$$W \cong S_{h+1} \circledS \mathbb{Z}_2^{h+1}$$

where S_{h+1} is the symmetric group on $h+1$ objects. In other words, the Weyl group of $\mathrm{Sp}(n)$ consists of all sign changes and permutations of the weight components.

The finite-dimensional irreducible representations of $\mathrm{Sp}(n)$ are uniquely characterized by their highest weights $\lambda = (\lambda_1,\lambda_2,\ldots,\lambda_{h+1})$, whose components λ_r are to be integers satisfying the inequalities

$$\lambda_1\geqslant\lambda_2\geqslant\cdots\geqslant\lambda_{h+1}\geqslant 0. \tag{12}$$

We let $V(\lambda)$ [resp. $V(\lambda_0)$] denote the finite-dimensional irreducible module over $\mathrm{Sp}(n)$ [resp. $\mathrm{Sp}(n-2)$] with the highest weight λ (resp. λ_0) and we denote the corresponding representation by π_λ (respectively, π_{λ_0}). We let H^* (resp. H_0^*) denote the dual of the Cartan subalgebra of $\mathrm{Sp}(n)$ [resp. $\mathrm{Sp}(n-2)$] and we let $\Lambda^+ \subseteq H^*$ (resp. $\Lambda_0^+ \subseteq H_0^*$) be the lattice of dominant integral weights, i.e., Λ^+ consists of those weights λ whose components are integers satisfying the inequalities (12). Finally, we denote elements of $H^*\times H_0^*$, herein referred to as *extended weights*, by (λ,λ_0), where $\lambda\in H^*$, $\lambda_0\in H_0^*$.

We call an extended weight $(\lambda,\lambda_0)\in\Lambda^+\times\Lambda_0^+$ *lexical* if the finite-dimensional irreducible $\mathrm{Sp}(n-2)$ module $V(\lambda_0)$ is contained in the finite-dimensional irreducible $\mathrm{Sp}(n)$ module $V(\lambda)$. The set of lexical (extended) weights $\mathscr{L}\subset\Lambda^+\times\Lambda_0^+$ may clearly be deduced from the $\mathrm{Sp}(n)\supset\mathrm{Sp}(n-2)$ branching rules to which we now turn.

III. BRANCHING RULES AND THE GK BASIS

Following Zhelobenko,[15] the branching rules for the reduction $\mathrm{Sp}(n)\downarrow\mathrm{Sp}(n-2)$ may be obtained by associating with each $\mathrm{Sp}(n)$ weight $\lambda\in\Lambda^+$ the patterns

$$\begin{array}{ccccccccc}
\lambda_1 & & \lambda_2 & & \cdots & & \lambda_h & & \lambda_{h+1} \\
& \mu_1 & & \mu_2 & & \cdots & & \mu_h & & \mu_{h+1}, \\
& & \lambda_{0_1} & & \lambda_{0_2} & & \cdots & & \lambda_{0_h}
\end{array} \tag{13}$$

where μ_i, λ_{0_i} are integers satisfying the inequalities

$$\begin{aligned}
&\lambda_{h+1}\geqslant\mu_{h+1}\geqslant 0, \quad \lambda_i\geqslant\mu_i\geqslant\lambda_{i+1}, \\
&\mu_i\geqslant\lambda_{0_i}\geqslant\mu_{i+1}, \quad 1\leqslant i\leqslant h.
\end{aligned} \tag{14}$$

The weights $\lambda_0 = (\lambda_{0_1},\lambda_{0_2},\ldots,\lambda_{0_h})$ occurring in Eq. (13) determine the highest weights of the irreducible $\mathrm{Sp}(n-2)$ modules which may occur in the irreducible $\mathrm{Sp}(n)$ module $V(\lambda)$. From the betweenness conditions (14) it thus follows that an extended weight $(\lambda,\lambda_0)\in\Lambda^+\times\Lambda_0^+$ is lexical if and only if the components of λ and λ_0 satisfy the inequalities

$$\lambda_i\geqslant\lambda_{0_i}\geqslant\lambda_{i+2}\,(1\leqslant i<h), \quad \lambda_h\geqslant\lambda_{0_h}\geqslant 0. \tag{15}$$

The multiplicity with which the irreducible L_0 module $V(\lambda_0)$ occurs in $V(\lambda)$ is then given by the number of tuples $(\mu_1,\ldots,\mu_{h+1})$ of non-negative integers whose components satisfy Eq. (14). For a given lexical weight (λ,λ_0) we see that the components μ_i must satisfy the inequalities

$$\begin{aligned}
&\lambda_1\geqslant\mu_1\geqslant\lambda_{0_1}\vee\lambda_2, \quad \lambda_{h+1}\wedge\lambda_{0_h}\geqslant\mu_{h+1}\geqslant 0, \\
&\lambda_{i+1}\wedge\lambda_{0_i}\geqslant\mu_{i+1}\geqslant\lambda_{i+2}\vee\lambda_{0_{i+1}}, \quad 1\leqslant i\leqslant h-1,
\end{aligned} \tag{16}$$

where $a\wedge b$ (resp. $a\vee b$) denotes $\min(a,b)$ [resp. $\max(a,b)$].

Following Gould and Kalnins[16] let $V(\hat\lambda)$ denote the finite-dimensional irreducible module over $\mathrm{U}(n=2h+2)$, with the highest weight

$$\hat\lambda = (\lambda_1,\lambda_2,\ldots,\lambda_{h+1},0,\ldots,0).$$

It is well known[15,25] that the irreducible $\mathrm{Sp}(n)$ module $V(\lambda)$ occurs in $V(\hat\lambda)$ with unit multiplicity and hence may be obtained by central projection from $V(\hat\lambda)$. To this end let $\Pi(\hat\lambda)$ denote the set of all $\mathrm{Sp}(n)$ highest weights occurring in $V(\hat\lambda)$, but excluding $\lambda=(\lambda_1,\ldots,\lambda_{h+1})$. Then set

$$p^\lambda = \prod_{\nu\in\Pi(\hat\lambda)}\left(\frac{\sigma_2-\langle\sigma_2\rangle_\nu}{\langle\sigma_2\rangle_\lambda-\langle\sigma_2\rangle_\nu}\right), \tag{17}$$

where $\sigma_2 = \alpha_j^i\,\alpha_i^j$ is the second-order invariant of $\mathrm{Sp}(n)$ and[26]

$$\langle\sigma_2\rangle_\nu = 2\sum_{r=1}^{h+1}\nu_r(\nu_r+n+2-2r)$$

is the eigenvalue of σ_2 in the irreducible module $V(\nu)$. We have the following result.

Lemma 1: The Casimir invariant σ_2 separates λ from the weights in $\Pi(\hat{\lambda})$ and, in particular,

$$V(\lambda) = p^{\lambda} V(\hat{\lambda}).$$

Proof: Following Hammermesh,[25] the highest weights of the irreducible $\mathrm{Sp}(n)$ modules occurring in $V(\hat{\lambda})$ are either equal to λ or obtained from λ via *contraction*. Thus the highest $\mathrm{Sp}(n)$ weights occurring in $\Pi(\hat{\lambda})$ are of the form

$$\nu = \lambda - \sum_{r=1}^{h+1} m_r \Delta_r, \tag{18}$$

where each m_r is a non-negative integer (such that $m_1 + m_2 + \cdots m_{h+1}$ is an even positive integer). Thus for ν as in Eq. (18) we have

$$\langle \sigma_2 \rangle_{\lambda} - \langle \sigma_2 \rangle_{\nu} = 2 \sum_{r=1}^{h+1} m_r (\lambda_r + \nu_r + n + 2 - 2r).$$

Since $\nu \in \Lambda^+$ (by assumption) it follows that

$$\lambda_r + \nu_r + n + 2 - 2r > 0, \quad r = 1,\ldots,h+1;$$

since $m_r \geqslant 0$ $(1 \leqslant r \leqslant h+1)$, the rhs above must be non-negative and only vanishes when $m_r = 0$ $(1 \leqslant r \leqslant h+1)$, in which case $\nu = \lambda$.

This proves the desired result that σ_2 separates λ from the weights in $\Pi(\hat{\lambda})$. It follows immediately that the projection operator (17) is well defined and projects $V(\hat{\lambda})$ onto the subspace $V(\lambda)$ as required.

Q.E.D.

The above result shows that we may obtain a set of vectors spanning the irreducible $\mathrm{Sp}(n)$ module $V(\lambda)$ by considering the central projector (17) applied to the $\mathrm{U}(n)$ GT basis states of the space $V(\hat{\lambda})$. Following Ref. 16, in order to obtain a complete set of linearly independent basis states it suffices to restrict to GT vectors of the form (herein referred to as *allowed* GT states)

$$\begin{pmatrix}
\lambda_{1h+1} & \lambda_{2h+1} & \cdots & \lambda_{h+1\,h+1} & 0 & \cdots & 0 \\
\mu_{1h+1} & \mu_{2h+1} & \cdots & \mu_{h+1\,h+1} & 0 & \cdots & 0 \\
 & \lambda_{1h} & \lambda_{2h} & \cdots\ \lambda_{hh} & 0 & \cdots & 0 \\
 & \mu_{1h} & \mu_{2h} & \cdots\ \mu_{hh} & 0 & \cdots & 0 \\
 & & \ddots & \ddots & & \iddots & \\
 & & \lambda_{12} & \lambda_{22} & 0 & 0 & \\
 & & \mu_{12} & \mu_{22} & 0 & & \\
 & & & \lambda_{11} & 0 & & \\
 & & & \mu_{11} & & &
\end{pmatrix},$$

which, for simplicity, we write in the form

$$\begin{pmatrix}
\lambda_{1\,h+1} & \lambda_{2\,h+1} \cdots \lambda_{h+1\,h+1} \\
\mu_{1\,h+1} & \mu_{2\,h+1} \cdots \mu_{h+1\,h+1} \\
\lambda_{1h} & \lambda_{2h} \cdots \lambda_{hh} \\
\mu_{1h} & \mu_{2h} \cdots \mu_{hh} \\
\vdots & \\
\lambda_{11} & \\
\mu_{11} &
\end{pmatrix}. \tag{19}$$

We denote the space spanned by all allowed $\mathrm{U}(n)$ GT states (19) by $A(\lambda)$.

From the $\mathrm{U}(n)$ GT betweenness conditions[9] the integers $\mu_{i,j}$ and $\lambda_{i,j}$ in the pattern (19) must satisfy the inequalities

$$\lambda_{1m} \geqslant \mu_{1m} \geqslant \lambda_{2m} \geqslant \cdots \geqslant \lambda_{mm} \geqslant \mu_{mm} \geqslant 0,$$
$$\mu_{1m} \geqslant \lambda_{1\,m-1} \geqslant \mu_{2m} \geqslant \cdots \geqslant \lambda_{m-1\,m-1} \geqslant \mu_{mm} \geqslant 0,$$

in agreement with the betweenness conditions of Eq. (14). We sometimes find it more convenient to write the allowed $\mathrm{U}(n)$ GT state (19) in dual-pattern notation:

$$\begin{pmatrix} (\lambda) \\ (\mu) \end{pmatrix}, \tag{20}$$

where (λ), (μ) denote the patterns

$$(\lambda)=\begin{pmatrix}\lambda_{1h+1} & \lambda_{2h+1} & \cdots & \lambda_{h+1\,h+1}\\ \lambda_{1h} & \lambda_{2h} & \cdots & \lambda_{hh}\\ \vdots & & \iddots & \\ \lambda_{12} & \lambda_{22} & & \\ \lambda_{11} & & & \end{pmatrix},\quad (\mu)=\begin{pmatrix}\mu_{1h+1} & \mu_{2h+1} & \cdots & \mu_{h+1\,h+1}\\ \mu_{1h} & \mu_{2h} & \cdots & \mu_{hh}\\ \vdots & & \iddots & \\ \mu_{12} & \mu_{22} & & \\ \mu_{11} & & & \end{pmatrix},$$

respectively.

Associated with each row λ_m in the (upper) (λ) pattern we construct the associated Sp($2m$) projector p^{λ_m} in direct analogy with Eq. (17). We then construct the compound projector

$$P_{(\lambda)}=\prod_{m=1}^{h+1}p^{\lambda_m}.$$

It follows from Ref. 16 that the Sp(n) states

$$\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle = P_{(\lambda)}\left|\begin{matrix}(\lambda)\\(\mu)\end{matrix}\right\rangle \tag{21}$$

form a full set of linearly independent basis states for $V(\lambda)$ which is symmetry adapted to the subgroup chain

$$\mathrm{Sp}(n)\supset \mathrm{Sp}(n-2)\supset\cdots\supset \mathrm{Sp}(2).$$

This is the state labeling scheme obtained by Gould and Kalnins.[16] We note that the states (21), although not an orthonormal set, are orthogonal w.r.t. their upper patterns.

The GK states (21) are weight states of weight $\nu=(\nu_1,\nu_2,\ldots,\nu_{h+1})$, where the components ν_i are given by

$$\nu_i=2\sum_{j=1}^{i}\mu_{j,i}-\sum_{j=1}^{i}\lambda_{j,i}-\sum_{j=1}^{i-1}\lambda_{j,i-1}.$$

This result is of importance since it allows us to obtain the branching rules for the reduction $\mathrm{Sp}(n)\downarrow\mathrm{Sp}(n-2)\times\mathrm{Sp}(2)$. Associated with each lexical weight $(\lambda,\lambda_0)\in\mathscr{L}$ and pattern (13) is an Sp($n-2$) maximal state of weight λ_0 whose corresponding Sp(2) weight (given by the eigenvalue of the Cartan generator h_{h+1}) is determined by the integer

$$\Omega=2\int(\mu)-\int(\lambda)-\int(\lambda_0), \tag{22}$$

where we define

$$\int(\lambda)=\sum_{i=1}^{h+1}\lambda_i,\quad \int(\lambda_0)=\sum_{i=1}^{h}\lambda_{0_i},\quad \text{etc.} \tag{23}$$

The multiplicity $m(\Omega)$ with which the Sp(2) weight Ω occurs is clearly given by the number of $h+1$ tuples $(\mu_1,\mu_2,\ldots,\mu_{h+1})$ whose components are non-negative integers satisfying the inequalities (16) and Eq. (22).

We note that for a given lexical weight (λ,λ_0) the maximum possible values for the integers μ_r satisfying the inequalities (16) are given by

$$\mu_1=\lambda_1,\quad \mu_{r+1}=\lambda_{r+1}\wedge\lambda_{0_r},\quad r=1,\ldots,h. \tag{24}$$

It follows that the maximum possible Sp(2) weight for a given Sp($n-2$) weight $\lambda_0\in\Lambda_0^+$ occurring in $V(\lambda)$ is given by

$$\begin{aligned}\Omega_{\lambda,\lambda_0}&=2\left[\lambda_1+\sum_{r=1}^{h}\lambda_{r+1}\wedge\lambda_{0_r}\right]-\int(\lambda)-\int(\lambda_0)\\ &=\lambda_1+\sum_{r=1}^{h}(\lambda_{r+1}\wedge\lambda_{0_r}-\lambda_{r+1}\vee\lambda_{0_r})\\ &=\lambda_1-\sum_{r=1}^{h}|\lambda_{0_r}-\lambda_{r+1}|,\end{aligned} \tag{25}$$

which can be shown to be a non-negative integer.

It thus follows that the irreducible module over $\mathrm{Sp}(n-2)\times\mathrm{Sp}(2)$ with the highest weight $\lambda_0\times\Omega_{\lambda,\lambda_0}$ occurs exactly once in the irreducible Sp(n) module $V(\lambda)$. The remaining possible $\mathrm{Sp}(n-2)\times\mathrm{Sp}(2)$ highest weights occurring for a given lexical weight $(\lambda,\lambda_0)\in\mathscr{L}$ are thus necessarily of the form $\lambda_0\times\Omega$, with

$$\Omega=\Omega_{\lambda,\lambda_0},\ \Omega_{\lambda,\lambda_0}-2,\ldots,\quad\begin{cases}0, & \int(\lambda)+\int(\lambda_0)\ \text{even},\\ 1, & \int(\lambda)+\int(\lambda_0)\ \text{odd}.\end{cases}$$

The multiplicity $M(\Omega)$ (possibly zero) with which the corresponding irreducible $\mathrm{Sp}(n-2)\times\mathrm{Sp}(2)$ module occurs in $V(\lambda)$ is clearly given by

$$M(\Omega)=m(\Omega)-m(\Omega+2),$$

with $m(\Omega)$ as before. We note that the multiplicities $m(\Omega)$ and hence $M(\Omega)$ can be obtained directly from patterns (13).

Throughout this paper we denote the irreducible module over $\mathrm{Sp}(n-2)\times\mathrm{Sp}(2)$ with the highest weight $\lambda_0\times\Omega_{\lambda,\lambda_0}$ by $V(\lambda,\lambda_0)$, herein referred to as a *canonical submodule* of $V(\lambda)$. Thus the irreducible $\mathrm{Sp}(n-2)\times\mathrm{Sp}(2)$ canonical submodule $V(\lambda,\lambda_0)$ of $V(\lambda)$ is uniquely determined as that submodule whose Sp(2) representation label takes the maximum possible value, given by Eq. (25), for given Sp(n) and Sp($n-2$) representation labels λ,λ_0, respectively.

IV. CANONICAL STATES AND SIGNATURES

Let $\mathscr{C}$ denote the centralizer of $\mathrm{Sp}(n-2)\times\mathrm{Sp}(2)$ in the universal enveloping algebra U of Sp(n), viz.,

$$\begin{aligned}\mathscr{C}=\{c\in\mathrm{U}\,|\,&[\alpha_j^i,c]=[\gamma_{\nu}^{\mu},c]\\ &=0,1\leqslant i,j\leqslant n-2,1\leqslant\mu,\nu\leqslant 2\}.\end{aligned} \tag{26}$$

We remark that Eq. (26) is the algebra from which "missing labeling invariants" may be chosen to resolve the multiplicities occurring in the reduction $\mathrm{Sp}(n)\downarrow\mathrm{Sp}(n-2)\times\mathrm{Sp}(2)$ (the so-called symplectic group state labeling problem) which has been previously discussed by several authors.[23,27]

In view of the work of Joseph,[28] we note that the center of the algebra (26), viz.,

$$\mathscr{C}_0 = \{c_0 \in \mathscr{C} \,|\, [c_0, c] = 0,\ \forall\, c \in \mathscr{C}\}$$

is given explicitly by

$$\mathscr{C}_0 = ZZ_0\Sigma,$$

where Z (resp. Z_0, Σ) is the center of the enveloping algebra of $\mathrm{Sp}(n)$ [resp. $\mathrm{Sp}(n-2)$, $\mathrm{Sp}(2)$].

If $V(\lambda,\lambda_0)$ is an $\mathrm{Sp}(n-2)\times\mathrm{Sp}(2)$ irreducible canonical submodule of $V(\lambda)$ it follows, since the $\mathrm{Sp}(n-2)\times\mathrm{Sp}(2)$ highest weight $\lambda_0\times\Omega_{\lambda,\lambda_0}$ occurs exactly once in $V(\lambda)$, that elements of the centralizer $\mathscr{C}$ must reduce to scalar multiples when acting on $V(\lambda,\lambda_0)$. Thus corresponding to every lexical weight $(\lambda,\lambda_0) \in \mathscr{L}$, we have an algebra homomorphism

$$\chi_{\lambda,\lambda_0}: \mathscr{C} \to \mathbb{C}, \quad c \to \chi_{\lambda,\lambda_0}(c), \tag{27}$$

where $\chi_{\lambda,\lambda_0}(c)$ is the eigenvalue of the centralizer element $c \in \mathscr{C}$ on the canonical submodule $V(\lambda,\lambda_0)$. Conversely, every centralizer element $c \in \mathscr{C}$ determines a function on $\mathscr{L}$ defined by

$$f_c: \mathscr{L} \to \mathbb{C}, \quad (\lambda,\lambda_0) \to \chi_{\lambda,\lambda_0}(c). \tag{28}$$

We call the algebra homomorphism determined by Eq. (27) a *generalized* (or *extended*) infinitesimal character. This definition extends the definition of Harish-Chandra[29] for infinitesimal characters over the center Z of the universal enveloping algebra U of $\mathrm{Sp}(n)$ to infinitesimal characters over the centralizer $\mathscr{C}$. It turns out that the form of the (polynomial) function (28) depends explicitly on the *signature* of a lexical weight (λ,λ_0) which is defined as follows.

Definition 1: The signature of an extended weight $(\lambda,\lambda_0) \in H^*\times H_0^*$ is defined to be the collection of h integers,

$$\mathbf{s} = (s_1, s_2, \ldots, s_h),$$

defined by

$$s_r = \mathrm{sgn}(\lambda_{0_r} - \lambda_{r+1}), \quad 1 \leqslant r \leqslant h \tag{29}$$

with $\mathrm{sgn}(0) = 1$. We denote the set of all possible signatures by S: clearly, $|S| = 2^h$.

We say that two extended weights (λ,λ_0), (μ,μ_0) are equivalent (mod S) (or are *S equivalent*) if and only if they have the same signature.

Definition 1 determines an equivalence relation on $H^*\times H_0^*$ and partitions $H^*\times H_0^*$ into 2^h distinct equivalence classes, herein referred to as *S classes*. Throughout this paper we denote the set of lexical weights with signature $\mathbf{s}$ simply by $\mathscr{L}_{\mathbf{s}}$.

Signatures play a fundamental role in the remaining two papers and are intimately connected with the (generalized) Weyl group symmetries of $\mathrm{Sp}(n)$: $\mathrm{Sp}(n-2)\times\mathrm{Sp}(2)$ reduced Wigner coefficients and reduced matrix elements.[10,12]

We wish now to generalize the above procedure down the subgroup chain

$$\mathrm{Sp}(n) \supset \mathrm{Sp}(n-2)\times\mathrm{Sp}(2) \supset \mathrm{Sp}(n-4)\times\mathrm{Sp}(2) \times\mathrm{Sp}(2) \supset \cdots \supset \mathrm{Sp}(4)\times\mathrm{Sp}(2)\times\cdots\times\mathrm{Sp}(2) \supset \mathrm{G}, \tag{30}$$

where G denotes the subgroup

$$\mathrm{G} = \mathrm{Sp}(2)\times\mathrm{Sp}(2)\times\cdots\times\mathrm{Sp}(2) \ (h+1\text{-fold product}). \tag{31}$$

The corresponding $\mathrm{Sp}(2)$ subgroups, denoted $\mathrm{Sp}_m(2)$ $(1 \leqslant m \leqslant h+1)$, have infinitesimal generators given by

$$h_m = \alpha_{2m-1}^{2m-1}, \quad \alpha_{2m}^{2m-1}, \quad \alpha_{2m-1}^{2m}.$$

We now consider states symmetry adapted to the subgroup chain (30), for which the $\mathrm{Sp}_m(2)$ algebras have representation labels which are maximal for given $\mathrm{Sp}(2m)$ and $\mathrm{Sp}(2m-2)$ representation labels $(1 \leqslant m \leqslant h+1)$. Such states may be represented by a (λ) pattern:

$$|(\lambda)\rangle_\omega = \left| \begin{array}{ccccc} \lambda_{1h+1} & \lambda_{2h+1} & \cdots & \lambda_{hh+1} & \lambda_{h+1\,h+1} \\ \lambda_{1h} & \lambda_{2h} & \cdots & \lambda_{hh} & \\ & & \iddots & & \\ \lambda_{12} & \lambda_{22} & & & \\ \lambda_{11} & & & & \end{array} \right\rangle_\omega, \tag{32}$$

where the λ_{ij} satisfy the betweenness conditions [cf. Eq. (15)]

$$\lambda_{r\,m+1} \geqslant \lambda_{rm} \geqslant \lambda_{r+2\,m+1} \ (1 \leqslant r \leqslant m-1),$$
$$\lambda_{mm+1} \geqslant \lambda_{mm} \geqslant 0, \quad \lambda_{1m} \geqslant \lambda_{2m} \geqslant \cdots \geqslant \lambda_{mm} \geqslant 0, \tag{33}$$

and where $\omega = (\omega_1, \ldots, \omega_{h+1})$ is the weight of the state. As usual, the rows of the pattern (λ) correspond to the highest weights of the irreducible representations of the subgroups $\mathrm{Sp}(2m)$ $(1 \leqslant m \leqslant h+1)$, with the top row corresponding to the highest weight of the irreducible $\mathrm{Sp}(n)$ module under consideration. Since the $\mathrm{Sp}_m(2)$ representation labels corresponding to state (32) are to be maximal, for given $\mathrm{Sp}(2m)$ labels $\boldsymbol{\lambda}_m$ and $\mathrm{Sp}(2m-2)$ labels $\boldsymbol{\lambda}_{m-1}$, the $\mathrm{Sp}_m(2)$ representation labels corresponding to state (32) are given explicitly by [cf. Eq. (25)]

$$\Omega_m = \lambda_{1,m} - \sum_{r=1}^{m-1} |\lambda_{r,m-1} - \lambda_{r+1,m}|, \quad 2 \leqslant m \leqslant h+1,$$

with

$$\Omega_1 = \lambda_{11}.$$

It follows that the components of the weight ω must satisfy

$$\omega_m = \Omega_m, \Omega_m - 2, \ldots, -\Omega_m + 2, -\Omega_m. \tag{34}$$

Following Eq. (29), we associate with each pattern (λ) the signature array

$$(s_{rm}) = \begin{pmatrix} s_{1h} & s_{2h} & \cdots & s_{hh} \\ s_{1h-1} & s_{2h-1} & \cdots & s_{h-1\,h-1} \\ \vdots & & & \\ s_{12} & s_{22} & & \\ s_{11} & & & \end{pmatrix}, \tag{35}$$

where $s_{rm} = \mathrm{sgn}\,(\lambda_{rm} - \lambda_{r+1\,m+1})$, $1 \leqslant m \leqslant h$, $1 \leqslant r \leqslant m$, with $\mathrm{sgn}(0) = 1$. The rows $\mathbf{s}_m$ of the array of signatures (35) clearly determine the signatures of the lexical weights $(\boldsymbol{\lambda}_{m+1}, \boldsymbol{\lambda}_m)$, in the sense of Definition 1, for each pair of canonical subgroups $\mathrm{Sp}(2m+2)$, $\mathrm{Sp}(2m)$ $(1 \leqslant m \leqslant h)$ of $\mathrm{Sp}(n)$.

Throughout this paper we refer to the states (32) as *canonical states* since they play a role in the representation theory of Sp(n) equivalent to the canonical GT states familiar in the representation theory of the orthogonal and unitary groups.[9,30] We denote the subspace of $V(\lambda)$ spanned by the canonical states (32) simply by $V_0(\lambda)$. The remainder of this paper is devoted to the structure of the space $V_0(\lambda)$ and the properties of canonical states.

We note that the set of all GK states (21) with a fixed (upper) (λ) pattern constitutes a reducible representation of the subgroup G of Eq. (31), herein denoted $V^{(\lambda)}$: The action of the generators of the subgroup G on the GK basis states (21) is given in Ref. 16. The canonical states $|(\lambda)\rangle_\omega$ are uniquely characterized as spanning that irreducible G submodule of $V^{(\lambda)}$ whose G representation labels are maximal. We denote this unique G submodule of $V^{(\lambda)}$ by $V_0^{(\lambda)}$. We thus obtain the following orthogonal G-module decomposition for the space $V_0(\lambda)$ of canonical states:

$$V_0(\lambda) = \underset{\nu}{\oplus} V_0^{(\nu)},$$

where the sum is over all allowable upper patterns (ν).

Definition 2: We say that the canonical state (32) is *G maximal* if the weight labels take maximum allowed values $\omega_i = \Omega_i$ $(1 \leqslant i \leqslant h+1)$. More generally, we say that state (32) is *G extremal* if the G-weight labels take extreme values $\omega_i = \pm \Omega_i$ $(1 \leqslant i \leqslant h+1)$.

We denote the above G-extremal states by

$$|(\lambda)\rangle_\epsilon, \tag{36}$$

where the components of the vector $\epsilon = (\epsilon_1,\ldots,\epsilon_{h+1})$ are determined by $\epsilon_i = \pm 1$ according to whether $\omega_i = \pm \Omega_i$, respectively. The extremal states (36) are uniquely characterized by the fact that their weights occur with unit multiplicity in the subspace $V^{(\lambda)}$. In the maximal case $\epsilon_i = 1$ $(1 \leqslant i \leqslant h+1)$ we denote the state (36) by the special convention

$$|(\lambda)\rangle_+ \tag{37}$$

while in the minimal case $\epsilon_i = -1$ $(1 \leqslant i \leqslant h+1)$ we denote the state (36) by

$$|(\lambda)\rangle_-. \tag{38}$$

Clearly the state (37) [respectively, (38)] is the G-maximal (respectively, minimal) weight state of the irreducible G module $V_0^{(\lambda)}$.

Associated with the G-extremal state (36) is the GK state

$$\left| \begin{matrix} (\lambda) \\ (\text{ext}) \end{matrix} \right\rangle \tag{38'}$$

whose lower (μ) pattern is extremely connected to the pattern (λ) and is determined by the vector ϵ in accordance with $(1 \leqslant m \leqslant h)$:

$$\mu_{1\,m+1} = \lambda_{1\,m+1}, \mu_{r+1\,m+1} = \lambda_{r+1\,m+1} \wedge \lambda_{rm} (1 \leqslant r \leqslant h), \quad \epsilon_{m+1} = +1,$$

$$\mu_{m+1\,m+1} = 0, \mu_{r\,m+1} = \lambda_{r+1\,m+1} \vee \lambda_{rm} (1 \leqslant r \leqslant h), \quad \epsilon_{m+1} = -1,$$

with

$$\mu_{11} = \begin{cases} \lambda_{11}, & \epsilon_1 = +1, \\ 0, & \epsilon_1 = -1, \end{cases}$$

i.e., the entries of the lower (μ) pattern take maximum or minimum values. We denote this extremal (μ) pattern by the special convention $(\lambda)_\epsilon$ and we denote the GK state (38') by

$$\left| \begin{matrix} (\lambda) \\ (\lambda)_\epsilon \end{matrix} \right\rangle. \tag{39}$$

We note that the pattern $(\lambda)_\epsilon$ depends on signature (35) of the pattern (λ), as well as on the vector ϵ, since

$$\lambda_{r+1\,m+1} \wedge \lambda_{rm} = \begin{cases} \lambda_{r+1\,m+1}, & s_{rm} = +1, \\ \lambda_{rm}, & s_{rm} = -1, \end{cases} \qquad 1 \leqslant m \leqslant h$$

$$\lambda_{r+1\,m+1} \vee \lambda_{rm} = \begin{cases} \lambda_{r+1\,m+1}, & s_{rm} = -1, \\ \lambda_{rm}, & s_{rm} = +1. \end{cases}$$

In the maximal case $\epsilon_i = 1$ $(1 \leqslant i \leqslant h+1)$ we denote the pattern $(\lambda)_\epsilon$ by the special convention $(\lambda)_+$; similarly, when $\epsilon_i = -1$ $(1 \leqslant i \leqslant h+1)$ we denote this pattern by $(\lambda)_-$.

By our construction the GK basis state (39) has the same weight as the extremal state (36) and, since this weight occurs with unit multiplicity in $V^{(\lambda)}$, states (36) and (39) must coincide (up to scalar multiples). All remaining canonical states may be obtained from the extremal states (36) and (39) by application of raising or lowering operators from the group G. From the point of view of overlap coefficients,[16] the above shows that the GK states (39) are orthogonal to all remaining GK states of the space $V(\lambda)$.

The above determines precisely the space of canonical states $V_0(\lambda)$ with which we are working. Although only a subspace of the irreducible Sp(n) module $V(\lambda)$, the space $V_0(\lambda)$ nevertheless contains a relatively large number of states, as can be seen from the following.

Theorem 1: (i) All distinct weights of $V(\lambda)$ occur in $V_0(\lambda)$ at least once.

(ii) In particular, all states with weight W conjugate to the highest weight occur in $V_0(\lambda)$: This includes the maximal and minimal weight vectors.

(iii) If all weights occur in $V(\lambda)$ with unit multiplicity we must have $V(\lambda) = V_0(\lambda)$.

Proof: Part (iii) of Theorem 1 follows immediately from part (i), as does (ii), if we take into account the fact[29] that all weights W conjugate to the highest weight λ must occur with unit multiplicity in $V(\lambda)$. In reference to part (i) we note that every weight state in $V(\lambda)$ must be a linear combination of weight vectors from each $V^{(\lambda)}$. On the other hand, it follows from the maximal nature of the space $V_0^{(\lambda)}$ that the vectors in $V^{(\lambda)}$ have weights of the form $\nu = (\nu_1,\nu_2,\ldots,\nu_{h+1})$, where [cf. Eq. (34)]

$$\Omega_m \geqslant \nu_m \geqslant -\Omega_m \quad \text{(in steps of two)},$$

which equals the weight of the canonical state $|(\lambda)\rangle_\nu$. This is enough to establish the result.

Q.E.D.

In addition to the results above, we note that in the degenerate cases $\lambda = (p,0)$ or (p), considered in the work of Wong and Yeh,[19] we necessarily have $V_0(\lambda) = V(\lambda)$. Thus

our results on canonical matrix elements include those of Wong and Yeh[19] as a special case.

V. S-DOMINANT (λ) PATTERNS

In the work of Gould and Kalnins[16] it was shown that the Sp(n) maximal weight state of the space $V(\lambda)$ is given by the U(n) GT state (notation as in Secs. III and IV)

$$|\overset{\lambda}{(\max)}\rangle_+ = \left|\begin{matrix}\overset{\lambda}{(\max)}\\ \overset{\lambda}{(\max)}\end{matrix}\right\rangle = \left|\begin{matrix}\overset{\lambda}{(\max)}\\ \overset{\lambda}{(\max)}\end{matrix}\right) \in V(\lambda) \subseteq V(\hat{\lambda})$$

whose upper and lower patterns take maximal allowed values. In the single-pattern notation of Eq. (19), the above U(n) GT state may be expressed as

$$\left|\begin{matrix} \lambda_1 & \lambda_2 & \cdots & \lambda_h & \lambda_{h+1} \\ \lambda_1 & \lambda_2 & \cdots & \lambda_h & \lambda_{h+1} \\ \lambda_1 & \lambda_2 & \cdots & \lambda_h & \\ \lambda_1 & \lambda_2 & \cdots & \lambda_h & \\ \vdots & & & & \\ \lambda_1 & \lambda_2 & & & \\ \lambda_1 & & & & \\ \lambda_1 & & & & \end{matrix}\right)$$

and it is easily verified that it indeed constitutes an Sp(n) maximal weight state of weight λ. It is our aim in this section to demonstrate that there is a relatively large class of canonical states which are represented by U(n) GT states: In such a case the U(n) GT states (20) [or (19)] must coincide with the projected states (21).

The signature **s** of a lexical weight $(\lambda,\lambda_0) \in \mathscr{L}$ was defined in Sec. IV (see Definition 1): We denote the space of all lexical weights of signature **s** by $\mathscr{L}_{\mathbf{s}}$. In particular, we have the lexical weights with maximal signature:

$$(\dot{1}) = (1,1,\ldots,1).$$

We call lexical weights of maximal signature *S dominant*: They play an important role in the remaining two papers since such lexical weights have similar properties to lexical weights[12] for the normal canonical imbeddings U(n) $\supset$ U($n-1$), O(n) $\supset$ O($n-1$). Because of the special nature of this case we denote the space $\mathscr{L}_{(\dot{1})}$ of *S*-dominant lexical weights by $\mathscr{L}_+$. We similarly have the space $\mathscr{L}_-$ of lexical weights with minimal signature $(-1,-1,\ldots,-1)$. By definition it follows that an extended weight $(\lambda,\lambda_0) \in \Lambda^+ \times \Lambda_0^+$ is *S* dominant if and only if the components of λ and λ_0 satisfy the betweenness conditions

$$\lambda_1 \geqslant \lambda_{0_1} \geqslant \lambda_2 \geqslant \lambda_{0_2} \geqslant \cdots \geqslant \lambda_{0_h} \geqslant \lambda_{h+1} \geqslant 0. \tag{40}$$

We note that such an extended weight satisfies the inequalities (15) and hence is automatically lexical.

In terms of signatures, we note that the maximum Sp(2) label of Eq. (25) may be expressed as

$$\Omega_{\lambda,\lambda_0} = \lambda_1 + \sum_{r=1}^{h} s_r(\lambda_{r+1} - \lambda_{0_r}), \tag{41}$$

where $\mathbf{s} = (s_1,\ldots,s_h)$ is the signature of the lexical weight (λ,λ_0). Thus if (λ,λ_0) is *S* dominant we have

$$\Omega_{\lambda,\lambda_0} = \int(\lambda) - \int(\lambda_0), \quad (\lambda,\lambda_0) \in \mathscr{L}_+, \tag{42}$$

with $\int(\lambda)$ as in Eq. (22).

Definition 3: Proceeding down the subgroup chain (30) we say that the canonical state $|(\lambda)\rangle_\omega$ of Eq. (32) is *S* dominant if, in addition to Eq. (33), the entries of the pattern (λ) satisfy the inequalities

$$\lambda_{rm} \geqslant \lambda_{r+1,m+1}, \quad 1 \leqslant r \leqslant m, \quad 1 \leqslant m \leqslant h. \tag{43}$$

Throughout this paper we denote the space of *S*-dominant canonical states by $V_+(\lambda) \subseteq V_0(\lambda)$.

Clearly the above definition of *S*-dominant (λ) patterns is equivalent to requiring that the lexical weights $(\lambda_m,\lambda_{m-1})$ be *S* dominant for each pair of canonical subgroups Sp($2m$), Sp($2m-2$) ($2 \leqslant m \leqslant h+1$) of Sp($n$). In such a case the representation labels of the subgroups Sp$_m$(2) ($1 \leqslant m \leqslant h+1$) are uniquely determined by

$$\Omega_m = \int(\lambda_m) - \int(\lambda_{m-1}).$$

It follows that if the pattern (λ) is *S* dominant, then the extremal state $|(\lambda)\rangle_\epsilon$ has the weight ω given by

$$\omega_1 = \pm\lambda_{11}, \quad \omega_m = \pm\left[\int(\lambda_m) - \int(\lambda_{m-1})\right],$$

$$2 \leqslant m \leqslant h+1$$

according to whether $\epsilon_m = \pm 1$, respectively ($1 \leqslant m \leqslant h+1$).

The importance of extremal *S*-dominant canonical states lies in the following result [notation as in Eq. (39)].

Theorem 2: Suppose the pattern (λ) is *S* dominant. Then the extremal canonical state $|(\lambda)\rangle_\epsilon$ is given by the U(n) GT state

$$|(\lambda)\rangle_\epsilon = \left|\begin{matrix}(\lambda)\\ (\lambda)_\epsilon\end{matrix}\right). \tag{44}$$

□

Before going on to the proof of Theorem (2) we denote the allowable U(n) GT state on the rhs of Eq. (44) by $|(\lambda)_\epsilon\rangle$ and note that this state may be expressed in the single-pattern notation of Eq. (19) according to

$$|(\lambda)_\epsilon) = \left|\begin{matrix} \lambda_{1\,h+1} & \cdots & \lambda_{h\,h+1} & \lambda_{h+1\,h+1} \\ \lambda'_{1\,h+1} & \cdots & \lambda'_{h\,h+1} & \lambda'_{h+1\,h+1} \\ \lambda_{1h} & \cdots & \lambda_{hh} & \\ \lambda'_{1h} & \cdots & \lambda'_{hh} & \\ \vdots & & \ddots & \\ \lambda_{12} & \lambda_{22} & & \\ \lambda'_{12} & \lambda'_{22} & & \\ \lambda_{11} & & & \\ \lambda'_{11} & & & \end{matrix}\right), \tag{45}$$

where

$$\lambda'_{rm}=\begin{cases}\lambda_{rm}, & (1\leqslant r\leqslant m);\quad \epsilon_m=+1,\\ \lambda_{rm-1}, & (1\leqslant r\leqslant m-1);\quad 0,\ (r=m);\quad \epsilon_m=-1.\end{cases}\qquad 1\leqslant m\leqslant h+1$$

In particular, for the maximal case $\epsilon_i=1$ $(1\leqslant i\leqslant h+1)$ the above U(n) GT state reduces to

$$|(\lambda)_+)=\begin{vmatrix}\lambda_{1\,h+1} & \cdots & \lambda_{h\,h+1} & \lambda_{h+1\,h+1}\\ \lambda_{1\,h+1} & \cdots & \lambda_{h\,h+1} & \lambda_{h+1\,h+1}\\ \lambda_{1\,h} & \cdots & \lambda_{hh} & \\ \lambda_{1\,h} & \cdots & \lambda_{hh} & \\ \vdots & \iddots & & \\ \lambda_{12} & \lambda_{22} & & \\ \lambda_{11} & 0 & & \\ \lambda_{11} & & & \end{vmatrix}\qquad (46)$$

The main step in the proof of Theorem 2 is the following result (notation as above).

Lemma 2: Suppose $(\lambda,\lambda_0)\in\mathscr{L}_+$. Then the canonical module $V(\lambda,\lambda_0)\subseteq V(\lambda)$ is the unique irreducible Sp($n-2$)×Sp(2) module with the highest weight $\lambda_0\times[\,f(\lambda)-f(\lambda_0)]$ occurring in the irreducible U(n) module $V(\hat{\lambda})$.

Proof: Suppose $V_{\lambda,\lambda_0}\subseteq V(\hat{\lambda})$ is an Sp($n-2$)×Sp(2) irreducible module with the highest weight $\lambda_0\times[\,f(\lambda)-f(\lambda_0)]$. We show that $V_{\lambda,\lambda_0}\subseteq V(\lambda)$. To see this we note that the only other irreducible Sp(n) modules $V(\mu)$ occurring in $V(\hat{\lambda})$ have highest weights μ obtained from λ via contraction: For such a case we have $\lambda_r\geqslant\mu_r$ $(1\leqslant r\leqslant h+1)$. It follows that if $V(\mu)\subseteq V(\hat{\lambda})$ is such that $(\mu,\lambda_0)\in\mathscr{L}$ [i.e., $V(\lambda_0)$ is contained in $V(\mu)$], then

$$\lambda_{0_r}\geqslant\lambda_{r+1}\geqslant\mu_{r+1},\quad 1\leqslant r\leqslant h,$$

so that (μ,λ_0) is also S dominant. Thus the maximum possible Sp(2) label for given $(\mu,\lambda_0)\in\mathscr{L}_+$ is given by

$$\int(\mu)-\int(\lambda_0)\leqslant\int(\lambda)-\int(\lambda_0),$$

with equality if and only if $\mu=\lambda$. This shows the required inclusion $V_{\lambda,\lambda_0}\subseteq V(\lambda)$. On the other hand, we have already seen in Sec. III that the canonical submodule $V(\lambda,\lambda_0)$ is the unique irreducible Sp($n-2$)×Sp(2) submodule of $V(\lambda)$ with highest weight $\lambda_0\times[\,f(\lambda)-f(\lambda_0)]$, from which we obtain the result.

Q.E.D.

Corollary: Suppose $(\lambda,\lambda_0)\in\mathscr{L}_+$ and $\Psi\in V(\hat{\lambda})$ transforms as a state in the irreducible Sp($n-2$) module $V(\lambda_0)$. If

$$h_{h+1}\,\Psi=\pm\left[\int(\lambda)-\int(\lambda_0)\right]\Psi,$$

then $\Psi\in V(\lambda,\lambda_0)\subseteq V(\lambda)$.

Proof: From the proof of Lemma 1 we know that given λ_0 as above, the maximum possible Sp(2) representation label occurring in $V(\hat{\lambda})$ is $\Omega=f(\lambda)-f(\lambda_0)$. Thus if Ψ satisfies the conditions of the lemma it follows, from maximality of Ω, that Ψ can only belong to the irreducible Sp($n-2$)×Sp(2) module with the highest weight $\lambda_0\times\Omega$. On the other hand, from Lemma 2, $V(\lambda,\lambda_0)\subseteq V(\lambda)$ is the unique irreducible Sp($n-2$)×Sp(2) module with highest weight $\lambda_0\times\Omega$ occurring in $V(\hat{\lambda})$, from which the result follows.

Q.E.D.

We are now in a position to prove Theorem 2 by recursion down the subgroup chain (30). By the inductive hypothesis we assume that the GT state obtained from the state (45) by omission of the top two rows constitutes an extremal canonical state of Sp($n-2$) belonging to the irreducible Sp($n-2$) module $V(\lambda_h)$. However, for $\epsilon_{h+1}=\pm1$ the state (45) has Sp(2) weight $\pm[f(\lambda_{h+1})-f(\lambda_h)]$, respectively. Thus the conditions of the above corollary are satisfied, hence the GT state (45) constitutes a state in the Sp(n) module $V(\lambda_{h+1})$, which is symmetry adapted to the subgroup chain (30) and whose G-representation labels are all maximal. This is enough to ensure that the state (45) determines a canonical state of Sp(n) (which is necessarily S dominant and G extremal), as required. This completes the proof of the theorem.

Theorem 2 shows that all G-extremal S-dominant canonical states belong to the space $A(\lambda)$ of allowable U(n) GT states. All remaining S-dominant canonical states may be obtained from the extremal states by application of raising or lowering operators from the subgroup G. On the other hand, Gould and Kalnins[16] have shown that the space of allowable U(n) GT states $A(\lambda)$ is stable under the action of the subgroup G. It follows that the space $V_+(\lambda)$ of S-dominant canonical states is contained in the space $A(\lambda)$. In fact, we conjecture the result:

$$A(\lambda)\cap V(\lambda)=V_+(\lambda).$$

We note that the maximal state of the Sp(n) module $V(\lambda)$ is a G-maximal S-dominant canonical state and hence is represented by a U(n) GT state, as noted earlier. We conclude this section by showing that all basis states with weight Weyl group conjugate to the highest weight belong to the space of G-extremal, S-dominant canonical states and hence are represented by U(n) GT states.

For each index $m\in\{1,\ldots,h+1\}$ we may define the Sp($n-2$) weight

$$\lambda(m)=(\lambda_1,\ldots,\lambda_{m-1},\lambda_{m+1},\ldots,\lambda_h,\lambda_{h+1}),$$

which is obtained from the Sp(n) maximal weight by omission of the mth component. Similarly, for each pair of indices $1\leqslant m\neq k\leqslant h$ we have the Sp($n-4$) weight ($m<k$, assumed)

$$\lambda(m,k)=(\lambda_1,\ldots,\lambda_{m-1},\lambda_{m+1},\ldots,\lambda_{k-1},\lambda_{k+1},\ldots,\lambda_{h+1}).$$

More generally, for any set of k distinct indices $1\leqslant i_1,\ldots,i_k\leqslant h+1$ we have the Sp($n-2k$) weight $\lambda(i_1,\ldots,i_k)$, which is obtained from the highest weight λ by omission of the components $i_1,i_2,\ldots,i_k$. We note that the weight $\lambda(i_1,\ldots,i_k)$

constitutes the highest weight of an irreducible $\mathrm{Sp}(n-2k)$ module.

Now let $(i_1,i_2,\ldots,i_{h+1})$ be any permutation of the numbers $(1,2,\cdots,h+1)$ and consider the $\mathrm{Sp}(2m)$ representation labels λ_m defined by

$$\lambda_{h+1}=\lambda,\lambda_m=\lambda(i_{m+1},\ldots,i_h,i_{h+1}),\quad 1\leqslant m\leqslant h.$$

Then these vectors determine the rows of an S-dominant (λ) pattern

$$(\lambda)=\begin{pmatrix}\lambda_{h+1}\\ \lambda_h\\ \vdots\\ \lambda_2\\ \lambda_1\end{pmatrix}$$

and the corresponding G-maximal state $|(\lambda)\rangle_+$ has weight

$$(\lambda_{i_1},\lambda_{i_2},\ldots,\lambda_{i_{h+1}}),$$

which is W conjugate to the highest weight. More generally, the G-extremal states $|(\lambda)\rangle_\epsilon$ have weight

$$(\epsilon_1\lambda_{i_1},\epsilon_2\lambda_{i_2},\ldots,\epsilon_{h+1}\lambda_{i_{h+1}}),\quad \epsilon_i=\pm 1,$$

which is necessarily W conjugate to the highest weight λ. This shows that all (normalized) $\mathrm{Sp}(n)$ states with weight W conjugate to the highest weight are G-extremal S-dominant canonical states and hence, by Theorem 2, are given by (allowed) $\mathrm{U}(n)$ GT states.

The above results show that a relatively large class of states are simply represented by allowable $\mathrm{U}(n)$ GT states. This then enables us to determine the action of the elementary $\mathrm{Sp}(n)$ generators on such states by exploiting the known action of the $\mathrm{U}(n)$ generators on GT states.

VI. ACTION OF Sp(n) GENERATORS ON S-DOMINANT CANONICAL STATES

It is the purpose of this section to derive the matrix elements of the elementary $\mathrm{Sp}(n)$ generators (9) between S-dominant canonical states by exploiting the known matrix element formulas of the $\mathrm{U}(n)$ generators. This work, of interest in its own right, also serves the role of providing a detailed check on our later matrix element formulas to be developed for the $\mathrm{Sp}(n)$ generators between arbitrary canonical basis states.

We note that the elementary generators α_n^{n-1}, α_{n-1}^n constitute the generators of the subgroup $\mathrm{Sp}(2)$ of $\mathrm{Sp}(n)$ and hence their action on arbitrary canonical states $|(\lambda)\rangle_\omega$ is easily determined from the known $\mathrm{Sp}(2)$ matrix element formulas (see Appendix B):

$$\alpha_n^{n-1}|(\lambda)\rangle_\omega=\sqrt{(\Omega-\omega_{h+1})(\Omega+\omega_{h+1}+2)}\,|(\lambda)\rangle_{w+2\Delta_{h+1}},\tag{47}$$

where $\Omega=f(\lambda_{h+1})-f(\lambda_h)$ is the $\mathrm{Sp}_{h+1}(2)$ representation label. Thus we concentrate here on obtaining the matrix elements of the elementary generators α_{2i+1}^{2i-1}, α_{2i-1}^{2i+1} on S-dominant canonical states.

We recall that our $\mathrm{Sp}(n)$ generators α_l^k may be expressed in terms of the $\mathrm{U}(n)$ generators a_{ij} according to (cf. Appendix A)

$$\alpha_l^k=a_{k,l}-(-1)^{k+l}a_{\bar l\bar k},\quad 1\leqslant k,\quad l\leqslant n,$$

where $\bar k$ is given by Eq. (7). In particular, we have the generators

$$\alpha_{2j-1}^{2i-1}=a_{2i-1,2j-1}-a_{2j,2i},\quad 1\leqslant i,\quad j\leqslant h+1.\tag{48}$$

Now let $|(\lambda)\rangle_+$ be a G-maximal S-dominant canonical state which [cf. Theorem 2] is represented by the $\mathrm{U}(n)$ GT state (46). It is easily verified from the known[30] action of the $\mathrm{U}(n)$ generators that the even $\mathrm{U}(n)$ generators

$$a_{2j,2i},\quad 1\leqslant i,\quad j\leqslant h+1$$

necessarily vanish on the $\mathrm{U}(n)$ GT state (46). Thus the action of the $\mathrm{Sp}(n)$ generators (48) on the state $|(\lambda)\rangle_+$ reduces to the action of the $\mathrm{U}(n)$ generator $a_{2i-1,2j+1}$ on the GT state (46), which follows from the known $\mathrm{U}(n)$ matrix element formulas. We obtain immediately the result (cf. Appendix C)

$$\alpha_{2m+1}^{2m-1}|(\lambda)\rangle_+=a_{2m+1}^{2m-1}|(\lambda)\rangle_+=\sum_{r=1}^m N_r^m|(\lambda+\Delta_r^m)\rangle_+,\tag{49}$$

where $(\lambda+\Delta_r^m)$ denotes the pattern obtained from (λ) by increasing the representation label λ_{rm} by one unit and leaving the remaining pattern labels unchanged. The coefficients N_r^m appearing in expansion (49) are given by

$$N_r^m=\left((-1)^m\frac{\prod_{p=1}^{m+1}(\lambda_{p\,m+1}-\lambda_{rm}+r-p)\prod_{q=1}^{m-1}(\lambda_{rm}-\lambda_{qm-1}+q-r+1)}{\prod_{\substack{p=1\\ \neq r}}^{m}(\lambda_{rm}-\lambda_{pm}+p-r)(\lambda_{rm}-\lambda_{pm}+p-r+1)}\right)^{1/2}.\tag{50}$$

Similarly, for the lowering elementary generators α_{2m-1}^{2m+1}, we obtain (see Appendix C)

$$\alpha_{2m-1}^{2m+1}|(\lambda)\rangle_+=a_{2m-1}^{2m+1}|(\lambda)\rangle_+=\sum_{r=1}^m \bar N_r^m|(\lambda-\Delta_r^m)\rangle_+,\tag{51}$$

where

$$\bar{N}_r^m = \left((-1)^m \frac{\prod_{p=1}^{m+1}(\lambda_{pm+1} - \lambda_{rm} + r - p + 1) \prod_{q=1}^{m-1}(\lambda_{rm} - \lambda_{qm-1} + q - r)}{\prod_{\substack{p=1 \\ \neq r}}^{m}(\lambda_{rm} - \lambda_{pm} + p - r)(\lambda_{rm} - \lambda_{pm} + p - r - 1)} \right)^{1/2}. \tag{52}$$

Our analysis is simplified by noting that the Sp(n) generators

$$\beta_j^i = \alpha_{2j-1}^{2i-1}, \quad 1 \leqslant i, \quad j \leqslant h+1 \tag{53}$$

constitute the generators of a U($h+1$) subgroup of Sp($n = 2h+2$). Also, as seen from Eqs. (49) and (51), the elementary U($h+1$) generators

$$\beta_{m+1}^m = \alpha_{2m+1}^{2m-1}, \quad \beta_m^{m+1} = \alpha_{2m-1}^{2m+1} \tag{54}$$

take an S-dominant G-maximal canonical state $|(\lambda)\rangle_+$ to a linear combination of such states. Since every $U(h+1)$ generator β_j^i may be obtained by repeated commutation with the elementary generators (54) it follows that the space of G-maximal S-dominant states $|(\lambda)\rangle_+$ is to constitute a module over U($h+1$). Also, the requirement that the pattern (λ) be S dominant is equivalent to the requirement that (λ) be a U($h+1$) GT pattern[9] for the irreducible representation of U($h+1$) with highest weight $\lambda = \lambda_{h+1}$ (top row of pattern). Moreover, we note that the action of equations (49) and (51) and the matrix element formulas (50) and (52) coincide precisely with the usual U($h+1$) elementary matrix element formulas[30] (between GT states), viz.,

$$\beta_{m+1}^m |(\lambda)\rangle = \sum_{r=1}^m N_r^m |(\lambda + \Delta_r^m)\rangle,$$

$$\beta_m^{m+1} |(\lambda)\rangle = \sum_{r=1}^m \bar{N}_r^m |(\lambda - \Delta_r^m)\rangle,$$

where $|(\lambda)\rangle$ denotes a U($h+1$) GT state, with $N_r^m, \bar{N}_r^m$ as in Eqs. (50) and (52), respectively.

It follows that the space of S-dominant G-maximal canonical states is to constitute an irreducible module over U($h+1$) with the highest weight $\lambda = \lambda_{h+1}$. In particular, the nonzero matrix elements of the Sp(n) generators (49) between S-dominant G-maximal states are given by

$${}_+\langle(\lambda')|\alpha_{2j-1}^{2i-1}|(\lambda)\rangle_+ = \langle(\lambda')|\beta_j^i|(\lambda)\rangle, \tag{55}$$

which may be obtained from the known[31] matrix element formulas of the U($h+1$) generators.

Similarly, with regard to the G-minimal S-dominant canonical states (38) which are given by the U(n) GT states $|(\lambda)_-\rangle$, we note that the U(n) generators $a_{2i-1,2j-1}$ always vanish on the GT states (38). Hence the action of the Sp(n) generators (48) reduces to the action of the U(n) generators $-a_{2j,2i}$ acting on the U(n) GT states (38). In this case we obtain immediately (cf. Appendix C)

$$\alpha_{2m+1}^{2m-1}|(\lambda)\rangle_- = -a_{2m+2,2m}|(\lambda)\rangle_-$$

$$= -\sum_{r=1}^m \bar{N}_r^m |(\lambda - \Delta_r^m)\rangle_-, \tag{56}$$

$$\alpha_{2m-1}^{2m+1}|(\lambda)\rangle_- = -a_{2m,2m+2}|(\lambda)\rangle_-$$

$$= -\sum_{r=1}^m N_r^m |(\lambda + \Delta_r^m)\rangle_-, \tag{57}$$

with $N_r^m, \bar{N}_r^m$ as in Eqs. (50) and (52), respectively. It follows in this case that the space of S-dominant G-minimal canonical states $|(\lambda)\rangle_-$ is to constitute an irreducible module over U($h+1$), which is dual (or contragradient) to the representation afforded by the G-maximal S-dominant canonical states $|(\lambda)\rangle_+$. In particular, the nonzero matrix elements of the Sp(n) generators (48) between S-dominant G-minimal canonical states are given by

$${}_-\langle(\lambda')|\alpha_{2j-1}^{2i-1}|(\lambda)\rangle_- = -\langle(\lambda')|\beta_i^j(\lambda)\rangle$$

$$= -\langle(\lambda)|\beta_j^i|(\lambda')\rangle$$

[cf. Eq. (55)], where the rhs is given by the known[30] U($h+1$) matrix element formulas.

In this way we may obtain the action of the elementary Sp(n) generators on G-maximal or -minimal S-dominant canonical states. The action of these generators may similarly be determined on arbitrary S-dominant canonical states by expressing these states in terms of U(n) GT states [in accordance with Theorem 2] and using the known action of the U(n) generators. However, apart from the G-maximal and -minimal cases discussed above, this approach generally results in a linear combination of U(n) GT states which do not correspond to S-dominant canonical Sp(n) states, so that direct comparisons with our symmetry adapted formalism will be difficult. We shall not pursue this line of thought any further here since an alternative method for obtaining these matrix elements between arbitrary canonical (not necessarily S dominant) states will be developed in the subsequent papers of the present series.

VII. CONCLUSIONS

We have shown that every irreducible finite-dimensional representation of the symplectic group Sp(n) admits a relatively large number of states, herein referred to as canonical states, which possess properties analogous to the GT states for the orthogonal and unitary groups. In particular, the space of canonical states contains all distinct weights of the representation at least once. Moreover, in the S-dominant case it was shown that every G-extremal canonical state is given simply by a U(n) GT state and that the space of G-extremal S-dominant canonical states contains all states with weight Weyl group conjugate to the highest weight.

By exploiting the known action of the U(n) generators on U(n) GT states, the matrix elements of the elementary Sp(n) generators between G-maximal or -minimal S-dominant canonical states were determined. Unfortunately, however, this method is difficult to extend to the general canonical states and moreover, the structure of the matrix elements (as a product of a reduced matrix element and a Wigner coefficient) is not apparent in this approach. This deficiency will be removed in a subsequent paper, where an alternative direct evaluation of the Sp(n) generator matrix elements

between arbitrary canonical states will be given. The results of this paper will then provide a useful check on our later matrix element formulas.

ACKNOWLEDGMENTS

The author gratefully acknowledges the financial support of an ARGS research fellowship.

The author also thanks the referee for bringing to his attention the work of Wong and Yeh.[19]

APPENDIX A: COMMUTATION RELATIONS

In the opposite index notation of Eq. (7) our Sp(n) metric may be expressed as

$$g_{ij} = g^{\bar{j}i} = -(-i)^i \delta_{j\bar{i}},$$

i.e.,

$$g_{i\bar{i}} = g^{\bar{i}i} = -(-1)^i, \quad 1 \leqslant i \leqslant n,$$

where all other entries are zero. Thus for our Sp(n) generators (4), we obtain

$$\begin{aligned}\alpha^i_j &= g^{ik}\alpha_{kj} = g^{i\bar{i}}\alpha_{\bar{i}j} \\ &= g^{i\bar{i}}(g_{\bar{i}p}a_{pj} + g_{jp}a_{p\bar{i}}) \\ &= g^{i\bar{i}}(g_{\bar{i}\bar{i}}a_{ij} + g_{j\bar{j}}a_{\bar{j}\bar{i}}) \\ &= a_{ij} - (-1)^{i+j}a_{\bar{j}\bar{i}}, \quad 1 \leqslant i, \quad j \leqslant n. \end{aligned} \tag{A1}$$

We thus obtain, in particular,

$$\alpha^{2i-1}_{2j-1} = a_{2i-1,2j-1} - a_{2j,2i}, \quad 1 \leqslant i, \quad j \leqslant h+1.$$

In view of Eq. (A1) we obtain

$$\begin{aligned}\alpha^i_j &= a_{ij} - (-1)^{i+j}a_{\bar{j}\bar{i}} \\ &= -(-1)^{i+j}(a_{\bar{j}\bar{i}} - (-1)^{\bar{i}+\bar{j}}a_{ij}) \\ &= -(-1)^{i+j}\alpha^{\bar{j}}_{\bar{i}},\end{aligned}$$

which is the symmetry condition of Eq. (6). Finally, we note that the Sp(n) commutation relations (5) may be expressed in opposite index notation according to

$$[\alpha^i_j, \alpha^k_l] = \delta^k_j\alpha^i_l - \delta^i_l\alpha^k_j - (-1)^{i+j}(\delta^k_{\bar{i}}\alpha^{\bar{j}}_l - \delta^{\bar{j}}_l\alpha^k_{\bar{i}}),$$

as may be readily verified using Eq. (A1).

APPENDIX B: REPRESENTATION THEORY OF Sp(2)

The Sp(2) generators γ^μ_ν ($1 \leqslant \mu,\nu \leqslant 2$) of Eq. (10) satisfy the commutation relations (cf. Appendix A)

$$[\gamma^\mu_\nu, \gamma^\rho_\sigma] = \delta^\rho_\nu \gamma^\mu_\sigma - \delta^\mu_\sigma \gamma^\rho_\nu - (-1)^{\mu+\nu}(\delta^\rho_{\bar{\mu}} \gamma^{\bar{\nu}}_\sigma - \delta^{\bar{\nu}}_\sigma \gamma^\rho_{\bar{\mu}}),$$

where $\bar{\mu}$ is given by $\bar{\mu} = 1$ (resp. 2) according to whether $\mu = 2$ (resp. 1). The finite-dimensional irreducible representations of Sp(2) are uniquely characterized by the highest weight Ω, which is a non-negative integer. The corresponding representation space has dimension $\Omega + 1$ and admits a basis of weight vectors $|\Omega\omega\rangle$ ($\omega = \Omega, \Omega-2, \ldots, -\Omega+2, -\Omega$):

$$\gamma^1_1|\Omega\omega\rangle = -\gamma^2_2|\Omega\omega\rangle = \omega|\Omega\omega\rangle.$$

The second-order invariant

$$\sigma_2 = \sum_{\mu,\nu=1}^{2} \gamma^\mu_\nu \gamma^\nu_\mu$$

takes the constant value

$$\langle\sigma_2\rangle_\Omega = 2\Omega(\Omega+2)$$

in the irreducible representation with the highest weight Ω.

It follows that the nonzero matrix elements of the raising Sp(2) generator γ^1_2 are given by

$$\langle\Omega,\omega+2|\gamma^1_2|\Omega\omega\rangle = \langle\Omega\omega|\gamma^2_1\,\gamma^1_2|\Omega\omega\rangle^{1/2}.$$

Using

$$\begin{aligned}\sigma_2 &= (\gamma^1_1)^2 + (\gamma^2_2)^2 + \gamma^1_2\,\gamma^2_1 + \gamma^2_1\,\gamma^1_2 \\ &= 2(\gamma^1_1)^2 + 4\gamma^1_1 + 2\gamma^2_1\,\gamma^1_2,\end{aligned}$$

we thus obtain

$$\begin{aligned}\langle\Omega,\omega+2|\gamma^1_2|\Omega\omega\rangle &= \langle\Omega\omega|\tfrac{1}{2}\sigma_2 - \omega(\omega+2)|\Omega\omega\rangle^{1/2} \\ &= [(\Omega-\omega)(\Omega+\omega+2)]^{1/2},\end{aligned}$$

in agreement with Eq. (48).

APPENDIX C: MATRIX ELEMENTS

In the dual-pattern notation of Eq. (20), the nonzero matrix elements of the U(n) generator $a_{2m-1,2m+1}$ between allowed GT states are given by

$$\begin{aligned}N^m_{r,l} &= \begin{pmatrix}(\lambda+\Delta^m_r)\\(\mu+\Delta^m_l)\end{pmatrix}\Bigg|a_{2m-1,2m+1}\Bigg|\begin{pmatrix}(\lambda)\\(\mu)\end{pmatrix}, \quad 1 \leqslant r,l \leqslant m, \\ &= \mathrm{sn}(r-l)N^{2m}_r N^{2m-1}_l[(\lambda_{rm} - \mu_{lm} + l - r)(\lambda_{rm} - \mu_{lm} + l - r + 1]^{-1/2},\end{aligned} \tag{C1}$$

where, in the notation of Gould,[30] N^{2m}_r, N^{2m-1}_l denote the matrix elements of the elementary U(n) generators $a_{2m,2m+1}$, $a_{2m-1,2m}$, respectively. In dual-pattern notation, we obtain, from the formulas of Ref. 30, the results

$$N^{2m}_r = (\lambda_{rm} + m + 1 - r)^{-1/2}\left[\frac{(-1)^m \prod^{m+1}_{p=1}(\mu_{pm+1} - \lambda_{rm} + r - p)\prod^m_{q=1}(\lambda_{rm} - \mu_{qm} + q - r + 1)}{\prod^m_{\substack{q=1\\ \neq r}}(\lambda_{rm} - \lambda_{qm} + q - r)(\lambda_{rm} - \lambda_{qm} + q - r + 1)}\right]^{1/2},$$

$$N^{2m-1}_l = (\mu_{lm} + m + 1 - l)^{1/2}\left[\frac{(-1)^{m+1} \prod^m_{p=1}(\lambda_{pm} - \mu_{lm} + l - p)\prod^{m-1}_{q=1}(\mu_{lm} - \lambda_{qm-1} + q - l + 1)}{\prod^m_{\substack{q=1\\ \neq l}}(\mu_{lm} - \mu_{qm} + q - l)(\mu_{lm} - \mu_{qm} + q - l + 1)}\right]^{1/2},$$

from which we deduce

$$N^m_{r,l} = \mathrm{sn}(r-l)\left[(-1)\frac{\prod_{p=1}^{m+1}(\mu_{pm+1}-\lambda_{rm}+r-p)\prod_{\substack{q=1\\ \neq l}}^{m}(\lambda_{rm}-\mu_{qm}+q-r+1)}{\prod_{\substack{q=1\\ \neq r}}^{m}(\lambda_{rm}-\lambda_{qm}+q-r)(\lambda_{rm}-\lambda_{qm}+q-r+1)}\right.$$

$$\left.\times\frac{\prod_{\substack{p=1\\ \neq r}}^{m}(\lambda_{pm}-\mu_{lm}+l-p)\prod_{q=1}^{m-1}(\mu_{lm}-\lambda_{qm-1}+q-l+1)}{\prod_{\substack{q=1\\ \neq l}}^{m}(\mu_{lm}-\mu_{qm}+q-l)(\mu_{lm}-\mu_{qm}+q-l+1)}\right]^{1/2}\left(\frac{\mu_{lm}+m+1-l}{\lambda_{rm}+m+1-r}\right)^{1/2}. \tag{C2}$$

To determine the action of the $U(n)$ generators $a_{2m-1,2m+1}$ on the G-maximal S-dominant state $|(\lambda)\rangle_+$ we observe, in view of Theorem 2, that this state is given by the $U(n)$ GT state (46), which may be expressed in dual-pattern notation according to

$$|(\lambda)\rangle_+ = \begin{vmatrix}(\lambda)\\(\lambda)\end{vmatrix},$$

where the lower (μ) pattern coincides precisely with the upper pattern (λ). It follows, in view of the Eq. (C1), that the action of the generator $a_{2m-1,2m+1}$ on the above state is given by

$$a_{2m-1,2m+1}\begin{vmatrix}(\lambda)\\(\lambda)\end{vmatrix} = \sum_{r=1}^{m} N^m_r \begin{vmatrix}(\lambda+\Delta^m_r)\\(\lambda+\Delta^m_r)\end{vmatrix}, \tag{C3}$$

where we note that the $U(n)$ GT states on the rhs have maximal lower (μ) patterns and hence represent G-maximal S-dominant symplectic group states, as required.

The matrix elements N^m_r of Eq. (C3) are given, in the notation of Eq. (C1), by

$$N^m_r = N^m_{r,r},$$

which may be evaluated from the $r=l$ case of Eq. (C2) with the substitutions

$$\lambda_{rm}=\mu_{rm}, \quad 1\leqslant r\leqslant m, \quad 1\leqslant m\leqslant h+1.$$

We thereby obtain

$$N^m_r = \left[(-1)^m\frac{\prod_{p=1}^{m+1}(\lambda_{pm+1}-\lambda_{rm}+r-p)\prod_{l=1}^{m-1}(\lambda_{rm}-\lambda_{lm-1}+l-r+1)}{\prod_{\substack{p=1\\ \neq r}}^{m}(\lambda_{rm}-\lambda_{pm}+p-r)(\lambda_{rm}-\lambda_{pm}+p-r+1)}\right]^{1/2},$$

in agreement with Eq. (50). In a similar way we deduce the result

$$a_{2m+1,2m-1}\begin{matrix}(\lambda)\\(\lambda)\end{matrix}\Big) = \sum_{r=1}^{m}\bar{N}^m_r\begin{vmatrix}(\lambda-\Delta^m_r)\\(\lambda-\Delta^m_r)\end{vmatrix}, \tag{C4}$$

with $\bar{N}^m_r$ as in Eq. (52). This demonstrates the matrix element formulas of Eqs. (49)–(52), as required. The matrix element formulas of Eqs. (56) and (57) follow from a similar analysis.

We remark that Eqs. (C3) and (C4) demonstrate that the space of allowed $U(n)$ GT states of the form

$$\begin{vmatrix}(\lambda)\\(\lambda)\end{vmatrix}, \tag{C5}$$

with (λ) S dominant, is invariant under the action of the $U(n)$ generators $a_{2i-1,2j-1}$ $(1\leqslant i,j\leqslant h+1)$. Since the even $U(n)$ generators $a_{2i,2j}$ $(1\leqslant i,j\leqslant h+1)$ vanish on the states (C5), this implies that the above space of states is stable under the action of the $Sp(n)$ generators α^{2i-1}_{2j-1}, i.e., the space of G-maximal S-dominant canonical symplectic group states $|(\lambda)\rangle_+$ is stable under the action of the $Sp(n)$ generators (48), as noted in Sec. VI.

[1] M. Moshinsky, *Group Theory and the Many Body Problem* (Gordon and Breach, New York, 1968); M. Moshinsky, in *Documents on Modern Physics* (Gordon and Breach, New York, 1969); P. Kramer and M. Moshinsky, in *Group Theory and its Applications*, edited by E. M. Loebl (Academic, New York, 1968), Vol. I.

[2] Y. Ohnuki and S. Kamefuchi, *Quantum Field Theory and Parastatistics* (Springer, New York, 1982).

[3] E. M. Loebl, *Group Theory and its Applications* (Academic, New York, 1968), Vols. I and II.

[4] J. Paldus, in *Theoretical Chemistry: Advances and Perspectives*, edited by H. Eyring and D. J. Henderson (Academic, New York, 1972), Vol. 2, p. 131; Phys. Scr. **21**, 295 (1980); Int. J. Quant. Chem. Symp. **9**, 165 (1975).

[5] J. Hinze, Ed., *The Unitary Group*, Lecture Notes in Chemistry (Springer, New York, 1981), Vol. 22.

[6] W. G. Harter and C. W. Patterson, *A Unitary Calculus For Atomic Orbitals*, Lecture Notes in Physics (Springer, New York, 1976), Vol. 49.

[7] B. R. Brooks and H. F. Schaefer III, J. Chem. Phys. **70**, 5092 (1979); B. R. Brooks, W. D. Laidig, P. Saxe, N. C. Handy, and H. F. Schaefer III, Phys. Scr. **21**, 312 (1980); J. Chem. Phys. **72**, 2837, 4652 (1980); P. E. M. Siegbahn, J. Chem. Phys. **72**, 1647 (1980); Int. J. Quant. Chem. **18**, 1229 (1980); H. Lischka, R. Shepard, F. Brown, and I. Shavitt, Int. J. Quant. Chem. Symp. **15**, 91 (1981).

[8] I. M. Gelfand and M. L. Tsetlin, Dokl. Akad. Nauk. SSSR **71**, 825 (1950).

The results of this paper are an announcement only. The proof appears in I. M. Gelfand and M. I. Graev, Izv. Akad. Nauk. SSSR. **29**, 1329 (1965) [see, also, Am. Math. Soc. Transl. **65**, 116 (1967), Ser. 2].
[9]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **4**, 1449 (1963).
[10]L. C. Biedenharn and J. D. Louck, Commun. Math. Phys. **8**, 89 (1968); L. C. Biedenharn, A. Giovannini, and J. D. Louck, J. Math. Phys. **8**, 691 (1967).
[11]G. E. Baird and L. C. Biedenharn, J. Math. Phys. **12**, 1723 (1964); J. Math. Phys. **5**, 1730 (1964); J. Math. Phys. **6**, 1845 (1965).
[12]M. D. Gould, J. Math. Phys. **27**, 1944 (1986).
[13]M. D. Gould, J. Math. Phys. **27**, 1964 (1986).
[14]H. Weyl, *The Theory of Groups and Quantum Mechanics* (Methuen, London, 1931); *The Classical Groups* (Princeton U.P., Princeton, NJ, 1939).
[15]D. P. Zhelobenko, Russ. Math. Surv. **17**, 1 (1962); Transl. Math. Monogr. **40** (1973).
[16]M. D. Gould and E. G. Kalnins, J. Math. Phys. **26**, 1446 (1985).
[17]M. A. Lohe and C. A. Hurst, J. Math. Phys. **12**, 1882 (1971).
[18]A. U. Klymik, J. Phys. A: Math. Gen. **15**, 3009 (1982).
[19]M. K. F. Wong and H. Y. Yeh, J. Math. Phys. **21**, 630 (1980).
[20]J. Michelsson, Rep. Math. Phys. **3**, 193 (1972).
[21]A. M. Bincer, J. Math. Phys. **23**, 347 (1982).
[22]P. Pajas and R. Raczka, J. Math. Phys. **9**, 1188 (1968).
[23]A. M. Bincer, J. Math. Phys. **21**, 671 (1980).
[24]M. Cerkaski, J. Math. Phys. **28**, 989 (1987).
[25]M. Hamermesh, *Group Theory and its Applications* (Addison-Wesley, Reading, MA, 1962).
[26]H. S. Green, J. Math. Phys. **12**, 2106 (1971).
[27]H. S. Green, C. A. Hurst, and Y. Ilamed, J. Math. Phys. **17**, 1376 (1976).
[28]A. Joseph, Am. J. Math. **99**, 1167 (1977).
[29]J. E. Humphreys, *Introduction to Lie Algebras and Representation Theory* (Springer, Berlin, 1972).
[30]M. D. Gould, J. Math. Phys. **22**, 15 (1981).

Relations between spherical functions of compact groups

A. U. Klimyk and N. Ja. Vilenkin
Institute for Theoretical Physics, Kiev-130, Union of Soviet Socialist Republics

(Received 11 November 1988; accepted for publication 21 December 1988)

The associated spherical functions of the homogeneous spaces $SO(n)/SO(n-1)$, $U(n)/U(n-1)$, and $Sp(n)/Sp(n-1)$ are found in different coordinate systems: They are matrix elements of representation operators, right invariant with respect to the subgroups $SO(n-1)$, $U(n-1)$, and $Sp(n-1)$, respectively. The left index of these matrix elements corresponds to the reductions onto the subgroups $SO(p)\times SO(q)$; $U(p)\times U(q)$; and $Sp(p)\times Sp(q)$, $p+q=n$. The relations between these spherical functions are derived. These relations lead to the formulas connecting the Clebsch–Gordan coefficients for the groups $SO(n)$, $U(n)$, and $Sp(n)$.

I. INTRODUCTION

Let G be a compact Lie group and H and K its subgroups such that $G=HBK$, where B is a commutative subgroup. Let T be an irreducible unitary representation of G having class 1 with respect to[1] K and let $\mathbf{e}$ be a normalized vector in the carrier space V of T, which is invariant with respect to the operators $T(k)$, $k\in K$. The space V decomposes into the linear subspaces $V_1,...,V_m$, irreducible for the operators $T(h)$, $h\in H$. Let us choose an orthonormal basis $\{\mathbf{f}_j\}$ in V consisting of bases of the subspaces V_r, $r=1,...,m$. The matrix element $\langle \mathbf{f}_j|T(g)|\mathbf{e}\rangle$ is called an HK-associated spherical function of the representation T. If $\mathbf{f}$ is an orthonormal vector, invariant with respect to the operators $T(h)$, $h\in H$, then the matrix element $\langle \mathbf{f}|T(g)|\mathbf{e}\rangle$ is called an HK-zonal spherical function of T. If $H=K$ we then have the usual associated and zonal spherical functions.

In this paper we find HK-associated and HK-zonal spherical functions for the groups $SO(n)$, $U(n)$, and $Sp(n)$, where H coincides with $SO(p)\times SO(q)$, $U(p)\times U(q)$, $Sp(p)\times Sp(q)$, $p+q=n$ and K coincides with $SO(n-1)$, $U(n-1)$, and $Sp(n-1)$. Moreover, we show that spherical functions of the groups $U(n)$ and $SO(2n)$, as well as spherical functions of the groups $Sp(n)$ and $SO(4n)$, are related by simple formulas.

Our spherical functions are of great significance for applications. These spherical functions constitute the semicanonical bases of representation spaces for the groups $SO(n)$, $U(n)$, and $Sp(n)$ which are bases corresponding to reduction onto the subgroups $SO(p)\times SO(q)$, $U(n)\times U(q)$, $Sp(p)\times Sp(q)$, $p+q=n$, respectively. These bases are useful for solving some physical problems since they realize explicitly decomposition of irreducible representations of G into a sum of representations of H. Besides, HK-spherical functions are related with harmonic analysis on the homogeneous space G/K.

We should like to turn the reader's attention to the connection between spherical functions of the groups $SO(n)$, $U(n)$, and $Sp(n)$. This is a small part of the broad problem: relations between representations of these groups. These relations exist and are of great interest for physics and mathematics. If the irreducible representations of $SO(n)$ with the highest weights $(m,0,...,0)$, $m\geqslant 0$; of $U(n)$ with the highest weights $(m_1,0,...,0,m_2)$, $m_1\geqslant 0\geqslant m_2$; and of $Sp(n)$ with the highest weights $(m_1,m_2,0,...,0)$, $m_1\geqslant m_2\geqslant 0$ are considered, then we can state that in some meaning, representations of $U(n)$ are the Fourier transform of representations of $SO(2n)$ and representations of $Sp(n)$ are the spherical transform of representations of $SO(4n)$.

The connection between spherical functions of the groups $SO(n)$, $U(n)$, and $Sp(n)$ has some consequences. In particular, we show that the connection implies corresponding relations for Clebsch–Gordan coefficients (CGC's) of these groups.

There are relations between representations of compact and noncompact Lie groups with the same complexification.[2–5] In particular, the pairs $SO(p+q)$ and $SO_0(p,q)$, $U(p+q)$ and $U(p,q)$, and $Sp(p+q)$ and $Sp(p,q)$ have the same complex Lie groups. Therefore, it is natural that there are relations between HK-spherical functions of the groups $SO_0(p,q)$, $U(p,q)$, and $Sp(p,q)$: They are shown in Ref. 6.

II. DECOMPOSITIONS OF THE GROUPS SO(n), U(n), AND Sp(n)

Let R and C be the fields of real and complex numbers and let Q be the set of quaternions denoted by the common symbol F. If $q\in Q$, then $q=a+ib+jc+kd$, where a, b, c, and d are real numbers and i,j, and k are imaginary units. We have $|q|^2=a^2+b^2+c^2+d^2$.

The set of points $(w_1,...,w_n)\in F^n$, for which

$$|w_1|^2+\cdots+|w_n|^2=1$$

is the sphere S^F_{n-1} of unit radius. If $F=C$, then $w_j=a_j+ib_j$, where $a_j\in R$, $b_j\in R$, and we have

$$\sum_j |w_j|^2=\sum_j (a_j^2+b_j^2)\,.$$

Therefore, $S^C_{n-1}=S^R_{2n-1}$. In the same way we show that

$$S^Q_{n-1}=S^C_{2n-1}=S^R_{4n-1}\,.$$

The groups $SO(n)$, $U(n)$, and $Sp(n)$ consist of matrices that conserve the form $|w_1|^2+\cdots+|w_n|^2$, where $(w_1,...,w_n)$ belongs to R^n, C^n, and Q^n, respectively. The group $Sp(n)$ consists of quaternion matrices which can be realized by $2n\times 2n$ complex matrices. For this reason, the notation $Sp(2n)$ is often used instead of $Sp(n)$.

If $\mathbf{w}_0=(0,...,0,1)$, then the stationary subgroup for $\mathbf{w}_0$ is $SO(n-1)$, $U(n-1)$, or $Sp(n-1)$. Therefore,

 0022-2488/89/061219-07$02.50

$$S^R_{n-1}=\frac{\mathrm{SO}(n)}{\mathrm{SO}(n-1)},\quad S^C_{n-1}=\frac{\mathrm{U}(n)}{\mathrm{U}(n-1)},\quad S^Q_{n-1}=\frac{\mathrm{Sp}(n)}{\mathrm{Sp}(n-1)}. \tag{1}$$

A real rotation by the angle θ in the (x_j, x_k) plane is denoted by $g_{jk}(\theta)$. For brevity, the rotation $g_{k,k+1}(\theta)$ will be denoted by $g_k(\theta)$ and the diagonal matrix $\mathrm{diag}(1,...,1,q,1,...,1)$, $|q|=1$ of $\mathrm{U}(n)$ or $\mathrm{Sp}(n)$ will be denoted by $d_k(q)$, where k means that q is on k th entry.

Let us introduce the spherical coordinates $\theta_1,...,\theta_{n-1}$ on the sphere S^R_{n-1}. The point $\mathbf{x}=(x_1,...,x_n)\in S^R_{n-1}$ can be represented as

$$\begin{aligned} x_1&=\sin\theta_{n-1}\cdots\sin\theta_2\sin\theta_1,\\ x_2&=\sin\theta_{n-1}\cdots\sin\theta_2\cos\theta_1,\\ &\vdots\\ x_n&=\cos\theta_{n-1}. \end{aligned} \tag{2}$$

With these spherical coordinates point $\mathbf{x}$ is obtained from point $\mathbf{x}_0=(0,...,0,1)$ as

$$\mathbf{x}=g_1(\theta_1)\cdots g_{n-1}(\theta_{n-1})\mathbf{x}_0. \tag{3}$$

The matrices $g_{n-1}(\theta_{n-1})$ form the one-parameter subgroup, which will be denoted by A. Since $g_1(\theta_1)\cdots g_{n-2}(\theta_{n-2})\in\mathrm{SO}(n-1)$, we have $S^R_{n-1}=\mathrm{SO}(n-1)A\mathbf{x}_0$: It follows from this and (1) that

$$\mathrm{SO}(n)=\mathrm{SO}(n-1)A\,\mathrm{SO}(n-1). \tag{4}$$

The point $\mathbf{z}=(z_1,...,z_n)\in S^C_{n-1}$ is represented as

$$\begin{aligned} z_1&=(\exp i\phi_1)\sin\theta_{n-1}\cdots\sin\theta_2\sin\theta_1,\\ z_2&=(\exp i\phi_2)\sin\theta_{n-1}\cdots\sin\theta_2\cos\theta_1,\\ &\vdots\\ z_n&=(\exp i\phi_n)\cos\theta_{n-1}. \end{aligned} \tag{5}$$

It is obtained from point $\mathbf{z}_0=(0,...,0,1)$ as

$$\mathbf{z}=d_1(\phi_1)d_2(\phi_2)g_1(\theta_1)\cdots d_n(\phi_n)g_{n-1}(\theta_{n-1})\mathbf{z}_0.$$

The elements $d_n(\phi_n)$ form the one-parameter subgroup, which will be denoted by D. It is clear that $D\sim\mathrm{U}(1)$. We have

$$S^C_{n-1}=\mathrm{U}(n-1)DA\mathbf{z}_0.$$

Therefore,

$$\mathrm{U}(n)=\mathrm{U}(n-1)DA\,\mathrm{U}(n-1). \tag{6}$$

The point $\mathbf{q}=(q_1,...,q_n)\in S^Q_{n-1}$ is represented as

$$\begin{aligned} q_1&=u_1\sin\theta_{n-1}\cdots\sin\theta_2\sin\theta_1,\\ q_2&=u_2\sin\theta_{n-1}\cdots\sin\theta_2\cos\theta_1,\\ &\vdots\\ q_n&=u_n\cos\theta_{n-1}, \end{aligned} \tag{7}$$

where $u_j\in Q$, $|u_j|=1$. The set of quaternions q, such that $|q|=1$, forms the group $\mathrm{Sp}(1)$. We have

$$S^Q_{n-1}=\mathrm{Sp}(n-1)\mathrm{Sp}(1)A\mathbf{q}_0,$$

where $\mathbf{q}_0=(0,...,0,1)\in S^Q_{n-1}$. Therefore,

$$\mathrm{Sp}(n)=\mathrm{Sp}(n-1)\mathrm{Sp}(1)A\,\mathrm{Sp}(n-1). \tag{8}$$

Let us introduce the polyspherical coordinates

$$\begin{aligned} x_1&=\sin\theta\cos\alpha_{p-1},\\ &\vdots\\ x_p&=\sin\theta\sin\alpha_{p-1}\cdots\sin\alpha_2\sin\alpha_1,\\ x_{p+1}&=\cos\theta\sin\beta_{q-1}\cdots\sin\beta_2\sin\beta_1,\\ &\vdots\\ x_n&=\cos\theta\cos\beta_{q-1} \end{aligned} \tag{9}$$

on the sphere S^R_{n-1}, where $q=n-p$. With these coordinates the point x is obtained from $\mathbf{x}_0=(0,...,0,1)$ as

$$\mathbf{x}=kk'g_{1n}(\theta)\mathbf{x}_0,\quad k\in\mathrm{SO}(p),\quad k'\in\mathrm{SO}(q).$$

The matrices $g_{1n}(\theta)$ form the one-parameter subgroup, which is denoted by B. We have

$$S^R_{n-1}=\mathrm{SO}(p)\mathrm{SO}(q)B\mathbf{x}_0.$$

Therefore,

$$\mathrm{SO}(n)=[\mathrm{SO}(p)\times\mathrm{SO}(q)]B\,\mathrm{SO}(n-1),\quad n=p+q. \tag{10}$$

In order to obtain polyspherical coordinates on S^C_{n-1} we have to multiply the lhs and rhs of Eqs. (9) by $\exp i\phi_j$, $j=1,...,n$, correspondingly. We obtain that $S^C_{n-1}=[\mathrm{U}(p)\times\mathrm{U}(q)]B\mathbf{z}_0$, where $\mathbf{z}_0=(0,...,0,1)$. Therefore,

$$\mathrm{U}(n)=[\mathrm{U}(p)\times\mathrm{U}(q)]B\,\mathrm{U}(n-1). \tag{11}$$

In the same way we show that

$$\mathrm{Sp}(n)=[\mathrm{Sp}(p)\times\mathrm{Sp}(q)]B\,\mathrm{Sp}(n-1). \tag{12}$$

In fact, the decompositions (4), (6), and (8) are special cases of the decompositions (10)–(12) with $p=n-1$.

III. INVARIANT MEASURE AND LAPLACE OPERATOR

The groups $\mathrm{SO}(n)$, $\mathrm{U}(n)$, and $\mathrm{Sp}(n)$ will be denoted by the symbol G; the subgroups $\mathrm{SO}(p)\times\mathrm{SO}(q)$, $\mathrm{U}(p)\times\mathrm{U}(q)$, and $\mathrm{Sp}(p)\times\mathrm{Sp}(q)$ will be denoted by H; and the subgroups $\mathrm{SO}(n-1)$, $\mathrm{U}(n-1)$, and $\mathrm{Sp}(n-1)$ will be denoted by K. Then the decompositions (10)–(12) can be written as $G=HBK$. The spheres S_{n-1} in real, complex, and quaternion spaces are the quotient spaces G/K. The invariant measure on group G can be represented as a product of the invariant measure on $S_{n-1}=G/K$ by the invariant measure on K. The invariant measure on S_{n-1}, expressed in spherical or polyspherical coordinates, is well known.[1] If elements $g\in G$ are represented as $g=hg_{1n}(\theta)k$, $h\in H$, $k\in K$, then the invariant integral on G can be written in the form[7]

$$\int_G f(g)dg=c\int_H\int_0^{\pi/2}\int_K f(hg_{1n}(\theta)k)D_{pq}(\theta)dh\,d\theta\,dk, \tag{13}$$

where dg, dh, and dk are the normalized invariant measures on G, H, and K, respectively, and

$$D_{pq}(\theta)=\sin^{rp-1}\theta\cos^{rq-1}\theta,$$

where $r=1$ for $G=\mathrm{SO}(n)$, $r=2$ for $G=\mathrm{U}(n)$, and $r=4$ for $G=\mathrm{Sp}(n)$. The constant c in (13) is

$$c=2\Gamma(rp/2)\Gamma(rq/2)/\Gamma(r(p+q)/2). \tag{14}$$

Formula (13) has to be changed for $G=\mathrm{SO}(n)$ and $p=n-1$. The explicit expression for the invariant integral in this case is shown in Ref. 1.

The Laplace operator Δ_n on S^R_{n-1}, expressed in polyspherical coordinates, is well known[1]: It can be written as

$$\Delta_n = D^{-1}_{pq}(\theta)\frac{d}{d\theta}D_{pq}(\theta)\frac{d}{d\theta} + \sin^{-2}\theta\Delta_p + \cos^{-2}\theta\Delta_q , \tag{15}$$

where Δ_p and Δ_q are the Laplace operators on S^R_{p-1} and S^R_{q-1}, respectively.

We can consider S^C_{n-1} and S^Q_{n-1} as the spheres S^R_{2n-1} and S^R_{4n-1}, respectively. Therefore, the Laplace operators on S^C_{n-1} and S^Q_{n-1} are obtained from (15) by replacing Δ_p and Δ_q by Δ_{rp} and Δ_{rq}, correspondingly.

IV. *HK*-SPHERICAL FUNCTIONS OF THE GROUP *G*

Let T be an irreducible unitary representation of G which has class 1 with respect to K and H. If $|O_K\rangle$ is a normalized vector, invariant with respect to K, and $|O_H\rangle$ is a normalized vector, invariant with respect to H, then

$$\langle O_H|T(g)|O_K\rangle \equiv \varphi^{HK}_T(g)$$

is the HK-spherical function of T. It is clear that

$$\varphi^{HK}_T(hbk) = \varphi^{HK}_T(b), \quad h\in H, \quad b\in B, \quad k\in K .$$

The usual spherical function φ_T (that is, the spherical function φ^{KK}_T) has the property $\varphi_T(e) = 1$, where e is the unit element of G. For φ^{HK}_T we have $\varphi^{HK}_T(e) = \langle O_H|O_K\rangle$. The relation

$$\overline{\varphi^{HK}_T(g)} = \varphi^{HK}_S(g) = \varphi^{KH}_T(g^{-1}) = \langle O_K|T(g^{-1})|O_H\rangle$$

holds, where S is the representation contragradient to T. We also have

$$|\varphi^{HK}_T(g)| \leqslant 1 ,$$

$$\int_H \varphi^{HK}_T(gh)dh = \langle O_H|O_K\rangle\varphi^{HH}_T(g) ,$$

$$\int_K \varphi^{HH}_T(gk)dk = \langle O_H|O_K\rangle^{-1}\varphi^{HK}_T(g) .$$

Let $L^2(B)$ be the Hilbert space of functions f on the interval $(0, \pi/2)$, with the scalar product

$$(f_1, f_2) = c\int_0^{\pi/2} f_1(\theta)f_2(\theta)D_{pq}(\theta)d\theta .$$

Then for $f\in L^2(B)$ we have the expansion

$$f(\theta) = \sum_T a_T\varphi^{HK}_T(g_{1n}(\theta)) , \tag{16}$$

where the sum is taken over all nonequivalent irreducible representations of G having class 1 with respect to K and H and

$$a_T = (\dim T)c\int_0^{\pi/2} f(\theta)\,\overline{\varphi^{HK}_T(g_{1n}(\theta))}D_{pq}(\theta)d\theta . \tag{17}$$

The Plancherel formula

$$\|f\|^2 = \sum_T (\dim T)^{-1}|a_T|^2 \tag{18}$$

holds. Formulas (16)–(18) are easily derived from the Peter–Weyl theorem.

Let T now have class 1 only with respect to the subgroup K. The basis elements f_j of Sec. I will be denoted by $|M,\alpha\rangle$, where M are the highest weights of the irreducible representations D_M of H and α labels basis elements of carrier spaces of these representations of H. The matrix element

$$t^{T,HK}_{M\alpha,O}(g) \equiv t^T_{M\alpha,O}(g) = \langle M,\alpha|T(g)|O_K\rangle \tag{19}$$

is an HK-associated spherical function of the representation T. It is clear that $t^T_{M\alpha,O}(hbk) = t^T_{M\alpha,O}(hb)$. We have

$$t^{T,HK}_{M\alpha,O}(hb) = D^M_{\alpha O}(h)t^{T,HK}_{MO}(b) , \tag{20}$$

where $t^{T,HK}_{MO}(b)$ does not depend on α and $D^M_{\alpha O}(h)$ is a usual associated spherical function of the representation D_M.

We have

$$(t^{T,HH}_{M\alpha,O} * \varphi^{HK}_S)(g) \equiv \int_G t^{T,HH}_{M\alpha,O}(g_1)\varphi^{HK}_S(g_1^{-1}g)dg_1 = (\dim T)^{-1}t^{T,HK}_{M\alpha,O}(g)\delta_{TS} .$$

Moreover, the convolution with φ^{HK}_T maps $L^2(G/H)$ into $L^2(G/K)$. For this reason, φ^{HK}_T is called an intertwining function[8,9] on G. The relation

$$\int_K t^{T,HK}_{M\alpha,O}(gh)dh = \langle O_H|O_K\rangle t^{T,HH}_{M\alpha,O}(g)$$

holds.

With the help of the Peter–Weyl theorem it is easy to prove that

$$\int_G |t^{T,HK}_{M\alpha,O}(g)|^2 dg = (\dim T)^{-1} .$$

This formula and Eq. (13) imply that

$$c\int_0^{\pi/2} |t^{T,HK}_{MO}(g_{1n}(\theta))|^2 D_{pq}(\theta)d\theta = \frac{\dim D_M}{\dim T} , \tag{21}$$

where c is given by Eq. (14).

V. *HK*-SPHERICAL FUNCTIONS OF SO(*n*)

Only the irreducible representations $T_m \equiv T^{SO(n)}_m$ of $SO(n)$ with the highest weights $(m,0,\ldots,0)$ have class 1 with respect to $K = SO(n-1)$. If m is even, then T_m has class 1 with respect to $H = SO(p)\times SO(q)$. We have the decompositions[1]

$$T_{m\downarrow SO(n-1)} = \sum_{l=0}^{m} T^{SO(n-1)}_l ,$$

$$T_{m\downarrow SO(p)\times SO(q)} = \sum_{r,s} (T^{SO(p)}_r \otimes T^{SO(q)}_s) ,$$

where the sums are taken over non-negative integral values of r and s such that

$$r + s \leqslant m, \quad (-1)^{r+s-m} = 1 .$$

For our case formula (20) can be written in the form

$$t^{m,HK}_{(rs,\alpha\beta)O}(hh'b) = D^r_{\alpha O}(h)D^s_{\beta O}(h')t^{m,HK}_{(rs)O}(b) , \quad h\in SO(p), \quad h'\in SO(q) , \tag{22}$$

where $D^r_{\alpha O}(h)$ is a usual associated spherical function.[1]

Function (22) is an eigenfunction of the Laplace operator (15) with the eigenvalue $-m(m+n-2)$. Since

$$\Delta_p D^r_{\alpha O}(h) = -r(r+p-2)D^r_{\alpha O}(h) ,$$

$$\Delta_q D^s_{\beta O}(h') = -s(s+q-2)D^s_{\beta O}(h') ,$$

operator (15), when acting on function (22), then leads to the differential equation

$$\left[\sin^{1-p}\theta\cos^{1-q}\theta\frac{d}{d\theta}\sin^{p-1}\theta\cos^{q-1}\theta\frac{d}{d\theta}\right.$$
$$\left.-\frac{r(r+p-2)}{\sin^2\theta}-\frac{s(s+q-2)}{\cos^2\theta}\right]u(\theta)$$
$$=-m(m+n-2)u(\theta),\quad n=p+q \tag{23}$$

for $t^{m,HK}_{(rs)O}(g_{1n}(\theta))$. This function is continuous at the point $\theta=0$. The solution of Eq. (23), continuous at this point, has the form

$$u(\theta)=\sin^r\theta\cos^s\theta P^{(r-1+p/2,s-1+q/2)}_{(m-r-s)/2}(\cos 2\theta)\,. \tag{24}$$

Therefore,

$$t^{m,HK}_{(rs)O}(g_{1n}(\theta))=c^m_{rs}u(\theta)\,.$$

In order to find the constant c^m_{rs} we have to take Eq. (21) into account, which in our case can be written as

$$c\int_0^{\pi/2}|c^m_{rs}u(\theta)|^2\sin^{p-1}\theta\cos^{q-1}\theta\,d\theta$$
$$=(\dim T^{SO(p)}_r)(\dim T^{SO(q)}_s)(\dim T^{SO(n)}_m)^{-1}\,.$$

Substituting expression (24) for $u(\theta)$ and the expression for dimensions of the representations, we obtain $|c^m_{rs}|$. The basis elements $|M,\alpha\rangle$ in (19) are defined uniquely up to constants a_M, $|a_M|=1$. Therefore, we can assume that $c^m_{rs}\geqslant 0$. As a result, we obtain

$$t^{m,HK}_{(rs)O}(g_{1n}(\theta))$$
$$=N\sin^r\theta\cos^s\theta P^{(r-1+p/2,s-1+q/2)}_{(m-r-s)/2}(\cos 2\theta)\,, \tag{25}$$

where

$$N=\left[\frac{\Gamma((m+r+s+n-2)/2)\Gamma((m-r-s+2)/2)\beta(p-1,q-1)(2r+p-2)}{2\Gamma((m+r-s+p)/2)\Gamma((m-r+s+q)/2)\beta(p/2,q/2)r!s!(m+n-3)!}\right.$$
$$\left.\times(2s+q-2)(r+p-3)!(s+q-3)!\right]^{1/2}. \tag{26}$$

For HK-zonal spherical functions we have the expression

$$\varphi^{HK}_m(\theta)=\frac{1}{2}\left[\frac{\Gamma((m+p-2)/2)\Gamma((m+2)/2)(n-3)!m!}{\Gamma((m+p)/2)\Gamma((m+q)/2)\beta(p/2,q/2)(m+n-3)!}\right]^{1/2}$$
$$\times P^{(-1+p/2,-1+q/2)}_{m/2}(\cos 2\theta)\,. \tag{27}$$

VI. *HK*-SPHERICAL FUNCTIONS OF U(*n*)

Only the irreducible representations of $U(n)$ with the highest weights $(m,0,\ldots,0,m')$, $m\geqslant 0\geqslant m'$ have class 1 with respect to $K=U(n-1)$. We denote these representations by $T_{mm'}\equiv T^{U(n)}_{mm'}$. We have the decompositions

$$T_{mm'\downarrow U(n-1)}=\sum_{m\geqslant l\geqslant 0\geqslant l'\geqslant m'}T^{U(n-1)}_{ll'} \tag{28}$$

and

$$T_{mm'\downarrow U(p)\times U(q)}=\sum(T^{U(p)}_{rr'}\otimes T^{U(q)}_{ss'}),$$
$$p+q=n,\quad p\neq 1,\quad q\neq 1\,, \tag{29}$$

where the summation is carried out over the highest weights $(r,0,\ldots,0,r')(s,0,\ldots,0,s')$ of irreducible representations of $U(p)\times U(q)$, for which

$$r+r'+s+s'=m+m',\quad r-r'+s-s'\leqslant m-m'\,. \tag{30}$$

In particular, $T^{U(n)}_{mm'}$ has class 1 with respect to $H=U(p)\times U(q)$ if and only if $m=m'$.

If $q=1$, then instead of (29) we have

$$T_{mm'\downarrow U(n-1)\times U(1)}=\sum_{m\geqslant s\geqslant 0\geqslant s'\geqslant m'}(T^{U(n-1)}_{ss'}$$
$$\otimes T^{U(1)}_{m+m'-s-s'})\,.$$

Formula (20) for $U(n)$ has the form

$$t^{mm',HK}_{(rr',ss')O}(hh'b)=D^{rr'}_{\alpha O}(h)D^{ss'}_{\beta O}(h')t^{mm',HK}_{(rr',ss')O}(b)\,,$$
$$h\in U(p),\quad h'\in U(q) \tag{31}$$

if $p\neq 1$, $q\neq 1$ and the form

$$t^{mm',HK}_{(ll',\alpha)O}(hh'b)=D^{ll'}_{\alpha O}(h)e^{-i(l+l'-m-m')\phi}t^{mm',HK}_{(ll')O}(b)\,,$$
$$h\in U(n-1),\quad h'=e^{i\phi} \tag{32}$$

if $p=n-1$, $q=1$.

The function $t^{mm',HK}_{(ll')O}(b)$ from (32) does not depend on the subgroup $U(1)$ of the group $H=U(n-1)\times U(1)$; therefore, it is really the function $t^{mm',KK}_{(ll')O}(b)$.

Using the results of Ref. 10, we prove that function (31) is the restriction onto $S^C_{n-1}=S^R_{2n-1}$ of a homogeneous harmonic polynomial of degree $m-m'$. Therefore, (31) is an eigenfunction of the operator Δ_{2n} with the eigenvalue $-(m-m')(m-m'+2n-2)$. Since $D^{rr'}_{\alpha O}(h)$ and $D^{ss'}_{\beta O}(h')$ are eigenfunctions of the operators Δ_{2p} and Δ_{2q}, respectively. we then obtain the differential equation

$$\left[\sin^{1-2p}\theta\cos^{1-2q}\theta\frac{d}{d\theta}\sin^{2p-1}\theta\cos^{2q-1}\theta\frac{d}{d\theta}\right.$$
$$\left.-\frac{J(J+2p-2)}{\sin^2\theta}-\frac{J'(J'+2q-2)}{\cos^2\theta}\right]u(\theta)$$
$$=-(m-m')(m-m'+2n-2)u(\theta) \tag{33}$$

for $t^{mm',HK}_{(rr',ss')O}(g_{1n}(\theta))$, where $J=r-r'$, $J'=s-s'$. In order to obtain the differential equation for $t^{mm',KK}_{(ll')O}(g_{1n}(\theta))$ we have to replace $J'(J'+2q-2)$ by s^2; p by $n-1$; and q by 1, $s=l+l'-m-m'$.

Comparing the differential equation for the function $t^{mm',KK}_{(ll')O}(g_{1n}(\theta))$, $K=U(n-1)$ with Eq. (23), we find that this function satisfies the differential equation for the function $t^{m-m',HK}_{(l-l',s)O}(g_{1,2n}(\theta))$, $H=SO(2n-2)\times SO(2)$,

$K = \mathrm{SO}(2n-1)$, corresponding to the group $\mathrm{SO}(2n)$. Repeating the reasonings of Sec. V for the group $\mathrm{U}(n)$, we obtain that

$$t^{\mathrm{U}(n),mm',KK}_{(ll')O}(g_{1n}(\theta)) = [(\dim T^{\mathrm{SO}(2n)}_{m-m'})(\dim T^{\mathrm{U}(n-1)}_{ll'})]^{1/2} \times [(\dim T^{\mathrm{U}(n)}_{mm'})(\dim T^{\mathrm{SO}(2n-2)}_{l-l'})]^{-1/2} \times t^{\mathrm{SO}(2n),m-m',HK}_{(l-l',s)O}(g_{1,2n}(\theta)), \quad (34)$$

where $K = \mathrm{U}(n-1)$ for the group $\mathrm{U}(n)$ and $K = \mathrm{SO}(2n-1)$, $H = \mathrm{SO}(2n-2)\times\mathrm{SO}(2)$ for the group $\mathrm{SO}(2n)$.

Using the formulas for dimensions of representations in (34), we obtain[11,12] that

$$t^{\mathrm{U}(n),mm',KK}_{(ll')O}(g_{1n}(\theta)) = N \sin^{l-l'}\theta \cos^{l+l'-m-m'}\theta \times P^{(l-l'+n-2,\,l+l'-m-m')}_{m-l}(\cos 2\theta), \quad (35)$$

where

$$N = \left[(n-2)(l-l'+n-2)\frac{m!(-m')!(m-l)!(l-m'+n-2)!(l+n-3)!(-l'+n-3)!}{(l'-m')!l!(-l')!(m-l'+n-2)!(m+n-2)!(-m'+n-2)!}\right]^{1/2}.$$

In particular, for the zonal spherical function of the representation $T_{mm'}$ of $\mathrm{U}(n)$ with respect to the subgroup $K = \mathrm{U}(n-1)$ we have

$$\varphi^{KK}_{mm'}(g_{1n}(\theta)) = [(n-2)!m!/(m+n-2)!]\cos^{-m-m'}\theta P^{(n-2,\,-m-m')}_{m}(\cos 2\theta). \quad (36)$$

Comparing the differential equations (23) and (33) we find that

$$t^{\mathrm{U}(n),mm',HK}_{(rr',ss')O}(g_{1n}(\theta)) = \left[\frac{(\dim T^{\mathrm{SO}(2n)}_{m-m'})(\dim T^{\mathrm{U}(p)}_{rr'})(\dim T^{\mathrm{U}(q)}_{ss'})}{(\dim T^{\mathrm{U}(n)}_{mm'})(\dim T^{\mathrm{SO}(2p)}_{r-r'})(\dim T^{\mathrm{SO}(2q)}_{s-s'})}\right]^{1/2} t^{\mathrm{SO}(2n),m-m',HK}_{(r-r',s-s')O}(g_{1,2n}(\theta)), \quad (37)$$

where $H = \mathrm{U}(p)\times\mathrm{U}(q)$, $K = \mathrm{U}(n-1)$ for the group $\mathrm{U}(n)$ and $H = \mathrm{SO}(2p)\times\mathrm{SO}(2q)$, $K = \mathrm{SO}(2n-1)$ for the group $\mathrm{SO}(2n)$. In particular, the HK-zonal spherical functions of the groups $\mathrm{U}(n)$ and $\mathrm{SO}(2n)$ are related by the formula

$$\varphi^{\mathrm{U}(n),HK}_{m,-m}(g_{1n}(\theta)) = (\dim T^{\mathrm{SO}(2n)}_{2m})^{1/2}(\dim T^{\mathrm{U}(n)}_{m,-m})^{-1/2}\varphi^{\mathrm{SO}(2n),HK}_{2m}(g_{1,2n}(\theta)).$$

Let us note that the multiplier in Eq. (37) is equal to

$$N = \left[\frac{\beta(m+n-1,-m'+n-1)\beta(p,p-1)\beta(q,q-1)m!(-m')!(s-s')!(r-r')!}{2\beta(r+p-1,-r'+p-1)\beta(s+q-1,-s'+q-1)\beta(n,n-1)(m-m')!s!r!(-s')!(-r')!}\right]^{1/2}. \quad (38)$$

VII. *HK*-SPHERICAL FUNCTIONS OF Sp(*n*)

Only the irreducible representations of $\mathrm{Sp}(n)$ with the highest weights $(m,m',0,\ldots,0)$, $m\geqslant m'\geqslant 0$ have class 1 with respect to $K = \mathrm{Sp}(n-1)$. We denote these representations by $T_{mm'} \equiv T^{\mathrm{Sp}(n)}_{mm'}$.

If $p>1$, $q>1$, we then have the decomposition

$$T_{mm'\downarrow \mathrm{Sp}(p)\times\mathrm{Sp}(q)} = \sum (T^{\mathrm{Sp}(p)}_{rr'} \otimes T^{\mathrm{Sp}(q)}_{ss'}),$$

where the summation is carried out over the highest weights $(r,r',0,\ldots,0)$ $(s,s',0,\ldots,0)$ of irreducible representations of $\mathrm{Sp}(p)\times\mathrm{Sp}(q)$, for which

$$m-m' \leqslant r+r'+s+s' \leqslant m+m',$$
$$|r-r'-s+s'| \leqslant m-m' \leqslant r-r'+s-s',$$
$$(-1)^{m+m'} = (-1)^{r+r'+s+s'}.$$

If $q=1$, then instead of the highest weights $(r,r',0,\ldots,0)$ $(s,s',0,\ldots,0)$ we have the highest weights $(r,r',0,\ldots,0)$ (s).

Formula (20) for $\mathrm{Sp}(n)$ has the form

$$t^{mm',HK}_{(rr',ss',\alpha\beta)O}(hh'b) = D^{rr'}_{\alpha O}(h)D^{ss'}_{\beta O}(h')t^{mm',HK}_{(rr',ss')O}(b), \quad h\in\mathrm{Sp}(p),\quad h'\in\mathrm{Sp}(q). \quad (39)$$

The function (39) is the restriction on $S^Q_{n-1} = S^R_{4n-1}$ of a harmonic polynomial of degree $m+m'$ in real parts of quaternion variables; therefore, it is an eigenfunction of the Laplace operator Δ_{4n} belonging to the eigenvalue $-(m+m')(m+m'+4n-2)$. Using formula (15) for Δ_{4n}, we obtain the differential equation

$$\left[\sin^{1-4p}\theta\cos^{1-4q}\theta\frac{d}{d\theta}\sin^{4p-1}\theta\cos^{4q-1}\theta\frac{d}{d\theta} - \frac{J(J+4p-2)}{\sin^2\theta} - \frac{J'(J'+4q-2)}{\cos^2\theta}\right]u(\theta) = -(m+m')(m+m'+4n-2)u(\theta) \quad (40)$$

for $t^{mm',HK}_{(rr',ss')O}(g_{1n}(\theta))$, where $J = r+r'$, $J' = s+s'$.

Comparing the differential equations (23) and (40) we find that

$$t^{\mathrm{Sp}(n),mm',HK}_{(rr',ss')O}(g_{1n}(\theta)) = \left[\frac{(\dim T^{\mathrm{SO}(4n)}_{m+m'})(\dim T^{\mathrm{Sp}(p)}_{rr'})(\dim T^{\mathrm{Sp}(q)}_{ss'})}{(\dim T^{\mathrm{Sp}(n)}_{mm'})(\dim T^{\mathrm{SO}(4p)}_{r+r'})(\dim T^{\mathrm{SO}(4q)}_{s+s'})}\right]^{1/2} \times t^{\mathrm{SO}(4n),m+m',HK}_{(r+r',s+s')O}(g_{1,4n}(\theta)) \quad (41)$$

for $q>1$, where $H = \mathrm{Sp}(p)\times\mathrm{Sp}(q)$, $K = \mathrm{Sp}(n-1)$ for $\mathrm{Sp}(n)$ and $H = \mathrm{SO}(4p)\times\mathrm{SO}(4q)$, $K = \mathrm{SO}(4n-1)$ for $\mathrm{SO}(4n)$. In particular,

$$\varphi^{\mathrm{Sp}(n),HK}_{mm'}(g_{1n}(\theta)) = (\dim T^{\mathrm{SO}(4n)}_{m+m'})^{1/2}(\dim T^{\mathrm{Sp}(n)}_{mm'})^{-1/2} \times \varphi^{\mathrm{SO}(4n),HK}_{m+m'}(g_{1,4n}(\theta)). \quad (42)$$

The multiplier in Eq. (41) is given by the formula

$$N=\left[\frac{\beta(m+2n-1,m'+2n-1)B(2p-1,2p)\beta(2q-1,2q)(m'+2n-2)(2p-2)(2q-2)}{\beta(r+2p-1,r'+2p-1)\beta(s+2q-1,s'+2q-1)\beta(2n-1,2n)(r'+2p-2)(2n-2)}\right.$$
$$\left.\times\frac{(r+r')!(s+s')!(r+r'+1)(s-s'+1)(m+1)!m'!}{(s'+2q-2)(m+m')!(m-m'+1)(r+1)!r'!(s+1)!s'!}\right]^{1/2}. \tag{43}$$

For $q=1$ we have

$$t^{\mathrm{Sp}(n),mm',HK}_{(rr',s)O}(g_{1n}(\theta))=Nt^{\mathrm{SO}(4n),m+m',HK}_{(r+r',s)O}(g_{1,4n}(\theta)), \tag{44}$$

where

$$N=\left[\frac{\beta(m+2n-1,m'+2n-1)(m'+2n-2)(r+r')!(r-r'+1)(m+1)m'!(4n-6)!}{\beta(r+2n-3,r'+2n-3)(r'+2n-4)(m+m')!(m-m'+1)(s+1)(r+1)r'!}\right.$$
$$\left.\times\frac{(2n-1)!}{(4n-2)!(2n-5)!}\right]^{1/2}. \tag{45}$$

Comparing the differential equations (33) and (40) we find that

$$t^{\mathrm{Sp}(n),mm',HK}_{(rr',ss')O}(g_{1n}(\theta))$$
$$=\left[\frac{(\dim T^{\mathrm{U}(2n)}_{m,-m'})(\dim T^{\mathrm{Sp}(p)}_{rr'})}{(\dim T^{\mathrm{Sp}(n)}_{mm'})(\dim T^{\mathrm{U}(2p)}_{(r+k,-r'+k)})}\right.$$
$$\left.\times\frac{(\dim T^{\mathrm{Sp}(q)}_{ss'})}{(\dim T^{\mathrm{U}(2q)}_{(s+l,-s'+l)})}\right]^{1/2}$$
$$\times t^{\mathrm{U}(2n),m,-m',HK}_{(r+k,-r'+k;s+l,-s'+l)O}(g_{1,2n}(\theta)), \tag{46}$$

where k and l are integers such that

$$r-r'+s-s'+2k+2l=m-m';$$

$H=\mathrm{Sp}(p)\times\mathrm{Sp}(q)$, $K=\mathrm{Sp}(n-1)$ for $\mathrm{Sp}(n)$; and $H=\mathrm{U}(2p)\times\mathrm{U}(2q)$, $K=\mathrm{U}(2n-1)$ for $\mathrm{U}(2n)$.

VIII. RELATIONS BETWEEN SOME CGC's OF THE GROUPS SO(n), U(n), AND Sp(n)

Relations between HK-associated spherical functions for the groups $\mathrm{SO}(n)$, $\mathrm{U}(n)$, and $\mathrm{Sp}(n)$ imply relations between CGC's for these groups. We express CGC's in terms of scalar factors with respect to subgroups. For this reason, below we consider scalar factors of the group G with respect to the subgroups H and K.

We can factorize CGC's of the group G into the product of scalar factors of G with respect to H and K by the CGC's of H and K. This factorization and Eq. (13) lead to the formula

$$c\frac{\dim T''}{\dim D_{M'}}\int_0^{\pi/2}t_{MO}(b)t'_{M'O}(b)\,\overline{t''_{M'O}(b)}D_{pq}(\theta)d\theta$$
$$=\left(\begin{matrix}T & T' & T'' & G\\ M & M' & M'' & H\end{matrix}\middle|\begin{matrix}T & T' & T'' & G\\ O_K & O_K & O_K & K\end{matrix}\right)^*, \tag{47}$$

where $b=g_{1n}(\theta)$; M, M', and M'' are the highest weights of irreducible representations of H; and $*$ denotes a complex conjugation.

Writing relation (47) for the associated spherical functions from Eq. (37), we obtain the relation

$$\left(\begin{matrix}m_1m_1' & m_2m_2' & mm' & \mathrm{U}(n)\\ (r_1r_1',s_1s_1') & (r_2r_2',s_2s_2') & (rr',ss') & \mathrm{U}(p)\times\mathrm{U}(q)\end{matrix}\right)$$
$$\times\left(\begin{matrix}m_1m_1' & m_2m_2' & mm' & \mathrm{U}(n)\\ O & O & O & \mathrm{U}(n-1)\end{matrix}\right)^*$$
$$=N'N''N^{-1}\left(\begin{matrix}m_1-m_1' & m_2-m_2' & m-m' & \mathrm{SO}(2n)\\ O & O & O & \mathrm{SO}(2n-1)\end{matrix}\right)^*$$
$$\times\left(\begin{matrix}m_1-m_1' & m_2-m_2' & m-m' & \mathrm{SO}(2n)\\ (r_1-r_1',s_1-s_1') & (r_2-r_2',s_2-s_2') & (r-r',s-s') & \mathrm{SO}(2p)\times\mathrm{SO}(2q)\end{matrix}\right), \tag{48}$$

where N, N', and N'' are expressions (38) taken for the representations $T^{\mathrm{U}(n)}_{mm'}$, $T^{\mathrm{U}(n)}_{m_1m_1'}$, and $T^{\mathrm{U}(n)}_{m_2m_2'}$, respectively.

In the case of the groups $\mathrm{Sp}(n)$ and $\mathrm{SO}(4n)$ we obtain, from the spherical functions (42), that

$$\left(\begin{matrix}m_1m_1' & m_2m_2' & mm' & \mathrm{Sp}(n)\\ (r_1r_1',s_1s_1') & (r_2r_2',s_2s_2') & (rr',ss') & \mathrm{Sp}(p)\times\mathrm{Sp}(q)\end{matrix}\right)$$
$$\times\left(\begin{matrix}m_1m_1' & m_2m_2' & mm' & \mathrm{Sp}(n)\\ O & O & O & \mathrm{Sp}(n-1)\end{matrix}\right)^*$$
$$=N'N''N^{-1}\left(\begin{matrix}m_1+m_1' & m_2+m_2' & m+m' & \mathrm{SO}(4n)\\ O & O & O & \mathrm{SO}(4n-1)\end{matrix}\right)^*$$
$$\times\left(\begin{matrix}m_1+m_1' & m_2+m_2' & m+m' & \mathrm{SO}(4n)\\ (r_1+r_1',s_1+s_1') & (r_2+r_2',s_2+s_2') & (r+r',s+s') & \mathrm{SO}(4p)\times\mathrm{SO}(4q)\end{matrix}\right), \tag{49}$$

where N, N', and N'' are expressions (43) or (45) for the representations $T^{\mathrm{Sp}(n)}_{mm'}$, $T^{\mathrm{Sp}(n)}_{m_1m_1'}$, and $T^{\mathrm{Sp}(n)}_{m_2m_2'}$, respectively.

Using the concept of dual groups, Ališauskas[13] has found similar relations for CGC's; however his CGC's are not attached to some bases. Our CGC's are related with bases. The HK-associated spherical functions can be taken as these bases. Moreover, we explain the structure of coefficients relating CGC's of different groups.

It is very important to generalize relations (48) and (49) for more general representations of the groups $\mathrm{SO}(n)$, $\mathrm{U}(n)$, and $\mathrm{Sp}(n)$: It would give us a chance to apply the powerful results found by Biedenharn, Louck, and others (see Refs. 14 and 15 and the literature quoted therein) to the CGC's of the groups $\mathrm{SO}(n)$ and $\mathrm{Sp}(n)$.

[1]N. Ja. Vilenkin, *Special Functions and the Theory of Group Representations* (Am. Math. Soc., Providence, RI, 1968).

[2]L. C. Biedenharn, J. Nuyts, and N. Straumann, Ann. Inst. H. Poincaré **3A**, 13 (1965).

[3]L. C. Biedenharn, in *Cargese Lectures in Theoretical Physics* (Gordon and Breach, London, 1967).

[4]A. U. Klimyk, *Matrix Elements and Clebsch–Gordan Coefficients of Group Representations* (Naukova Dumka, Kiev, 1979).

[5]A. U. Klimyk and A. M. Gavrilik, J. Math. Phys. **20**, 1624 (1979).

[6]N. Ja. Vilenkin and A. U. Klimyk, Dokl. Akad. Nauk Ukr. SSR Ser. A **7**, 9 (1987).

[7]S. Helgason, *Differential Geometry and Symmetric Spaces* (Academic, New York, 1962), Chap. 10.

[8]B. Hoogenboom, *Intertwining Functions on Compact Lie Groups* (Mathematische Centrum, Amsterdam, 1983).

[9]C. F. Duncl, Symp. Math. **22**, 145 (1977).

[10]W. Rudin, *Function Theory in the Unit Ball of C^n* (Springer, Berlin, 1980), Chap. 12.

[11]N. Ja. Vilenkin and R. L. Shapiro, Izv. Vyshch. Uchebn. Zaved. Mat. **62**, 9 (1967).

[12]R. L. Shapiro, Izv. Vyshch. Uchebn. Zaved. Mat. **71**, 97 (1968).

[13]S. J. Ališauskas, Sov. J. Part. Nucl. **14**, 563 (1983).

[14]L. C. Biedenharn, M. A. Lohe, and J. D. Louck, J. Math. Phys. **26**, 1458 (1985).

[15]J. D. Louck and L. C. Biedenharn, J. Math. Phys. **11**, 2368 (1970).

Some results from a Mellin transform expansion for the heat kernel

A. P. C. Malbouisson and F. R. A. Simão
Centro Brasileiro de Pesquisas Físicas, CBPF/CNPq, Rua Dr. Xavier Sigaud, 150, 22290 Rio de Janeiro, RJ, Brazil

A. Ferraz de Camargo Fo.
Instituto de Física Teórica, Universidade Estadual Paulista, Rua Pamplona, 145, 01405 São Paulo, SP, Brazil

(Received 23 June 1988; accepted for publication 21 December 1988)

In the case of a differential operator containing a gauge field, coefficients of a new heat kernel expansion obtained in a preceding paper (A. P. C. Malbouisson, M. A. R. Monteiro, and F. R. A. Simão, CBPF-NF-024/88, to be published in J. Math. Phys.) are calculated. The prior expansion allows it to be shown that the meromorphic structure of the generalized zeta function is much richer than was known previously. Also, an application to anomalies is done, resulting in a general formula for the arbitrary dimension D. The special cases $D = 2$ and $D = 3$ are investigated.

I. INTRODUCTION

In a previous paper[1] an asymptotic expansion was obtained for the diagonal part of the heat kernel associated with a given elliptic operator H of order m, based on the connection, through a Mellin transform, between the heat kernel and the Seeley's kernel $K(s;x,y)$[2] of the complex sth power H^s of the operator H and the meromorphic properties of $K(s;x,x)$. We recall that "heat kernel" means the solution of the "heat equation"

$$\frac{\partial}{\partial t} F(t;x,y) = HF(t;x,y), \tag{1.1}$$

where t is a "time" or "temperature" parameter and x and y are, in the case we are interested in, points of a D-dimensional compact manifold M. The Seeley's kernel is defined for $\mathrm{Re}(s) < -D/m$ such that

$$H^s f(x) = \int_M dy\, K(s;x,y) f(y).$$

The expansion mentioned above is obtained by analytic continuation of K in the variable s and reads as

$$F(t;x,x) = -\sum_{l=0}^{\infty} t^l \left(\frac{d\phi}{ds}\right)_{s=l} - \sum_j t^{(j-D)/m} \Gamma\left(\frac{D-j}{m}\right) R_j(x). \tag{1.2}$$

The sum over j is such that we take $j = 0,1,2,...$ *excluding the terms such that* $(j-D)/m = 0,1,2,...$ and $R_j(x)$ is the residue of $K(s;x,x)$ at the pole $s = (j-D)/m$:

$$R_j(x) = \frac{1}{im(2\pi)^{D+1}} \int_{\|\xi\|=1} \int_\Gamma d\lambda \times \lambda^{(j-D)/m} b_{-m-j}(x,\xi,\lambda), \tag{1.3}$$

where Γ is a curve coming from ∞ along a ray of minimal growth, clockwise on a small circle around the origin, and then going back to ∞. The quantities b_{-m-j} are obtained from the coefficients of the symbol of H (see Sec. III) and $\|\xi\| = 1$ means that the set of variables $\{\xi\}$ is constrained to be at the surface of the unit sphere in D-dimensional space. The function $\phi(s)$ is introduced to account for the coincidence of the poles of the gamma function $\Gamma(-s)$ and those of $K(s;x,x)$ at the positive integers l and is defined by

$$\Gamma(-s)K(s;x,x) \approx \phi(s)/(s-l)^2 \tag{1.4}$$

for $s \approx l$.

As was remarked in Ref. 1, the expansion (1.2) is rather different from de Witt's ansatz currently used.[3] In particular (1.2) contains fractionary powers at even dimension and even operator order, coming from the second term in the expansion.

In the rest of the paper we explore some consequences of the new expansion (1.2). In Sec. II we show that the generalized zeta function $\zeta(s)$ has an infinity of poles at real values of s. In Sec. III we calculate the coefficients of the leading and next-to-leading terms in (1.2). In Sec. IV we obtain a general formula for the anomaly in arbitrary dimension D and particularize to the special cases $D = 2$ and $D = 3$.

II. MEROMORPHY OF THE GENERALIZED ZETA FUNCTION

One of the implications of the series (1.2) is of a mathematical character and concerns the meromorphic structure of the Hawking's generalized zeta function,[4] which is much richer than the structure known previously. This may be easily seen as follows.

The generalized zeta function is written as

$$\zeta(s) = \frac{1}{\Gamma(s)} \int_0^1 dt\, t^{s-1} \int d^D x\, F(t;x,x) + Q(s), \tag{2.1}$$

where $Q(s)$ converges for all s.

Let us take $D = 4$ and consider an operator of order $m = 2$. Replacing $F(t;x,x)$ in (2.1) by the series (1.2) we see that the first term of the expansion gives no poles as a result of the factor $1/\Gamma(s)$ in front of the integral in (2.1). From the second term of the expansion we have the sum

 0022-2488/89/061226-04$02.50

$$-\frac{1}{\Gamma(s)}\sum_j \Gamma\left(\frac{4-j}{2}\right)R_j(x)\int_0^1 dt\, t^{s+j/2-3}, \qquad (2.2)$$

which gives poles at $s=2-j/2$ for integer values of j and $(j-4)/2\neq 0,1,2,\ldots$.

Thus the poles of the generalized zeta function are not situated only at $s=1$ ($j=2$) and $s=2$ ($j=0$). We also have poles at $s=\frac{3}{2}$ ($j=1$) and $s=\frac{1}{2}$ ($j=3$); for $j=5,7,\ldots$ we have an infinity of poles in s at the negative half-integers. There are no poles at negative integers, as a result of the vanishing of the residues of $K(s;x,x)$ at those values.[2] The residues at the poles are given by the corresponding coefficients $-[1/\Gamma(2-j/2)]\Gamma[(4-j)/2]R_j(x)$ in (2.2).

III. APPLICATION TO A DIFFERENTIAL OPERATOR

Let us consider a differential operator H of order $m=2$,

$$H=-[g^{\mu\nu}(x)(\partial_\mu+B_\mu(x))(\partial_\nu+B_\nu(x))+P(x)], \qquad (3.1)$$

acting on a D-dimensional compact manifold M and endowed with a metric $g_{\mu\nu}(x)$ ($\mu,\nu=1,2,\ldots,D$). In (3.1) $P(x)$ is a nondifferential operator and

$$B_\mu(x)=gA_\mu(x)+\eta_\mu(x), \qquad (3.2)$$

where $A_\mu(x)$ and g are, respectively, the gauge field and a coupling constant (not to be confused with the metric tensor or its determinant). The quantity $\eta_\mu(x)$ contains information about curvature and torsion. The usual convention of summation over repeated indices will be adopted.

In Seeley's notation[2] the operator H must be written in the form

$$H=\sum_{\{\alpha\}|\alpha|\leq 2}(-i)^{|\alpha|}H^{|\alpha|}_{\alpha_1\cdots\alpha_D}(x)\frac{\partial^{|\alpha|}}{\partial x_1^{\alpha_1}\cdots\partial x_D^{\alpha_D}}, \qquad (3.3)$$

where $|\alpha|=\alpha_1+\cdots+\alpha_D$.

Expanding (3.1) and comparing with (3.3) we obtain the set of coefficients $H^{|\alpha|}_{\alpha_1\cdots\alpha_D}(x)$:

$$H^{(2)}_{\alpha_1\cdots\alpha_D}(x)=H^{(2)}_{0\cdots01(\mu)0\cdots01(\nu)0\cdots0}(x)=g_{\mu\nu}(x), \qquad (3.4a)$$

$$H^{(1)}_{\alpha_1\cdots\alpha_D}(x)=H^{(1)}_{0\cdots01(\nu)\cdots0}=-2ig_{\mu\nu}(x)B^\mu(x), \qquad (3.4b)$$

$$H^{(0)}_{\alpha_1\cdots\alpha_D}(x)=H^{(0)}_{0\cdots0\cdots0}(x)=-g_{\mu\nu}(x)(2\partial^\mu B^\nu+B^\mu B^\nu)-P(x). \qquad (3.4c)$$

Now, to calculate the coefficients of the second term of expansion (1.2) we need the quantities b_{-2-j} [see Eq. (1.3)], which are expressed in terms of the coefficients $a_{2-k}(x,\xi)$ of the symbol of H[2],

$$a_{2-k}(x,\xi)=\sum_{|\alpha|=2-k}H^{|2-k|}_{\alpha_1\cdots\alpha_D}\xi_1^{\alpha_1}\cdots\xi_D^{\alpha_D}, \qquad (3.5)$$

by the following set of equations:

$l=0$:

$$b_{-2}[a_2(x,\xi)-\lambda]=1, \qquad (3.6a)$$

$l>0$:

$$b_{-2-l}(a_2-\lambda)+\sum_{j,k}(-i)^{|\alpha|}\sum_{\{\alpha\}}\frac{\partial^{|\alpha|}b_{-2-j}}{\partial\xi_1^{\alpha_1}\cdots\partial\xi_D^{\alpha_D}}\times\frac{\partial^{|\alpha|}a_{2-k}}{\alpha_1!\cdots\alpha_D!\partial_{x_1}^{\alpha_1}\cdots\partial_{x_D}^{\alpha_D}}=0,$$

$$\text{with } j<l,\quad j+k+|\alpha|=l. \qquad (3.6b)$$

The coefficients a_{2-k} are easily obtained from Eqs. (3.4):

$$a_2(x,\xi)=g_{\mu\nu}(x)\xi^\mu\xi^\nu\equiv\|\xi\|^2, \qquad (3.7a)$$

$$a_1(x,\xi)=-2ig_{\mu\nu}(x)B^\mu(x)\xi^\nu, \qquad (3.7b)$$

$$a_0(x,\xi)=-g_{\mu\nu}(x)(\partial^\mu B^\nu-B^\mu B^\nu)-P(x). \qquad (3.7c)$$

Then the first two quantities b_{-2-j} that we need for calculating the leading and next-to-leading contributions in the second term of expansion (1.2) are given by

$$b_{-2}(x,\xi,\lambda)=(\|\xi\|^2-\lambda)^{-1}, \qquad (3.8)$$

$$b_{-3}(x,\xi,\lambda)=\frac{2iB\cdot\xi}{(\|\xi\|^2-\lambda)^2}-\frac{2i\xi\cdot\partial\|\xi\|^2}{(\|\xi\|^2-\lambda)^3}, \qquad (3.9)$$

where the scalar product is defined with the metric $g_{\mu\nu}(x)$.

From (1.2), (1.3), (3.8), and (3.9), the contributions that are coefficients of the powers $t^{-D/2}$ and $t^{(1-D)/2}$, are given, respectively, by

$$-\Gamma\left(\frac{D}{2}\right)R_0(x)=-\Gamma\left(\frac{D}{2}\right)\frac{1}{2i(2\pi)^{D+1}}\int_{\|\xi\|=1}d\xi\times\int_\Gamma d\lambda\,\lambda^{-D/2}(\|\xi\|^2-\lambda)^{-1}, \qquad (3.10)$$

$$-\Gamma\left(\frac{D-1}{2}\right)R_1(x)=-\Gamma\left(\frac{D-1}{2}\right)\frac{1}{(2\pi)^{D+1}}\left[\int_{\|\xi\|=1}d\xi\,B\cdot\xi\times\int_\Gamma\frac{\lambda^{(1-D)/2}d\lambda}{(\|\xi\|^2-\lambda)^2}-\int_{\|\xi\|=1}d\xi\,\xi\cdot\partial\|\xi\|^2\times\int_\Gamma\frac{\lambda^{(1-D)/2}d\lambda}{(\|\xi\|^2-\lambda)^3}\right], \qquad (3.11)$$

where we take the integration path Γ as the curve coming from $-\infty$ along the negative real axis, then clockwise along the unit circle around the origin, and then backward to $-\infty$ along the negative real axis. Since we must restrict the ξ's to the surface of the unit D-dimensional sphere, the last integral in (3.11) vanishes; to avoid the singularity at $\lambda=1$, we introduce a regulator $p>1$.[2] Then (3.10) and (3.11) become

$$-\Gamma\left(\frac{D}{2}\right)R_0(x)=\Gamma\left(\frac{D}{2}\right)\frac{1}{2\cdot(2\pi)^{D+1}}\int d\xi\times\left[2\sin\left(\frac{\pi D}{2}\right)\int_{-1}^{-\infty}\frac{d\lambda\,|\lambda|^{-D/2}}{p-\lambda}-i\int_\pi^{-\pi}\frac{d\theta\,e^{i(1-D/2)\theta}}{p-e^{i\theta}}\right], \qquad (3.12)$$

$$-\Gamma\left(\frac{D-1}{2}\right)R_1(x)$$
$$=-\Gamma\left(\frac{D-1}{2}\right)\frac{1}{(2\pi)^{D+1}}\int d\xi \times B\cdot\xi\left[-2i\sin\left(\frac{\pi(1-D)}{2}\right)\int_{-1}^{-\infty}\frac{d\lambda\,|\lambda|^{(1-D)/2}}{(p-\lambda)^2}+i\int_{\pi}^{-\pi}\frac{d\theta\, e^{(i/2)(3-D)\theta}}{(p-e^{i\theta})^2}\right]. \tag{3.13}$$

In (3.12), (3.13), and the subsequent formulas, the integrations over the ξ's are constrained to the unit sphere $\|\xi\|\equiv\sqrt{g_{\mu\nu}(x)\xi^\mu\xi^\nu}=1$.

In dimension $D=4$, making the change of variables $p^{-1/2}e^{i\theta/2}=e^{i\phi}$, the integrations over λ and θ may be performed. The results, after suppression of the regularization, are

$$-\Gamma(2)R_0(x)=\frac{1}{2(2\pi)^4}\int d\xi \tag{3.14}$$

and

$$-\Gamma\left(\frac{3}{2}\right)R_1(x)=-\Gamma\left(\frac{3}{2}\right)\frac{3i}{2(2\pi)^4}\int d\xi\, B\cdot\xi. \tag{3.15}$$

Analogously, in dimension $D=2$, the coefficients of the two first powers of the second term in (1.2) (powers t^{-1} and $t^{-1/2}$, respectively) are obtained from (3.12) and (3.13):

$$-\Gamma(1)R_0(x)=\frac{1}{2(2\pi)^2}\int d\xi, \tag{3.16}$$

$$-\Gamma\left(\frac{1}{2}\right)R_1(x)=-\Gamma\left(\frac{1}{2}\right)\frac{i}{2(2\pi)^2}\int d\xi\, B\cdot\xi. \tag{3.17}$$

As an example, we calculate the coefficients (3.16) and (3.17) in the Penrose compactified two-dimensional Minkowski space,[5] which has the metric

$$\bar{g}_{\mu\nu}=\tfrac{1}{2}\begin{pmatrix}0 & 1\\ 1 & 0\end{pmatrix}.$$

In this case the unit sphere $\|\xi\|=1$ is the section of hyperbola depicted in Fig. 1. Using polar coordinates (r,θ) and the well-known formula for the induced metric on a $(D-1)$-dimensional surface embedded in D-dimensional metric space, it is easy to see that the integration on the "surface" $\|\xi\|=1$ reduces simply to integration over θ between the limits θ_1, θ_2 and $\theta_1+\pi$, $\theta_2+\pi$:

$$\int_{\|\xi\|=1}d\xi=\int_{\theta_1}^{\theta_2}d\theta+\int_{\theta_1+\pi}^{\theta_2+\pi}d\theta,$$

with

$$\theta_1=\arctan(1/\pi^2), \tag{3.18a}$$

$$\theta_2=\arctan\pi^2. \tag{3.18b}$$

We obtain

$$-\Gamma(1)R_0(x)=[1/(2\pi)^2](\theta_2-\theta_1), \tag{3.19}$$

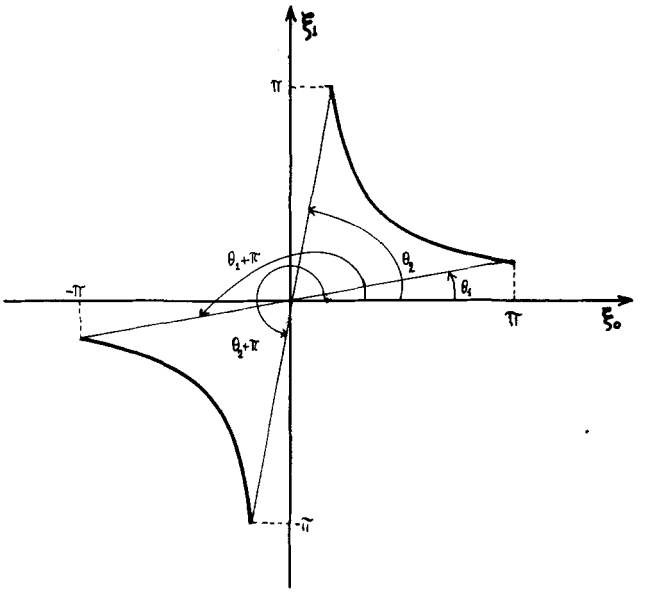


FIG. 1. We show ξ variables submitted to the constraint $g_{\mu\nu}\xi^\mu\xi^\nu=1$ in the Penrose compactified two-dimensional Minkowski space.[5]

$$-\Gamma\left(\frac{1}{2}\right)R_1(x)$$
$$=-\Gamma\left(\frac{1}{2}\right)\frac{i}{(2\pi)^2}\sqrt{2}\left\{\left(\frac{1+i}{2}\right)\Pi\left(\alpha,\frac{1+2}{2},\frac{1}{\sqrt{2}}\right)+F\left(\alpha,\frac{1}{\sqrt{2}}\right)-2E\left(\alpha,\frac{1}{\sqrt{2}}\right)\right\}\Bigg|_{\theta_1}^{\theta_2}$$
$$+(\theta_1\to\theta_1+\pi,\theta_2\to\theta_2+\pi), \tag{3.20}$$

where Π and E are the elliptic integrals of the third and second kinds, respectively; F is the generalized hypergeometric series.

IV. ANOMALIES

In this section we apply expansion (1.2) to study anomalies using the heat kernel method.[6] We borrow some of the notation and methods employed in a recent work by Cognola and Zerbini[7] since they are suitable for our purposes. Using the generalized zeta-function regularization, the anomaly may be written in the form

$$\mathrm{A}=-q\lim_{s\to 0}\mathrm{Tr}\left\{(X+Y)\frac{1}{\Gamma(s)}\int_0^1 dt\, t^{s-1}\times\left[F(t;x,x)-P_0(x,x)\right]\right\}, \tag{4.1}$$

where $q=-1$, $\frac{1}{2}$, or 1 for fermions, neutral or charged bosons, respectively; $X=X_1+X_2$ and $Y=Y_1+Y_2$ are operators satisfying the relation $\delta K(J)=(\delta JX_1+Y_1\delta J)K+K(Y_2\delta J+\delta JX_2)$ and $K(J)$ is such that $H(J)\propto K(J)$ for bosons and $H(J)=K^2(J)$ for fermions, where J is a classical source. Here P_0 is the projector onto the zero modes. For the axial anomaly, $X=Y=i\gamma_5$.

In (4.1) we replace $F(t;x,x)$ by (1.2); after some simple manipulations we see that the sole contribution to the anom-

aly comes from the coefficient of the power t^0, giving, for arbitrary dimension D,

$$A = -q\,\mathrm{Tr}\left\{(X+Y)\left[-\left(\frac{d\phi}{ds}\right)\Big|_{s=0} - P_0(x,x)\right]\right\}. \tag{4.2}$$

Now, from (1.4) and the formula

$$\Gamma(z) = \frac{\Gamma(z+l+1)}{z+l}\prod_{n=1}^{l}\frac{1}{z+l-n},$$

we have, for integer $l \geqslant 0$,

$$\frac{d\phi}{ds}\Big|_{s=l} = -2\frac{(-1)^l}{l!}K(l;x,x), \tag{4.3}$$

where the Seeley's kernel for integer l is[2]

$$K(l;x,x) = \frac{1}{(-1)^l 2(2\pi)^D}\int d\xi \int_0^\infty dt \times t^l b_{-2-2l-D}(x,\xi,te^{i\theta}). \tag{4.4}$$

Thus taking $\arg\lambda = \theta = \pi$ in (4.4) the anomaly may be obtained for *arbitrary* dimension D from (4.2), with

$$\frac{d\phi}{ds}\Big|_{s=0} = \frac{-1}{(2\pi)^D}\int d\xi \int_0^\infty dt\, b_{-2-D}(x,\xi,-t). \tag{4.5}$$

Next we apply (4.2) to the cases $D=2$ and $D=3$. The case $D=3$ is particularly interesting since, in spite of the well-known difficulties in defining the matrix γ_5 in odd dimension,[8] certain aspects of even-dimensional axial anomaly could appear in odd-dimensional field theories (see Niemi and Semenoff[9] and the references therein). This results from the fact that the connection between zero modes of Dirac operators and nontriviality of the background field topology is valid for any value of D, as shown by Callias.[10]

Moreover, there is a technical difficulty to (formally) calculating anomalies in odd dimension using the de Witt ansatz in the heat kernel method which is not present with our expansion: When one uses the de Witt ansatz for expanding $F(t)$, the anomaly depends on the coefficient of the power $t^{D/2}$, which does not exist for odd values of D, while with our expansion the anomaly depends directly on the coefficient of the zeroth power of t, given by (4.5), for any even or odd dimension.

Calculations for a general coordinate-dependent metric are extremely involved. Here, we restrict ourselves to the simpler situation of a symmetric, coordinate-independent metric tensor $g_{\mu\nu}$. In this case we obtain

for $D=2$:

$$A_2 = \frac{q}{(2\pi)^2}\mathrm{Tr}\left\{(X+Y)\int d\xi\left[4\xi^\mu(\partial_\mu B_\nu)\xi^\nu + ig_{\mu\nu}(\partial^\mu B^\nu - B^\mu B^\nu) + P(x) + 2(B\cdot\xi)^2\right]\right\}, \tag{4.6}$$

for $D=3$:

$$A_3 = \frac{q}{(2\pi)^3}\mathrm{Tr}\left\{(X+Y)\int d\xi\left[-g^{\mu\nu}(\partial_\mu\partial_\nu B^\sigma)\xi_\sigma + 2i(B\cdot\xi)(g_{\mu\nu}(\partial^\mu B^\nu - B^\mu B^\nu) + P(x)) - 2iB^\mu(\partial_\mu B^\nu)\xi_\nu + i\xi^\mu g_{\rho\sigma}(\partial_\mu\partial^\rho B^\sigma + B^\rho\partial_\mu B^\sigma + (\partial_\mu B^\rho)B^\sigma) + i\xi^\mu\partial_\mu P(x) - \frac{8}{3}(B\cdot\xi)^3 - \frac{4}{3}\xi^\mu\xi_\sigma(B\cdot\xi)(\partial_\mu B^\sigma) - \frac{16}{3}i\xi^\mu\xi^\nu\xi_\sigma(\partial_\mu\partial_\nu B^\sigma)\right]\right\}. \tag{4.7}$$

In the Penrose compactified two-dimensional Minkowski space,[5] (4.6) gives the result

$$A_2 = \frac{q}{(2\pi)^2}\mathrm{Tr}\left\{(X+Y)\left[(\theta_2-\theta_1)\left((1/2+i)\times\left(\frac{\partial B_1}{\partial x^0} - \frac{\partial B_0}{\partial x^1}\right) + (2-i)B_0B_1 + P(x)\right) + \tfrac{1}{2}\ln\frac{\sin\theta_2}{\sin\theta_1}\left(\frac{B_1^2}{2} + \frac{\partial B_1}{\partial x^1}\right) - \tfrac{1}{2}\ln\frac{\cos\theta_2}{\cos\theta_1}\left(\frac{B_0^2}{2} + \frac{\partial B_0}{\partial x^0}\right)\right]\right\},$$

where the angles θ_1, θ_2 are given by (3.18a) and (3.18b) and B_0, B_1 are the components of $B_\mu(x)$ given by (3.2).

[1] A. P. C. Malbouisson, M. A. R. Monteiro, and F. R. A. Simão, to be published in J. Math Phys.
[2] R. T. Seeley, Am. Math. Soc. Proc. Symp. Pure Math. **10**, 288 (1967).
[3] See B. de Witt, *The Dynamical Theory of Groups and Fields* (Gordon and Breach, New York, 1965); see, also, Birrel and Davies in Ref. 5.
[4] S. W. Hawking, Commun. Math. Phys. **55**, 133 (1977).
[5] R. Penrose, in *Relativity, Groups and Topology*, edited by B. de Witt and C. de Witt (Gordon and Breach, New York, 1964); N. D. Birrel and P. C. W. Davies, *Quantum Fields in Curved Spaces* (Cambridge U.P., Cambridge, 1984).
[6] K. Fujikawa, Phys. Rev. Lett. **42**, 1195 (1979); Phys. Rev. D **21**, 2848 (1980); A. M. Polyakov and P. B. Wilyman, Phys. Lett. B **131**, 121 (1983); O. Alvarez; Nucl. Phys. B **238**, 61 (1984); R. E. Gamboa Saravi, F. A. Schaposnik, and J. E. Solomin, Phys. Rev. D **30**, 1353 (1984); L. C. L. Botelho and M. A. do Rego Monteiro, *ibid.* **30**, 2242 (1984); A. P. Balachandran, G. Marmo, Y. P. Nair, and C. G. Traherm, *ibid.* **25**, 2713 (1982).
[7] G. Cognola and S. Zerbini, Phys. Lett. B **195**, 435 (1988); Mod. Phys. Lett. A **3**, 599 (1988).
[8] T. L. Trueman, Phys. Lett. B **88**, 331 (1979); G. 't Hooft and M. Veltman, Nucl. Phys. B **44**, 189 (1972); G. Bonneau, *ibid.* **171**, 477 (1980).
[9] A. J. Niemi and G. W. Semenoff, Phys. Rev. Lett. **54**, (9), 873 (1985).
[10] C. Callias, Commun. Math. Phys. **62**, 213 (1978).

Comment on "A metric space construction for the boundary of space-time" by D. A. Meyer [J. Math. Phys. 27, 124 (1986)]

Marcus Kriele
Technische Universität Berlin, Fachbereich 3-Mathematik, 1000 Berlin 12, West Germany

(Received 10 September 1988; accepted for publication 18 January 1989)

For his metric space construction Meyer incorrectly derived several results. Meyer's proofs are corrected or disproved by counterexamples. The physical relevance of his construction is further examined.

I. INTRODUCTION

Meyer[1] has introduced a new boundary for space-time. His construction is described as follows: Let (M,g) be a paracompact, connected C^∞-Hausdorff manifold with a Lorentz metric of signature $(-,+,\ldots,+)$. Furthermore, suppose that the timelike diameter of (M,g) is finite and that (M,g) is a distinguishing space-time. The height of a set $U\subset M$ is defined as $d(U) := \sup\{L(\gamma)/\gamma \text{ is a causal curve in } U\}$, where $L(\gamma)$ is the arclength of γ.

Define $D^+: M\times M\to\mathbb{R}$

$(p,q)\mapsto d((I^+(p)-I^+(q))\cup(I^+(q)-I^+(p)))$;

D^- is defined similarly. Then $D := D^+ + D^-$ is a topological metric on M. Denote by $\mathscr{M}$ the manifold topology and by $\mathscr{D}$ the metric topology of M. When $(M,\bar{D})$ is the metric completion of (M,D) (with topology $\overline{\mathscr{D}}$), $\partial_D M := \bar{M}-M$ is the D boundary of space-time.

For this construction Meyer derived several results on the topology of space-time. Unfortunately, a majority of these results remains erroneous: partly they are wrong and partly their proof is not correct. The purpose of this work is to correct the wrong proofs and to disprove the wrong statements by counterexamples. Furthermore, I shall sketch a generalization of the D boundary to space-times with arbitrary timelike diameter and discuss its physical relevance.

II. CORRECTIONS

In Theorem 3.1 Meyer states that D is continuous iff $\mathscr{M}$ agrees with $\mathscr{D}$. But Meyer's proof that $\mathscr{D}$ is finer than $\mathscr{M}$ whenever D is continuous is wrong. In "$\Rightarrow$. (ii)" he claims that, given a neighborhood of a point $p\in M$ with compact closure, the following statement holds for all $r\in M$: If $I^-(r)\cap I^-(p)\cap\partial U=\emptyset$ and $I^-(r)\cap I^-(p)\neq\emptyset$, then there exists $q\in \mathrm{cl}_{/\!/}(\partial U\cap I^-(p))$ with $q\in I^+(t)$ for all $t\in I^-(p)\cap I^-(r)$. [$\mathrm{cl}_{/\!/}(\cdot)$ denotes closure in $\mathscr{M}$ topology.] Cylindrical two-dimensional Minkowski space $((0,1)\times S^1, -dt^2+dx^2)$ provides a counterexample (Fig. 1).

$I^+(t_1)\cap I^+(t_2)\cap \mathrm{cl}_{\mathscr{M}}(\partial U\cap I^-(p))=\emptyset$, so his assertion cannot be correct. Fortunately, Rübe has proven that in any case $\mathscr{D}$ is finer than $\mathscr{M}$. I sketch her proof, because in some applications it might be useful to have a finer metric topology than $\mathscr{M}$ which depends only on the causal structure of (M,g).

Theorem: Let (M,g) be a distinguishing space-time with finite timelike diameter. Then for every sequence $\{q_n\}$ in M with $D(q_n,p)\to 0$ $(n\to\infty)$ one has $q_n\to p$ with respect to $\mathscr{M}$.

Proof: Suppose there exists a sequence $\{q_n\}$ with $D(q_n,p)\to 0$, but $q_n\not\to p$. By taking a subsequence one can choose an open neighborhood U of p with compact closure contained in some convex neighborhood such that no q_n hits U. Since $I^+(q_n)\cap I^-(q_n)=\emptyset$, we can assume without loss of generality that $p\notin I^+(q_n)$ for all n. Let γ: $[0,1]\to U$ be a future directed timelike curve with $\gamma(0)=p$ and $L(\gamma|[0,1/n))>D(q_n,p)$. We have $\gamma(1/n)\in I^+(q_n)$. Choose a timelike curve λ_n from q_n to $\gamma(1/n)$ and denote by x_n its last intersection with ∂U. If x is a limit point of the x_n, there exists[2] a causal curve λ from x to p. Clearly, $x\neq p$ and $I^+(p)\subset I^+(x)$. To get our contradiction, we show that also $I^+(x)\subset I^+(p)$. Assume that there exists $y\in I^+(x)-I^+(p)$. Let $z\in I^+(x)\cap I^-(y)$ and η a timelike curve from z to y. Since $I^-(z)$ is a neighborhood of x, we find a subsequence q_{n_k} and causal curves from q_{n_k} via x_{n_k} to z which can be lengthened by η to y. Denote the resulting curves by α_{n_k}. But then we have $D(q_{n_k},p)\geqslant L(\alpha_{n_k})\geqslant L(\eta)>0$ which contradicts $D(q_n,p)\to 0$.

Therefore $I^+(x)=I^+(p)$ holds, which is impossible because (M,g) is distinguishing.

Proposition 3.7 in Meyer's paper is wrong: He defines a C^0 topology on the space of causal future inextendible curves $\mathscr{C}$ by taking $\{\mathscr{C}_R(P,Q) := \{\gamma/\gamma \text{ is a future inextendible causal curve in } R \text{ with initial point in } P \text{ which reaches } Q \text{ and eventually remains in } Q\}/P,\ R\in\mathscr{M},\ Q=I^+(q),\ q\in M\}$ as a basis. In Proposition 3.7 he states: "D^+ is continuous at p iff L: $\mathscr{C}\to\mathbb{R}$ is continuous across the null cone of p." The "only if" part is false (Fig. 2). In order that the null geodesic γ reaches $Q=I^+(q)$ (in our example this choice for Q suffices), it is necessary that q lies in the past of some points on γ. But then there is a timelike curve λ which is in every C^0 neighborhood of γ, and L cannot be continuous in γ.

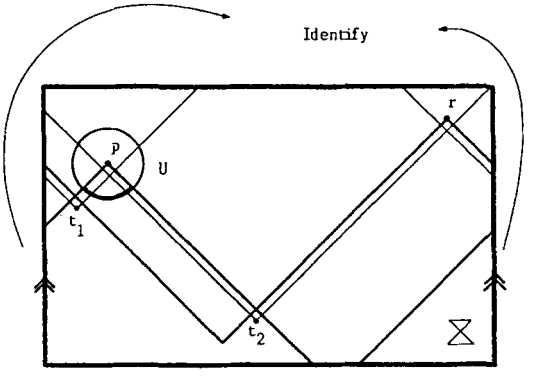


FIG. 1. Cylindrical Minkowski space.

 0022-2488/89/061230-03$02.50

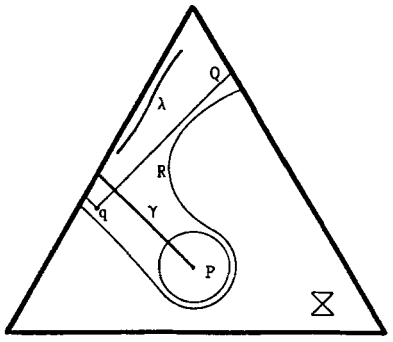


FIG. 2. L is not continuous across the null cone.

Meyer defines future boundary points (FBP's), but they do not exist: $p \in \partial_D M$ is said to be a FBP if there is a causal curve γ: $[a,b) \to M$, $c \in [a,b)$ such that for all $\overline{\mathscr{D}}$ neighborhoods U of p there exists $t_0 \in (a,c)$ with $\gamma(t) \in U$ for all $t \in (t_0,c)$. Since $\gamma(c) \in M$, there exists a $\mathscr{M}$-neighborhood V of $\gamma(c)$ with compact closure. Because $\mathscr{D}$ is finer than $\mathscr{M}$, $\mathrm{cl}_{\mathscr{D}} V \subset \mathrm{cl}_{\mathscr{M}} V$ is a neighborhood of $\gamma(c)$ and, consequently, $M - \mathrm{cl}_{\overline{\mathscr{D}}} V = M - \mathrm{cl}_{\mathscr{D}} V$ is a neighborhood of p in $(M,\overline{\mathscr{D}})$ which is not hit by any $\gamma(t)$ sufficiently close to $\gamma(c)$. A future boundary end point (FBE) is defined similarly, but now $c = b$ is required. For future inextendible causal curves FBE's may exist. Past boundary end points (PBE's) are defined analogously.

Assume now that the c-boundary construction of Geroch, Kronheimer, and Penrose[3] works (our notation is, unless explicitly defined, as in this paper). In order to compare D boundary with c boundary, Meyer defines the map $i:(\bar{M},\overline{\mathscr{D}}) \to (M_h^{\#},\mathscr{M}^{\#})$,

$$p \to \begin{cases} (I'^-(p) \cap M)^{\#}, & \text{if } I'^-(p) \neq \varnothing, \\ (I'^+(p) \cap M)^{\#}, & \text{if } I'^+(p) \neq \varnothing, \\ \varnothing, & \text{if } I'^-(p) = I'^+(p) = \varnothing, \end{cases}$$

where $(M_h^{\#},\mathscr{M}^{\#})$ ($= \overline{M}_c$ in Meyer's notation) is the Hausdorff space resulting from the causal completion procedure,[3] and $I'^-(p) := \{q \in \overline{M} /$ there is a nontrivial future directed causal curve in M from q (or with PBE q) to p (or with FBE p, if $p \in \partial_D M)\}$. $I'^+(p)$ and $J'^{\pm}(p)$ are defined analogously.

First, note that $\varnothing \notin M_h^{\#}$. But this problem can be solved by introducing an abstract point e: $M'^{\#} := M_h^{\#} \cup \{e\}$; let $\{e\}$ be open and $\mathscr{M}'^{\#}|M_h^{\#} = \mathscr{M}^{\#}$.

Second, it remains to check whether

$$(I'^+(p) \cap M)^{\#} = (I'^-(p) \cap M)^{\#},$$

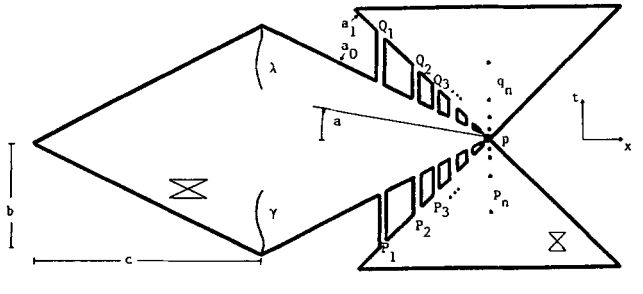


FIG. 3. Here i is not well defined. (This is preparing the example for the three-dimensional counterexample.)

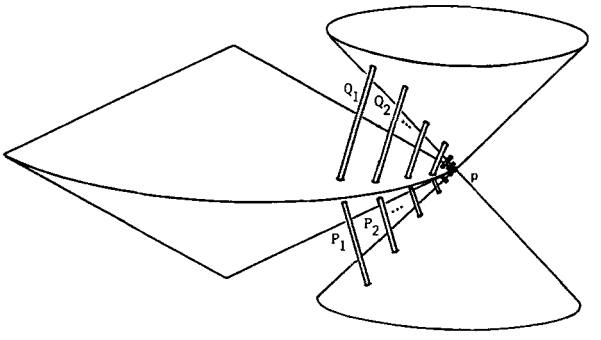


FIG. 4. Here i is not well defined (counterexample).

whenever $I'^+(p) \neq \varnothing$ and $I'^-(p) \neq \varnothing$. Meyer's proof is wrong because he claims that every open $\overline{\mathscr{D}}$ neighborhood of $I'^-(p) \cap M$ must intersect $I'^+(p) \cap M$, which is obviously false. By giving a three-dimensional counterexample I show that i is ill-defined. First, consider the following two-dimensional example (Fig. 3). We have two sequences of tubes $\{Q_n\}$, $\{P_n\}$ which connect region III with region I (region II, respectively) and accumulate at p. Our metric is $g := -dt^2 + f(a)dx^2$, where f is a smooth function of the angle a with $f(a) = 1$ ($a < -a_1$ or $a > a_1$) and $f(a) = c/b$ ($-a_0 < a \leqslant a_0$). Here $I^-(\{p_n\})$ is in $I^+(\gamma)^{\mathrm{ext}}$ but not in $I^-(\lambda)^{\mathrm{ext}}$, and for $\{q_n\}$ the dual statement holds. Moreover, we have $I^+(\gamma)^{\mathrm{ext}} \cap I^-(\lambda)^{\mathrm{ext}} = \varnothing$. If we had $p_n \to p$, $q_n \to p$ (in $\overline{\mathscr{D}}$ topology), this would imply that i was not well defined, because then $I'^-(p)\,M$ and $I'^+(p)\,M$ could be separated by two elements of $\mathscr{M}^{\#}$ which have an empty intersection. However, they do not converge, for a timelike future directed curve, starting at p_n, can reach the first tube Q_1, whence $D(p_n,q_n) \geqslant L(Q_1)$ for all n,m. Adding one dimension as in Fig. 4 we can achieve that the tubes $\{Q_n\}$, $\{P_n\}$ are spacelike separated, whenever $n \neq m$. The sectional view (Fig. 5) from above onto $t = 0$ is a circular area now. The tubes are arranged so that their causal shadows (in $t = 0$) are disjoint. Now we see that $p_n \to p$, $q_n \to p$ (in $\overline{\mathscr{D}}$ topology) and that the other properties remain unchanged.

Since our counterexample is a little bit artificial, one could consider exclusively space-times in which i is well defined. Therefore we examine the most important of Meyer's assertions about i. He claims that i would be continuous. He is wrong even when i is well defined (Fig. 6). Here $i(p) = i(\lim p_n)$ is in A^{ext}, because every subset S of M with $I'^-(p) \cap M = I^-(S)$ satisfies $S \subset \mathrm{cl}_{\mathscr{M}}(J'^-(p)M)$, but

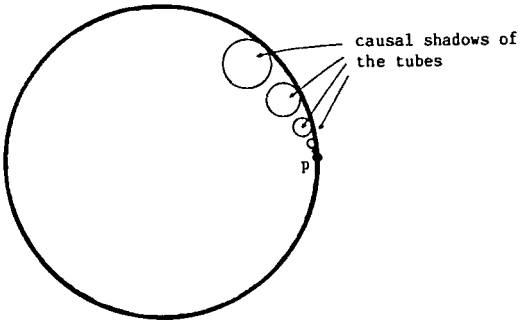


FIG. 5. Sectional view onto the plane $t = 0$ in Fig. 4.

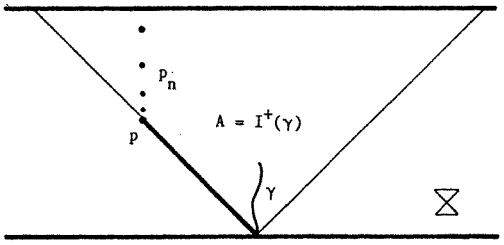


FIG. 6. Here i is not continuous. (Bold lines do not belong to M.)

$\mathrm{cl}_{/\!/}(J'^{-}(p)\cap M)\cap \mathrm{cl}_{/\!/}(A)=\varnothing$, so $I^{+}(S)\subset A$ is impossible. Clearly, $i(p_n)\notin A^{\mathrm{ext}}$ for all n, and i is not continuous. Meyer states: "The part of the causal boundary which is generated by $\overline{\mathscr{D}}$-uniformly continuous curves is homeomorphic to the part of the D boundary which has either past or future." He is wrong, even if D is continuous (Meyer uses this property in his proof) (Fig. 7). In Fig. 7, our metric is indicated by some null geodesics. We just have a part of Minkowski space, where $t<0$. At $x\to 0$ $(t>0)$, the metric degenerates so that in the limit all "causal" vectors consist of a single line and $g(\partial_t,\partial_t)=-1$ holds everywhere on M. The whole lightlike part of the c boundary does not belong to the D boundary, as is easily seen. So there cannot be a bijection between them.

III. D BOUNDARY COMPARED WITH PHYSICS

At first glance, the finite timelike diameter seems to be a serious obstruction to physics. Fortunately, by composing D with a homeomorphism h: $\mathbb{R}\to(a,b)$ $(a,b$ finite), one can save all correct theorems presented here or in Meyer's paper[1] (in Proposition 3.4 therein, one has to compose f with h, too). Of course, the $h\circ D$ topology does not depend on the particular homeomorphism h we have chosen, and the modified D boundary also works in space-times with infinite timelike diameter.

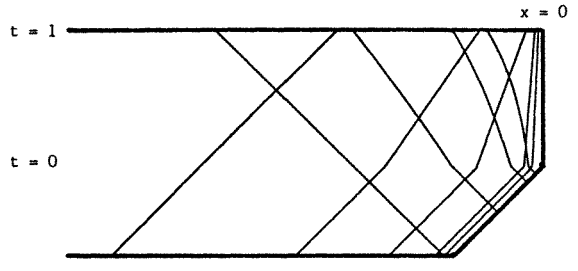


FIG. 7. The D boundary is not homeomorphic to the causal boundary.

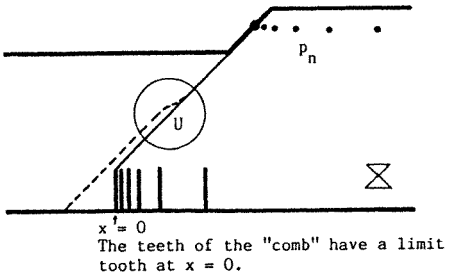


FIG. 8. The D boundary is not stable under small metric perturbations. (Bold lines do not belong to M.)

Although we could overcome this drawback, the D boundary is ruled out by physics (Fig. 8). We see that none of the intuitive lightlike boundary points is a D-boundary point, because to every p_n there exists a p_m $(m>n)$ and a timelike curve from p_m to the space between two teeth of the "comb" which cannot be reached from p_n. But by an arbitrary small perturbation of the metric in the arbitrary small set U we can widen the lightcone in U and thereby achieve that every p_n sufficiently close to p can reach the tooth at $x=0$. Consequently, $\{p_n\}$ becomes a Cauchy sequence and p a D-boundary point. Since general relativity is a classical theory which describes nature only approximately, every physically meaningful object must be stable under small metric perturbations. This implies that the D boundary can only serve as a technical tool.

ACKNOWLEDGMENTS

I want to thank Bernd Wegner for calling my attention to boundary constructions and for suggesting to examine Meyer's construction. I also had a very instructive discussion with Petra Rübe.

[1] D. A. Meyer, J. Math. Phys. **27**, 124 (1986).
[2] J. K. Beem and P. E. Ehrlich, *Global Lorentzian Geometry* (Dekker, New York, 1981), Corollary 2.19.
[3] R. P. Geroch, E. H. Kronheimer, and R. Penrose, Proc. R. Soc. London Ser. A **327**, 545 (1972).

Characteristic classes of super vector bundles

Ugo Bruzzo
Dipartimento di Matematica, Università di Genova, Via L. B. Alberti 4, 16132 Genova, Italy

Daniel Hernández Ruipérez
Departamento de Matemáticas, Universidad de Salamanca, Plaza de la Merced 1-4, 37008 Salamanca, Spain

(Received 7 September 1988; accepted for publication 4 January 1989)

A theory of characteristic classes for complex super vector bundles over supermanifolds is developed. Such characteristic classes are proved to fulfill the usual properties, and it is shown that, under suitable conditions, they can be represented in terms of supersmooth curvature forms.

I. INTRODUCTION

This paper is devoted to the development of a theory of characteristic classes for (complex) super vector bundles (SVB's) by generalizing the construction of the Chern classes of smooth complex vector bundles in terms of the cohomology of the projectivization of the bundle. Such a theory should be relevant to physics, since many techniques exploited in supersymmetric gauge theories or in superstring theory, mainly in connection with the anomaly problem, actually involve nontrivial super vector bundles over supermanifolds and the study of their cohomology ring.[1]

We consider supermanifolds in the framework of the theory initiated by De Witt and Rogers,[2–4] which was given a satisfactory setting by Rothstein.[5] Even though we shall use a particular category of supermanifolds (which were called $\mathcal{G}$ supermanifolds), it should be noticed that the results in the present paper could be easily extended to the more general setting described by Rothstein, provided that the graded-commutative Banach algebra B that enters the theory is finite dimensional and there is a surjective algebra morphism $B \to F$, where F is the ground field.

The resulting category of supermanifolds contains objects that are topologically richer than Berezin and Kostant's graded manifolds[6,7]; this of course reflects on the cohomology of supermanifolds and on the properties of SVB's on them.

The basic algebraic object in supermanifold theory is a real Grassmann algebra $B_L = (B_L)_0 \oplus (B_L)_1$ with L generators. (In dealing with complex SVB's we shall also use the complexification of $B_L, C_L = B_L \oplus_{\mathbb{R}} \mathbb{C}$.) The Cartesian product B_L^{m+n} is graded (by "graded" we always mean "$\mathbb{Z}_2$ graded") according to the rule

$$B_L^{m+n} = B_L^{m,n} \oplus B_L^{\bar{m},\bar{n}} \equiv \left[(B_L)_0^m \times (B_L)_1^n\right] \oplus \left[(B_L)_1^m \times (B_L)_0^n\right].$$

Since there is a direct sum splitting $B_L = \mathbb{R} \oplus N_L$, where N_L is the nilpotent ideal of B_L, maps $\sigma: B_L \to \mathbb{R}$ (body map) and $s: B_L \to N_L$ are defined.

The first step in introducing supermanifolds is to consider a distinguished class of functions $f: U \subset B_L^{m,n} \to B_L$; the most suitable choice seems to be given by the so-called GH^∞ functions introduced by Rogers.[4] One gets a sheaf $\mathcal{G}\mathcal{H}$ of graded commutative B_L algebras on $B_L^{m,n}$, where L' is a positive integer such that $L - L' \geqslant n$. The most natural definition of supermanifold now would seem to state that a (GH^∞) supermanifold is a pair $(M, \mathcal{A})$ where M is a topological space and $\mathcal{A}$ is a sheaf of graded commutative B_L algebras, such that $(M,\mathcal{A})$ and $(B_L^{m,n}, \mathcal{G}\mathcal{H})$ are locally isomorphic as ringed spaces. However, it was shown elsewhere[8] that the resulting category of supermanifolds is not suitable to develop a theory of SVB's that parallels the ordinary theory of vector bundles on real or complex manifolds, e.g., the graded tangent bundle to a GH^∞ supermanifold has no standard fiber. A solution to this problem is obtained by replacing $\mathcal{G}\mathcal{H}$ with the sheaf of graded commutative B_L algebras $\mathcal{G} = \mathcal{G}\mathcal{H} \oplus B_L \cdot B_L$, and defining $\mathcal{G}$ supermanifolds as pairs $(M,\mathcal{A})$ locally isomorphic with $(B_L^{m,n}, \mathcal{G})$.[8,9]

A particular class of supermanifolds is given by the so-called De Witt supermanifolds; they are characterized by the fact that they can be covered by means of $\mathcal{G}$ atlases $A = \{(U_i,\psi_i)\}$ such that the images $\psi_i(U_i)$ are of the type $\psi_i(U_i) \simeq V_i \times N_L^{m,n}$, where the V_i are open sets in $\mathbb{R}^m$. It can be shown that an (m,n) De Witt supermanifold M is a fiber bundle over a smooth m manifold M_0 (called the body of M) with standard fiber $N_L^{m,n}$. Therefore, M and M_0 have the same integer cohomology. Moreover, it was proved in Ref. 10 that the structure sheaf $\mathcal{A}$ of a De Witt supermanifold is acyclic, which is not the case for a general supermanifold.

The nonacyclicity of the structure sheaf of a generic supermanifold M has the consequence that the de Rham-type cohomology of the complex of supersmooth forms on M is quite distinct from the ordinary de Rham cohomology of M, and that a Čech–de Rham type isomorphism fails to exist. Indeed, introducing the sheaves Ω^k of k superforms on M by letting

$$\Omega^0 \equiv \mathcal{A}, \quad \Omega^1 \equiv \mathrm{Der}^*\,\mathcal{A}, \quad \Omega^k = \Lambda^k \Omega^1, \quad \text{for } k > 1,$$

where Λ is the graded–antisymmetrized graded tensor product over $\mathcal{A}$, and defining a sheaf morphism (Cartan's exterior differential) $d: \Omega^k \to \Omega^{k+1}$ in the usual way, one gets the following resolution of the locally constant sheaf B_L over M:

$$0 \to B_L \to \Omega^*. \tag{1.1}$$

The exactness of (1.1) is a consequence of a Poincaré lemma for $\mathcal{G}$ superforms, which was proved elsewhere.[11] The cohomology of the complex of graded B_L modules $(\Gamma\Omega^*, d)$ ("supersmooth de Rham cohomology") will be denoted by $H_{\mathrm{SDR}}(M)$ (Γ denotes the global section functor). Standard

 0022-2488/89/061233-05$02.50

cohomological arguments[12] entail the existence of a morphism

$$H_{\mathrm{SDR}}(M) \to H(M,B_L) \simeq H_{\mathrm{DR}}(M) \oplus B_L, \tag{1.2}$$

which in general is neither injective nor subjective. However, if M is De Witt, so that the sheaves Ω^k are acyclic, the morphism (1.2) is bijective.[13]

Within the category of $\mathscr{G}$ supermanifolds it is possible to develop a satisfactory theory of SVB's; the category of rank (r,s) SVB's on a fixed supermanifold $(M,\mathscr{A})$ turns out to be equivalent to the category of rank (r,s) locally free graded $\mathscr{A}$ modules. It is the purpose of the present paper to develop a theory of characteristic classes for such bundles in the complex case. We shall consider characteristic classes with integer coefficients, and we shall study their representation in terms of the curvature form of a connection on the bundle. More precisely, this paper is organized as follows. In Sec. II we introduce projective superspaces and analyze their integer cohomology; moreover, we review the definition of SVB and study the cohomology of the projectivizations of an SVB. It turns out that it is convenient to introduce two projectivizations, an even and an odd one, and correspondingly in Sec. III we introduce even and odd Chern classes; a Whitney product formula is then proved. In Sec. IV we investigate whether the characteristic classes of an SVB E over M can be represented by means of cohomology classes in $H_{\mathrm{SDR}}(M)$ defined in terms of the curvature of a connection on E (Chern–Weil theorem); we are able to answer in the affirmative in the cases where E has rank (1,0) or (0,1), or when E has arbitrary rank, but M is De Witt.

II. PROJECTIVIZATIONS OF A SUPER VECTOR BUNDLE

In this section we deal with some preparatory material that we shall need in the following section to introduce Chern classes of super vector bundles. We start by recalling the definition of SVB.[8]

Definition 2.1: A rank (r,s) complex super vector bundle is a triple (E, M, p) where E and M are supermanifolds and $p: E \to M$ is a $\mathscr{G}$ map, such that (i) M has a cover $\{U_i\}$ with $\mathscr{G}$ diffeomorphisms

$$\psi_i: p^{-1}(U) \to U_i \times F \tag{2.1}$$

satisfying $\mathrm{pr}_2 \circ \psi_i = p$, where F is a rank (r,s) free graded C_L module; (ii) the transition functions defined by letting

$$\psi_i \circ \psi_j^{-1}(x,u) = (x, g_{ij}(x)u)$$

are morphisms of graded C_L modules, i.e., they are $\mathscr{G}$ maps $g_{ij}: U_i \cap U_j \to \mathrm{GL}(r,s)$, where $\mathrm{GL}(r,s)$ is the super Lie group formed by the even automorphisms of the graded C_L module C_L^{r+s}.[14,15] □

We recall that an $X \in \mathrm{GL}(r,s)$ is an invertible matrix showing the block structure

$$X = \begin{pmatrix} A & B \\ C & D \end{pmatrix},$$

where A and D have entries in $(C_L)_0$ and are $r \times r$ and $s \times s$, respectively, while B and C have entries in $(C_L)_1$ and are $r \times s$ and $s \times r$, respectively.

With this definition, the sheaf of germs of sections of a complex SVB E over a $\mathscr{G}$ supermanifold $(M,\mathscr{A})$ is a locally free graded $\mathscr{S}$ module, where $\mathscr{S} = \mathscr{A} \oplus_{\mathbb{R}} \mathbb{C}$ may be identified with the sheaf of germs of $\mathscr{G}$ maps $M \to C_L$,[9] and one indeed can show that the category of rank (r,s) complex SVB's on M and the category of locally free graded $\mathscr{S}$ modules are equivalent.[8]

Let us recall some further details in the case of complex super line bundles (CSLB's).[16] These are defined as complex SVB's of rank (1,0) [equivalently, one could consider the rank (0,1) case, since the two types of bundles have the same standard fiber and structure group]. Here the transition functions are $\mathscr{G}$ maps $g_{ij}: U_i \cap U_j \to (C_L)_0^*$, where $(C_L)_0^* \simeq \mathrm{GL}(1,0) \simeq \mathrm{GL}(0,1)$ is the group of invertible elements of $(C_L)_0$; thus the set of isomorphism classes of CSLB's over a given supermanifold M is in a one-to-one correspondence with $H^1(M,\mathscr{S}_0^*)$, where $\mathscr{S}_0^*$ is the invertible subsheaf of $\mathscr{S}_0$. There is an exact sequence

$$0 \to \mathbb{Z} \to \mathscr{S}_0 \to \mathscr{S}_0^* \to 0,$$

and the associated cohomology sequence contains the segment

$$H^1(M,\mathscr{S}_0) \to H^1(M,\mathscr{S}_0^*) \xrightarrow{\delta} H^2(M,\mathbb{Z}).$$

If $E \in H^1(M,\mathscr{S}_0^*)$, we define $\delta(E) \in H^2(M,\mathbb{Z})$ to be the *obstruction class* of E. Unless M is De Witt, δ is not necessarily injective, so that CSLB's are not classified by their obstruction class.

Projective superspaces. After fixing non-negative integers r,s,h,k with $h \leqslant r$ and $k \leqslant s$, we define $\mathrm{GL}(h,k;r,s)$ as the subgroup of $\mathrm{GL}(r,s)$ whose elements are matrices of the form

$$\begin{pmatrix} a & b & c & d \\ 0 & e & 0 & f \\ g & i & l & p \\ 0 & q & 0 & u \end{pmatrix},$$

where the blocks have the following dimensions, both horizontal and vertical: $h, r-h, k, s-k$. Here, $\mathrm{GL}(h,k;r,s)$ is a De Witt supermanifold with body $\mathrm{Gl}(h;r) \times \mathrm{Gl}(k;s)$, where $\mathrm{Gl}(h;r)$ is the subgroup of matrices in $\mathrm{Gl}(r;\mathbb{C})$ (ordinary Lie group) of the form

$$\begin{pmatrix} a & b \\ 0 & c \end{pmatrix}.$$

Hence it follows that the quotient

$$G_{h,k}(r,s) = \mathrm{GL}(r,s)/\mathrm{GL}(h,k;r,s)$$

is a De Witt supermanifold, of even dimension $h(r-h) + k(s-k)$, odd dimension $k(r-h) + h(s-k)$, and body $G_h(r) \times G_k(s)$, where $G_h(r)$ is the Grassmann manifold of h planes in $\mathbb{C}^r$. It is otherwise obvious that $G_{h,k}(r,s)$ parametrizes the rank (h,k) free graded sub-C_L-modules of C_L^{r+s}.

Now, let W be a rank (r,s) free graded C_L module, and define

$$P^{1,0}(W) = \text{space of rank (1,0) free graded sub-}C_L\text{-modules of } W,$$

$$P^{0,1}(W) = \text{space of rank (0,1) free graded sub-}C_L\text{-modules of } W.$$

From the previous discussion it follows that $P^{1,0}(W)$ and $P^{0,1}(W)$ are both De Witt supermanifolds, with dimensions $(r-1,s)$ and $(s-1,r)$, respectively, and have bodies isomorphic with the complex projective spaces $\mathbb{P}^{r-1}$ and $\mathbb{P}^{s-1}$. It follows that $P^{1,0}(W)$ [resp. $P^{0,1}(W)$] has the same integer cohomology as $\mathbb{P}^{r-1}$ (resp. $\mathbb{P}^{s-1}$).

Tautological exact sequences. On $P^{1,0}(W)$ we may define a tautological bundle S_0, which is the rank (1,0) subbundle of $P^{1,0}(W)\times W$ formed by the pairs (u,v) such that $v\in u$; analogously, one defines a rank (0,1) tautological bundle S_1 on $P^{0,1}(W)$, which is a subbundle of $P^{0,1}(W)\times W$. Now, let V be the body of W, i.e., the vector space $V=W\oplus_{C_L}\mathbb{C}$, where $\mathbb{C}$ is given a C_L-module structure by means of the body map σ: $C_L\to\mathbb{C}$; V is graded, $V=V_0\oplus V_1$. Denoting by $S_i^{\#}$, $i=0,1$, the tautological bundles of the projective spaces $P(V_i)$, the body of S_i (in the sense of De Witt supermanifolds) in just $S_i^{\#}$, whence one has commutative diagrams

$$\begin{array}{ccccccc} 0\to & S_i & \to P^{1-i,i}(W)\times W\to & Q_i & \to 0, & & \\ & \downarrow & \downarrow & \downarrow & & i=0,1, & (2.2)\\ 0\to & S_i^{\#}\to & P(V_i)\times V_i & \to Q_i^{\#}\to 0, & & & \end{array}$$

where Q_i and $Q_i^{\#}$ are by definition the quotient (super) bundles. The following theorem is a straightforward consequence of (2.2) and of classical results concerning the cohomology of projective bundles.[17]

Theorem 2.1: The integer cohomology of $P^{1,0}(W)$ is freely generated over $\mathbb{Z}$ by $\{1,x,x^2,\ldots,x^{r-1}\}$, where x is the obstruction class of S_0. Analogously, the integer cohomology of $P^{0,1}(W)$ is freely generated over $\mathbb{Z}$ by $\{1,t,t^2,\ldots,t^{s-1}\}$, where t is the obstruction class of S_1.

Projectivizations of super vector bundles. Let us define the super Lie group

$$\mathrm{PGL}(r,s)=\mathrm{GL}(r,s)/(C_L)_0^{*}I$$

together with the canonical projection λ: $\mathrm{GL}(r,s)\to\mathrm{PGL}(r,s)$; $\mathrm{PGL}(r,s)$ acts in a natural way on $P^{1,0}(W)$ and $P^{0,1}(W)$. Given an SVB $p:E\to M$, whose transition functions relative to a fixed cover are g_{ij}, we define its even and odd projectivizations as follows: $P^{1,0}(E)$ [resp. $P^{0,1}(E)$] is the bundle on M whose standard fiber over $x\in M$ is $P^{1,0}(E_x)$ [resp. $P^{0,1}(E_x)$] and whose transition functions are the maps $\lambda\circ g_{ij}$. We shall denote by π_i: $P^{1-i,i}(E)\to M$, $i=0,1$, the bundle projections. The operation of taking the projectivizations is functorial, in the sense that if f: $M\to N$ is a morphism of $\mathscr{G}$ supermanifolds, and E is an SVB over N, there are $\mathscr{G}$ maps F_i: $P^{1-i,i}(f^{-1}E)\to P^{1-i,i}(E)$ such that the following diagram commutes:

$$\begin{array}{ccc} P^{1-i,i}(f^{-1}E), & \xrightarrow{F_i} & P^{1-i,i}(E), \\ \downarrow & & \downarrow \\ M & \xrightarrow{f} & N, \end{array} \qquad i=0,1.$$

$P^{1,0}(E)$ and $P^{0,1}(E)$ carry tautological bundles defined in the obvious way; $S_0(E)\to P^{1,0}(E)$ has rank (1,0), while $S_1(E)\to P^{0,1}(E)$ has rank (0,1). There are two tautological exact sequences,

$$0\to S_i(E)\to \pi_i^{-1}E\to Q_i(E)\to 0, \quad i=0,1.$$

The assignment of the tautological bundles is functorial as well, i.e., there are commutative diagrams

$$\begin{array}{ccccccc} 0\to & S_i(f^{-1}E) & \to \pi_i^{-1}(f^{-1}E)\to & Q_i(f^{-1}E) & \to 0, & & \\ & \downarrow & \downarrow & \downarrow & & i=0,1. & \\ 0\to & S_i(E) & \to \pi_i^{-1}E\to & Q_i(E) & \to 0, & & \end{array}$$

In order to get information about the integer cohomology of the projectivizations of E, we must use the Leray–Hirsch theorem.[18] We need it in the following weaker form than the one given in Ref. 18: if p: $Q\to M$ is a locally trivial topological bundle, with standard fiber F, K is a principal ring, and there are cohomology classes $\{a_1\cdots a_q\}$ that when restricted to the fibers generate freely over K the cohomology of the fibers with coefficients in K, then $H(Q,K)$ is a free $H(M,K)$ module generated by $\{a_1\cdots a_q\}$. If we consider the bundles $P^{1-i,i}(E)$ over M, the hypotheses of the Leray–Hirsch theorem are fulfilled as a consequence of Theorem 2.1, so that we have the following theorem.

Theorem 2.2: The following isomorphisms of $\mathbb{Z}$ modules hold:

$$H(P^{1-i,i}(E),\mathbb{Z}) \simeq H(M,\mathbb{Z})\otimes_{\mathbb{Z}} H(P^{1-i,i}(C_L^{m+n}),\mathbb{Z}), \quad i=0,1.$$

III. CHARACTERISTIC CLASSES OF SUPER VECTOR BUNDLES

Given a rank (r,s) SVB p: $E\to M$, we can straightforwardly introduce its even and odd Chern classes as follows: if x and t are, respectively, the obstruction classes of the even and odd tautological bundles of the projectivizations of E, we let (with reference to Theorem 2.2)

$$x^r=-\sum_{j=1}^{r}C_j^0(E)x^{r-j}, \quad t^s=-\sum_{k=1}^{s}C_k^1(E)t^{s-k}, \tag{3.1}$$

so that $C_j^0(E)$ and $C_k^1(E)$ are well determined elements in $H^{2j}(M,\mathbb{Z})$ and $H^{2k}(M,\mathbb{Z})$, respectively. Correspondingly there are two total Chern classes,

$$C^0(E)=\sum_{j=0}^{r}C_j^0(E), \quad C^1(E)=\sum_{k=0}^{s}C_k^1(E). \tag{3.2}$$

According to this definition, a rank (r,s) SVB has r even and s odd Chern classes. The normalization and functoriality properties of these classes are readily proved.

Theorem 3.1: If E has rank (1,0), then

$$C^0(E)=1-\delta(E); \tag{3.3}$$

if E has rank (0,1), then

$$C^1(E)=1-\delta(E). \tag{3.4}$$

Proof: If rank $E=(1,0)$, then E has only an even projectivization; moreover, $S_0(E)\simeq E$, so that (3.3) follows. A similar argument applies to the rank (0,1) case. □

Theorem 3.2: If f: $M\to N$ is a morphism of $\mathscr{G}$ supermanifolds, and E is an SVB over N, then

$$C^i(f^{-1}E)=f^{*}C^i(E), \quad i=0,1.$$

Proof: This property follows from the functoriality of the projectivized and tautological bundles. □

In order to prove a Whitney product formula, we need some further constructions. Recalling that $\mathscr{S}$ is the sheaf of germs of $\mathscr{G}$ maps $M \to C_L$, and denoting by $\mathscr{C}$ the sheaf of germs of smooth complex-valued functions on M, we can define a morphism γ: $\mathscr{S} \to \mathscr{C}$ by letting $\gamma(f) = \sigma \circ f$ for all $f \in \mathscr{S}(U)$. By means of γ, the sheaf $\mathscr{C}$ can be given an $\mathscr{S}$-module structure. Given a rank (r,s) SVB E over M, we let

$$E^{\#} = E \otimes_{\mathscr{S}} C. \tag{3.5}$$

$E^{\#}$ is a rank $r+s$ graded smooth complex vector bundle on M, which splits canonically into $E^{\#} = E_0^{\#} \oplus E_1^{\#}$, where the two summands have rank r and s, respectively. The bundles $E_i^{\#}$ can be described in terms of their transition functions as follows. If $\{g_{ij}\}$ is a set of transition functions for E, then the matrices $\{\sigma \circ g_{ij}\}$ have the block structure

$$\sigma \circ g_{ij} = \begin{pmatrix} h_{ij} & 0 \\ 0 & k_{ij} \end{pmatrix} \in \mathrm{Gl}(r;\mathbb{C}) \times \mathrm{Gl}(s;\mathbb{C}),$$

and the sets $\{h_{ij}\}$ and $\{k_{ij}\}$ provide transition functions for $E_0^{\#}$ and $E_1^{\#}$, respectively.

The definition (3.5) entails the existence of vector bundle maps $E \to E_i^{\#}$; these can be lifted to maps between the projectivized bundles $P^{1-i,i}(E) \to P(E_i^{\#})$ and between the tautological bundles, so that one obtains commutative diagrams

$$\begin{array}{ccccccc} 0 \to & S_i(E) & \to & \pi_i^{-1}E & \to & Q_i(E) & \to 0, \\ & \downarrow & & \downarrow & & \downarrow & \qquad i = 0,1, \\ 0 \to & S(E_i^{\#}) & \to & \pi_i^{\#\,-1}E & \to & Q(E_i^{\#}) & \to 0, \end{array}$$

where $\pi_i^{\#}$ is the bundle projection $E_i^{\#} \to M$. The commutativity of these diagrams implies that, for fixed i, $S_i(E)$ and $S(E_i^{\#})$ have the same obstruction class. This in turn implies the following lemma.

Lemma 3.1: $C^0(E) = c(E_0^{\#})$, $C^1(E) = c(E_1^{\#})$. □

It is now possible to prove Whitney's formula.

Theorem 3.3: If $0 \to E \to F \to G \to 0$ is an exact sequence of SVB's, then

$$C^i(F) = C^i(E)C^i(G), \quad i = 0,1, \tag{3.6}$$

where the product in the right-hand side is the cup product in $H(M,\mathbb{Z})$.

Proof: Since $0 \to E \to F \to G \to 0$ is an exact sequence of locally free modules, by tensoring it with $\mathscr{C}$ one gets an exact sequence of smooth vector bundles over M,

$$0 \to E^{\#} \to F^{\#} \to G^{\#} \to 0,$$

which splits, thus giving isomorphisms $F_i^{\#} \simeq E_i^{\#} \oplus G_i^{\#}$. The ordinary Whitney formula then yields $c(F_i^{\#}) = c(E_i^{\#})c(G_i^{\#})$, which, together with Lemma 3.1, implies the thesis. □

It should be noticed that we have stated the Whitney product formula in terms of exact sequences of SVB's rather than in terms of direct sums of SVB's, since, due to the nonacyclicity of the structure sheaf of the base supermanifolds, not all exact sequences of SVB's split (see Sec. IV).

We conclude this section by introducing the Chern character of an SVB; as we shall see in Sec. IV, the representation of the characteristic classes of an SVB in terms of curvatures is most simply exhibited by means of the Chern character. For a given rank (r,s) SVB E over M, by means of the formal factorizations[19]

$$\sum_{j=0}^{r} C_j^0(E)x^j = \prod_{j=1}^{r} (1 + \gamma_j x),$$

$$\sum_{k=0}^{s} C_k^1(E)t^k = \prod_{k=1}^{s} (1 + \delta_k t),$$

we define the even and odd Chern characters of E

$$\mathrm{Ch}^0(E) = \sum_{j=1}^{r} e^{\gamma_j}, \quad \mathrm{Ch}^1(E) = \sum_{k=1}^{s} e^{\delta_k},$$

and the total Chern character

$$\mathrm{Ch}(E) = \mathrm{Ch}^0(E) - \mathrm{Ch}^1(E).$$

Of course $\mathrm{Ch}(E) \in H(M,\mathbb{Z})$, and there is a decomposition

$$\mathrm{Ch}(E) = \sum_{i=0}^{\infty} \mathrm{Ch}_i(E), \quad \mathrm{Ch}_i(E) \in H^{2i}(M,\mathbb{Z});$$

in particular, one has $\mathrm{Ch}_0(E) = r - s$ (we assume that M is connected).

The analog of the Whitney product formula for Chern characters reads as follows: if $0 \to E \to F \to G \to 0$ is an exact sequence of SVB's, then

$$\mathrm{Ch}^i(F) = \mathrm{Ch}^i(E) + \mathrm{Ch}^i(G), \quad i = 0,1. \tag{3.7}$$

IV. REPRESENTATION OF CHARACTERISTIC CLASSES IN TERMS OF CURVATURE FORMS

Let E be a complex SVB of rank (r,s) on a supermanifold $(M,\mathscr{A})$; a connection Δ on E is an even morphism of sheaves of graded C_L-modules

$$\Delta\colon E \to \mathrm{Hom}(TM,E) \equiv \Omega^1 \otimes_{\mathscr{S}} E,$$

satisfying (recall that $\mathscr{S} = \mathscr{A} \otimes_{\mathbb{R}} \mathbb{C}$)

$$\Delta(f\xi) = f\Delta(\xi) + df \otimes \xi, \quad \forall f \in \mathscr{S}(U), \quad \xi \in E(U),$$

$$\text{and } \forall \text{ open } U \subset M.$$

In contrast with smooth bundles, and in analogy with holomorphic bundles, an SVB does not always carry a connection (a more detailed discussion of this point is to be found in Ref. 20). This is due to the nonacyclicity of the structure sheaf of a generic supermanifold; indeed, in the case of a De Witt base supermanifold, connections always exist.

Let

$$0 \to E \to F \to G \to 0 \tag{4.1}$$

be an exact sequence of complex SVB's over a supermanifold $(M,\mathscr{A})$; general arguments[21] show that the sequence (4.1) is split if and only if a certain cohomology class in $H^1(M,\mathrm{Hom}(G,E))$ vanishes. Since $\mathrm{Hom}(G,E)$ is a locally free graded $\mathscr{S}$ module, it is acyclic if M is De Witt,[20] so that we have that *all exact sequences of SVB's on a De Witt supermanifold split*. This implies that *an SVB over a De Witt supermanifold always admits connections*; indeed it can be shown that the connections on E are in a one-to-one correspondence with the splittings of the exact sequence of SVB's over M

$$0 \to \Omega^1 \otimes E \to D(E) \to E \to 0,$$

where $D(E)$ is $E \oplus (\Omega^1 \otimes E)$ endowed with the structure of graded $\mathscr{S}$ module given by

$$f(\xi \oplus a) = f\xi \oplus (fa + df \otimes \xi), \forall f \in \mathscr{S}(U),$$

$$\xi \in E(U), \quad a \in (\Omega^1 \otimes E)(U),$$

$$\text{and } \forall \text{ open } U \subset M.$$

The representation theorem has been proved elsewhere in the case of complex super line bundles. Let E be a CSLB over a (generic) supermanifold M, and assume that E admits a connection Δ, with curvature form Ω. The Bianchi identity states that Ω is closed, and one can prove[22] that the cohomology class $[\Omega] \in H^2_{\mathrm{SDR}}(M) \otimes_{\mathbb{R}} \mathbb{C}$ is independent of the choice of the connection. Let $j: H(M,\mathbb{Z}) \to H(M,C_L)$ be the morphism induced by the inclusion of sheaves $\mathbb{Z} \to C_L$, and let $\alpha: H_{\mathrm{SDR}}(M) \otimes_{\mathbb{R}} \mathbb{C} \to H(M,C_L)$ be the morphism induced by (1.2); then we have[16]

$$j(C_1^0(E)) = (i/2\pi)\alpha([\Omega]). \tag{4.2}$$

Equation (4.2) refers to the rank (1,0) case; obviously, in the rank (0,1) case we get $j(C_1^1(E)) = (i/2\pi)\alpha([\Omega])$. In terms of the Chern character, in both cases one has

$$\mathrm{Ch}_1(E) = (i/2\pi)[\mathrm{Str}\,\Omega], \tag{4.3}$$

where Str denotes the supertrace of a matrix in $\mathrm{GL}(r,s)$.[14] Equation (4.3) follows from the fact that $\mathrm{Str}\,\Omega = \Omega$ if rank $E = (1,0)$, while $\mathrm{Str}\,\Omega = -\Omega$ if rank $E = (0,1)$.

We are able to generalize this result to the case of a rank (r,s) SVB, provided that the base supermanifold is De Witt.

Theorem 4.1: Let E be a rank (r,s) SVB over an (m,n) dimensional De Witt supermanifold M. For any connection Δ on E with curvature Ω, one has

$$j(\mathrm{Ch}_k(E)) = (i/2\pi)^k[\mathrm{Str}\,\Omega^k], \quad 1 \leqslant k \leqslant m/2. \tag{4.4}$$

For $k > m/2$, both sides of (4.3) vanish identically.

Before proving this theorem, we need the following result, which holds also when the base supermanifold is not De Witt.

Lemma 4.1: Let $p: E \to M$ be a rank (r,s) complex SVB. There is a $\mathcal{G}$ supermanifold morphism $f: N \to M$ such that: (i) $f^*: H(M,\mathbb{Z}) \to H(N,\mathbb{Z})$ is a monomorphism; (ii) there is a chain of morphisms of complex SVB's over N, $g_j: F_{j-1} \to F_j$, with $j = 1 \cdots r+s$, $F_{r+s} = f^{-1}E$, and $F_0 = M \times \{0\}$, such that any quotient superbundle F_j/F_{j-1} has either rank (1,0) or (0,1).

Proof: This lemma is proved by double induction on the rank of E. If rank $E = (1,0)$ or $(0,1)$ the result is trivial. Suppose now that rank $E = (r+1,s)$ and consider the even projectivization of E, $\pi_0: P^{1,0}(E) \to M$; the cohomology map $\pi_0^*: H(M,\mathbb{Z}) \to H(P^{1,0}(E),\mathbb{Z})$ is injective by Leray–Hirsch. The pullback bundle $\pi_0^{-1}E \to P^{1,0}(E)$ has a tautological super line subbundle $S_0(E) \to P^{1,0}(E)$, and the quotient superbundle $Q_0(E)$ has rank (r,s). By the induction hypothesis, there is a $\mathcal{G}$ map $g: N \to P^{1,0}(E)$ satisfying the properties in the statement of this lemma. Then the composition $f = \pi_0 \circ g: N \to M$ yields the required map. The induction on the odd rank is proved in the same way. □

Finally, we may prove Theorem 4.1. As a consequence of Lemma 4.1, and using the fact that on a De Witt supermanifold all exact sequences of SVB's split, there is a $\mathcal{G}$ morphism $f: N \to M$ such that

$$f^{-1}E \simeq L_1 \oplus \cdots \oplus L_r \oplus K_1 \oplus \cdots \oplus K_s, \tag{4.5}$$

where the L's have rank (1,0) and the K's have rank (0,1). Using a "tubular partition of unity" over M (Ref. 20) it is possible to define a connection Δ on E which when pulled back to $f^{-1}E$ "splits" in accordance with (4.5), i.e., it defines connections Δ_j on L_j and Θ_h on K_h, $j = 1 \cdots r$, $h = 1 \cdots s$. Then we have [recalling that $\mathrm{Str}(\Omega^\Delta)^k$, where Ω^Δ is the curvature of Δ, can be regarded as a superform on M]

$$f^* \mathrm{Str}(\Omega^\Delta)^k = \sum_{j=1}^{r} (\Omega^{\Delta_j})^k - \sum_{h=1}^{s} (\Omega^{\Theta_h})^k. \tag{4.6}$$

Now Eqs. (3.7), (4.4), and (4.6) yield

$$f^* \circ j(\mathrm{Ch}_k(E)) = (i/2\pi)^k f^* \circ \alpha([\mathrm{Str}\,\Omega^k]).$$

Since f^* is injective, this implies Eq. (4.4).

ACKNOWLEDGMENTS

We would like to thank C. Bartocci and J. Muñoz Masqué for useful discussions and valuable suggestions.

This work was done within the joint Italian–Spanish CNR–CSIC research project "Methods and applications of differential geometry in mathematical physics." It was also supported by "Gruppo Nazionale per la Fisica Matematica" of CNR, by the Italian Ministry for Public Education through the research project "Geometria e Fisica," and by the Spanish CICYT through the research project "Geometría de las teorías gauge."

[1] L. Bonora, P. Pasti, and M. Tonin, Phys. Lett. B **156**, 341 (1985); Nucl. Phys. B **261**, 241 (1985); Phys. Lett. B **177**, 191 (1986); Nucl. Phys. B **286** 150 (1987); and in *Field and Geometry*, edited by A. Jadczyk (World Scientific, Singapore, 1987); S. Buckingham, "Weil triviality and anomalies in two dimensional supergravity," King's College Preprint, London, 1987; U. Bruzzo and G. Landi, Phys. Rev. D **39**, 1174 (1989); C. Bartocci, U. Bruzzo, and G. Landi, "Geometry of standard constraints and Weil triviality in supersymmetric gauge theories," to appear.
[2] B. De Witt, *Supermanifolds* (Cambridge U.P., London, 1984).
[3] A. Rogers, J. Math. Phys. **21**, 1352 (1980).
[4] A. Rogers, Commun. Math. Phys. **105**, 375 (1986).
[5] M. Rothstein, Trans. Am. Math. Soc. **297**, 159 (1986).
[6] F. A. Berezin and D. A. Lĕites, Sov. Math. Dokl. **16**, 1218 (1975).
[7] B. Kostant, in *Differential Geometric Methods in Mathematical Physics, Lecture Notes in Mathematics*, Vol. 570 (Springer, Berlin, 1977), p. 177.
[8] C. Bartocci and U. Bruzzo, J. Geom. Phys. **4**, 391 (1987).
[9] U. Bruzzo, C. Bartocci, and D. Hernández Ruipérez, "A remark on a new category of supermanifolds," to appear.
[10] C. Bartocci and U. Bruzzo, J. Math. Phys. **29**, 1789 (1988).
[11] U. Bruzzo, "Supermanifolds, supermanifold cohomology, and super vector bundles," in *Differential Geometric Methods in Theoretical Physics* (Como, 1987), edited by K. Bleuler and M. Werner (Kenwer, Dordrecht, 1988).
[12] R. Godement, *Théorie des Faisceaux* (Hermann, Paris, 1964).
[13] C. Bartocci and U. Bruzzo, J. Math. Phys. **28**, 2363 (1987).
[14] V. Rittenberg and M. Scheunert, J. Math Phys. **19**, 713 (1978).
[15] A. Rogers, J. Math. Phys. **22**, 939 (1981).
[16] C. Bartocci and U. Bruzzo, Lett. Math. Phys. **17**, 61 (1989).
[17] J. W. Milnor and J. D. Stasheff, *Characteristic Classes* (Princeton U.P., Princeton, NJ, 1974).
[18] D. Husemoller, *Fibre Bundles* (McGraw-Hill, New York, 1966).
[19] F. Hirzebruch, *Topological Methods in Algebraic Geometry* (Springer, Berlin, 1966).
[20] C. Bartocci and U. Bruzzo, to be published in Lett. Math. Phys.
[21] M. F. Atiyah, Trans. Am. Math. Soc. **85**, 181 (1957).
[22] C. Bartocci, U. Bruzzo, and G. Landi, "Chern–Simons forms on principal super fiber bundles." Preprint S.I.S.S.A. 109/87/FM, Trieste.

The Birkhoff–Gustavson normal form of one-dimensional double-well Hamiltonians

M. K. Ali and W. Robert Wood[a)]
Department of Physics, The University of Lethbridge, Lethbridge, Alberta T1K 3M4, Canada

(Received 12 September 1988; accepted for publication 8 February 1989)

The Birkhoff–Gustavson normal form (BGNF) is employed to study general double-well anharmonic oscillators with polynomial potentials of degree 4. Via an analytic continuation of the BGNF series, expressions are presented that provide classical and semiclassical results for all energies. A surprising observation of the classical period and an interesting feature of the semiclassical quantum numbers are reported. It is shown that except for a small region near the top of the potential barrier, the BGNF approach yields accurate quantum energies.

I. INTRODUCTION

Over the past few decades, one-dimensional anharmonic oscillators have been studied extensively. The work on the quartic oscillator has been particularly comprehensive; it has been studied as a mathematical and physical model in both classical and quantum theory. In classical theory, the solution to Newton's equation of motion has been given[1] in terms of Jacobi elliptic functions and the period and action have been given[2] in terms of generalized hypergeometric functions. In quantum theory, the literature on the quartic oscillator is very extensive.[3] Bender and Wu[4] showed that the Rayleigh–Schrödinger (RS) perturbation series diverges for all values of the coupling parameter: This result prompted resummation studies of the series, including the Borel[5] and Padé approximant[6] methods. In more recent years, the double-well system obtained by adding a cubic term to the quartic oscillator (cubic-quartic oscillator) has received considerable attention.[7] Many of the techniques employed in the study of the quartic oscillator (e.g., the use of elliptic functions and Borel summation) have been extended to the cubic-quartic oscillator. The classical period in the most general case, where the anharmonic term is any polynomial of order $n>4$, is represented in terms of a "hyperelliptic" integral which has just recently been shown[8] to be expressable in terms of a sum of multiple hypergeometric series. The enormity of work on one-dimensional anharmonic oscillators is the result of the applicability of this model in such fields of study as molecular dynamics and quantum field theory.[9]

In this paper, we study the cubic-quartic oscillator using the Birkhoff–Gustavson normal form (BGNF) approach. The double-well nature of this oscillator makes it an attractive model to study. The BGNF approach is useful because it provides the classical perturbation series, which is equivalent to the quantum RS perturbation series when the quantum operators are replaced by their corresponding classical functions. Since this classical series is a power series in the action, semiclassical results may easily be compared with exact quantum results once the series is properly summed. In Sec. II we define the Hamiltonian under consideration in two coordinate systems to facilitate the determination of the BGNF expansion about the equilibrium points of both the left and right wells. In Sec. II we also discuss some interesting properties of the action and period of the oscillator. In Sec. III we outline the algorithm used to perform the Lie transformations in the BGNF approach before developing the series. The analytic continuation of this series is also derived by identifying the inversion of the series with generalized hypergeometric functions. Padé approximants are presented in Sec. IV for inner and analytically continued series: Numeric classical results from these approximants are compared with results from truncated series, as well as with exact results. In Sec. IV we compare the semiclassical and quantum energies of the asymmetric double-well (ASDW) oscillator. A summary is given in Sec. V.

II. THE HAMILTONIAN

In general, the Hamiltonian under consideration defines an ASDW potential. Since either the right (R) or left (L) wells may be expanded to obtain the BGNF series, we define the Hamiltonian as

$$H_\Omega(x,p) = H_0 + H_{1\Omega} + \tfrac{1}{2} H_{2\Omega}, \tag{2.1}$$

where

$$H_0 = \tfrac{1}{2}(p^2+x^2),\ H_{1\Omega} = b_\Omega x^3,\ H_{2\Omega} = g_\Omega x^4,$$

with $\Omega = R$ or L. Here g_Ω is taken to be positive definite and b_Ω is taken to be arbitrary. The system defined by H_L is related to that defined by H_R through the coordinate shift

$$x \to x - (3b_R+\beta)/4g_R,\ \beta = [9b_R^2 - 8g_R]^{1/2},$$

along with the energy scaling

$$H_R = \sigma H_L + \Delta, \tag{2.2}$$

where

$$\sigma = [\alpha\beta/4g_R]^{1/2},\ \alpha = \beta + 3b_R,$$

$$\Delta = [\beta^2(g_R - \alpha b_R) + 3g_R b_R^2]/64g_R^3,$$

$$b_R = -(\beta + 2\sigma^{3/2} b_L),\ g_R = \sigma^2 g_L.$$

In the Hamiltonian (2.1) the choices $b_\Omega = 0$ and $b_\Omega = \pm\sqrt{g_\Omega}$ correspond to the single-well (SW) and symmetric double-well (SDW) cases, respectively. The BGNF

[a)] Present address: Department of Physics, University of Regina, Regina Saskatchewan, S4S 0A2 Canada.

 0022-2488/89/061238-07$02.50

approach has been applied to the SW and SDW oscillators in Refs. 10 and 11.

It is worthwhile to mention here that the period and action of the oscillator of Eq. (2.1) have some interesting properties. In the SDW case, the symmetry suggests that the period for a given energy in the left well (T_L) should be identically equal to the period in the right well (T_R), as indeed is true. One might think that for the asymmetric case, the period will depend on the well in which the classical particle executes its oscillatory motion. To our surprise, we observed that the oscillator has only one period for a given energy, i.e., $T_L = T_R$ in the asymmetric case as well! Let the energy E be measured from the bottom of the right well in Fig. 1 and let $V_R(x)$ represent the potential with the coordinate origin at the right well. The coordinates x_1, x_2, and x_3 of the critical points of the potential are given by

$$x_1 = 0, \quad x_2 = (-3b_R + \beta)/4g_R,$$
$$x_3 = (-3b_R - \beta)/4g_R.$$

In the "overlapping region" between the lines ASB and CWD in Fig. 1, the energy satisfies the condition $0 \leqslant E \leqslant V_R(x_2)$ and the particle can oscillate in either well. It is a simple matter to show that the period is the same irrespective of the well identification. Let d, c and b, a be the respective turning points in the left and right wells for any energy in the overlapping region (see Fig. 1). The periods T_L and T_R are given by

$$T_L = \sqrt{2}\int_d^c \frac{dx}{\sqrt{E - V_R(x)}}, \quad T_R = \sqrt{2}\int_b^a \frac{dx}{\sqrt{E - V_R(x)}}.$$

Using Eqs. (253.00) and (257.00) of Ref. 12, it can be shown that $T_L = T_R = \mu F(\pi/2|m)$, where

$$\mu = 4/\sqrt{g(a-c)(b-d)},$$
$$m = (a-b)(c-d)/(a-c)(b-d).$$

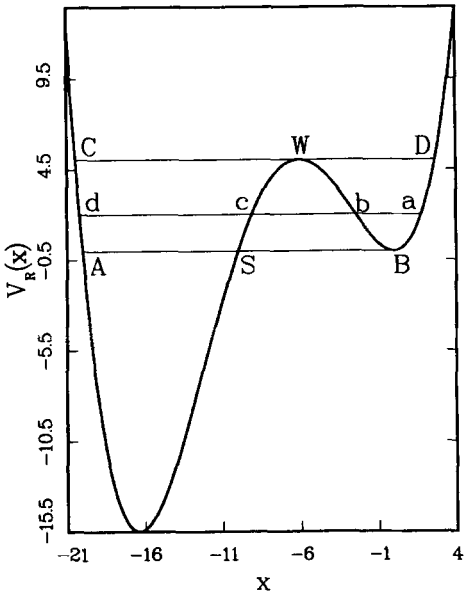


FIG. 1. The potential $V(x) = (\frac{1}{2})x^2 + bx^3 + (\frac{1}{2})gx^4$ as a function of x for $b = 0.075$ and $g = 0.005$, with the coordinate origin at the bottom of the right well.

The elliptic integral $F(\pi/2|m)$ can be expressed in terms of a hypergeometric function (see Abramowitz and Stegun[13]) as

$$F(\pi/2|m) = (\pi/2)\,{}_2F_1(\tfrac{1}{2},\tfrac{1}{2},1,m).$$

Since the action I and the period T are related by the expression $dI/dE = T/2\pi$, the difference in the actions for any two energies in the two wells is the same.

III. THE BGNF OF THE ASDW, SDW, AND SW OSCILLATORS

In this section we briefly outline the algorithm used in developing the BGNF series, referring the interested reader to Refs. 14–19 for a more detailed discussion. In the BGNF approach, a classical Hamiltonian function $H(\mathbf{x})$ where $\mathbf{x} = (x^i, p_i)$, $i = 1,...,n$, is canonically transformed to a normal ordered Hamiltonian function $K(\xi)$, $\xi = (\xi^i, \eta_i)$, $i = 1,...,n$. The transformation is implemented by expressing $\mathbf{x}$ as the following Taylor series in powers of an expansion parameter ϵ about $\epsilon = 0$, with $\mathbf{x} = \xi$ at $\epsilon = 0$:

$$\mathbf{x}(\epsilon) = \xi + \sum_{n=1}^{\infty} \frac{\epsilon^n}{n!}\left[\frac{d^n\mathbf{x}}{d\epsilon^n}\right]_{\epsilon=0}. \tag{3.1}$$

Following this "coordinate transformation," the Hamiltonian function is transformed by being expressed in the Taylor series form

$$H[\mathbf{x}(\epsilon)] = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!} H_n(\mathbf{x}),$$
$$H_n(\mathbf{x}) = \left[\frac{\partial^n H[\mathbf{x}(\epsilon)]}{\partial\epsilon^n}\right]_{\epsilon=0}, \tag{3.2}$$

rewriting the partial derivatives in $H_n(\mathbf{x})$ in terms of total derivatives evaluated at $\mathbf{x} = \xi$, and then writing these total derivatives recursively in terms of Lie derivatives or Poisson brackets. The generating function $w[\mathbf{x}(\epsilon)]$ for the Lie derivatives is defined through

$$\frac{dx^i}{d\epsilon} = \frac{\partial w[\mathbf{x}(\epsilon)]}{\partial p_i}, \tag{3.3a}$$

$$\frac{dp_i}{d\epsilon} = -\frac{\partial w[\mathbf{x}(\epsilon)]}{\partial x^i}, \quad i = 1,...,n \tag{3.3b}$$

and is assumed to have the series representation

$$w[\mathbf{x}(\epsilon)] = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!} w_{n+1}(\mathbf{x}),$$

where $w_{n+1}(\mathbf{x})$ are homogeneous polynomials of degree $n+3$. The transformed Hamiltonian assumes the form

$$K[\xi(\epsilon)] = \sum_{n=0}^{\infty} \frac{\epsilon^n}{n!} K_n(\xi). \tag{3.4}$$

The first three K_n's of Eq. (3.4) are given by

$$K_0 = H_0, \quad K_1 = H_1 + \{H_0, w_1\},$$
$$K_2 = H_2 + 2\{H_1, w_1\} + \{H_0, w_2\} + \{\{H_0, w_1\}, w_1\}.$$

Although $K(\xi)$ is the canonical transform of $H(\mathbf{x})$, it must also satisfy the additional condition

$$\{K_0(\xi), K_m(\xi)\} = \sum_{i=1}^{n}\left[\frac{\partial K_0}{\partial \xi_i}\frac{\partial K_m}{\partial \eta_i} - \frac{\partial K_0}{\partial \eta_i}\frac{\partial K_m}{\partial \xi_i}\right] = 0, \quad \forall m$$

Using this procedure we have generated the BGNF series for $H_\Omega(x,p)$ as given in Eq. (2.1) for arbitrary b_Ω and g_Ω (hereafter referred to as b and g). The first few terms of the infinite BGNF series with $b^2 = \lambda g$, $\kappa = gK$, and $\kappa_0 = gK_0$ are given

$$\begin{aligned}\kappa = \kappa_0\Big\{1 &+ \Big[\frac{3}{4} - \frac{15}{4}\lambda\Big]\kappa_0 + \Big[-\frac{17}{16} + \frac{225}{8}\lambda - \frac{705}{16}\lambda^2\Big]\kappa_0^2 \\ &+ \Big[\frac{375}{128} - \frac{24945}{128}\lambda + \frac{116325}{128}\lambda^2 - \frac{115755}{128}\lambda^3\Big]\kappa_0^3 \\ &+ \Big[\frac{10689}{1024} + \frac{338625}{256}\lambda - \frac{6383475}{512}\lambda^2 + \frac{8163729}{256}\lambda^3 - \frac{23968161}{1024}\lambda^4\Big]\kappa_0^4 \\ &+ \Big[\frac{87549}{2048} - \frac{18237765}{2048}\lambda + \frac{145625865}{1024}\lambda^2 - \frac{687587901}{1024}\lambda^3 + \frac{2424834657}{2048}\lambda^4 - \frac{1412410545}{2048}\lambda^5\Big]\kappa_0^5 \\ &+ \Big[-\frac{3132399}{16384} + \frac{489189285}{8192}\lambda - \frac{23869797345}{16384}\lambda^2 + \frac{44855525331}{4096}\lambda^3 \\ &- \frac{556379422209}{16384}\lambda^4 + \frac{374415391845}{8192}\lambda^5 - \frac{361809217935}{16384}\lambda^6\Big]\kappa_0^6 \\ &+ \Big[\frac{238225977}{262144} - \frac{104743883475}{262144}\lambda + \frac{3644984868525}{262144}\lambda^2 - \frac{39986908838415}{262144}\lambda^3 \\ &+ \frac{190295579741355}{262144}\lambda^4 - \frac{437037206771385}{262144}\lambda^5 + \frac{475728673084335}{262144}\lambda^6 \\ &- \frac{196443710834085}{262144}\lambda^7\Big]\kappa_0^7 + \cdots\Big\}. \end{aligned} \tag{3.5}$$

It has been shown[17] that the BGNF series may be recovered from the quantum RS perturbation series for the Hamiltonian given in Eq. (2.1) by replacing the quantum operators by their corresponding classical functions. However, this quantum to classical mapping is not invertible. The BGNF series *cannot* be quantized so as to yield the RS series,[20] but rather, the BGNF algorithm outlined above must be modified by replacing the Poisson brackets (Lie derivatives) with quantum commutators to obtain the *quantum normal form*[17] identical to the RS series.

The series in Eq. (3.5) is given in powers of gK_0, where the action K_0 defined by

$$K_0 = \frac{1}{2}(\xi^2 + \eta^2) = \frac{1}{2\pi}\oint p\,dx \tag{3.6}$$

is the constant of motion for our one-dimensional problem. Since the normal ordered Hamiltonian κ is given in terms of the action, the study of classical and semiclassical properties of the anharmonic system becomes simple. For example, Hamilton's equations of motion are given by

$$\frac{d\xi}{dt} = \omega\eta, \quad \frac{d\eta}{dt} = -\omega\xi,$$

where

$$\omega = \frac{d\kappa}{d\kappa_0} = \text{a constant for a given } K_0$$

and the period T is given by

$$T = 2\pi/\omega. \tag{3.7}$$

The Einstein–Brillouin–Keller[21] (EBK) semiclassical energies are obtained by substituting

$$K_0 = (m + \tfrac{1}{2}) \tag{3.8}$$

(where m is an integer) in Eq. (3.5). The simplicity of the BGNF approach is offset by the fact that as in any power series approach, the radius of convergence of the series limits the region in which accurate results—which may require a very large number of terms—can be obtained. In order to extend the range of applicability of our BGNF series, we now turn our attention to the study of Eq. (3.5).

We begin by noting that since the coefficients of κ_0^n are polynomials in λ of order $n-1$, the series for κ is in fact a double-power series. It can be shown[11] that Eq. (3.5) is the inversion of the double-series

$$\kappa_0 = \kappa\sum_{i,j}\frac{\Gamma(3i+2j+\frac{1}{2})}{\Gamma(i+\frac{1}{2})(i+j+1)!}\frac{(2\lambda\kappa)^i(-2\kappa)^j}{i!j!}. \tag{3.9}$$

In order to invert the series in Eq. (3.9), we express it as $\kappa_0 = \kappa\phi(\kappa)$, where $\phi(\kappa)$ represents the terms under the summation sign. The inversion is then given by

$$\kappa = \sum_{n=1}^{\infty}\frac{\kappa_0^n}{n!}\Big[\frac{d^{n-1}}{d\kappa^{n-1}}\frac{1}{\phi(\kappa)^n}\Big]_{\kappa=0}.$$

As pointed out in Ref. 11, the series in Eq. (3.9) has a small (albeit nonzero) region of convergence which puts severe limitations on the values of κ and κ_0 to compute each other. For the SDW case these limitations were overcome by using Padé approximants and a combination of the analytic continuation of the relevant series and Padé approximants. In this paper we are concerned with the general ASDW case. After some algebra Eq. (3.9) can be expressed in the following forms:

$$\kappa_0 = \kappa \sum_{m,n} \frac{\Gamma(2m+n+\frac{1}{2})}{\Gamma(n+\frac{1}{2})(m+1)(m-n)!} \frac{(-\lambda)^n(-2\kappa)^m}{n!m!}, \tag{3.10}$$

$$\kappa_0 = \kappa \sum_j \frac{\Gamma(2j+\frac{1}{2})}{\Gamma(\frac{1}{2})(j+1)!} {}_2F_1(-j,2j+\tfrac{1}{2},\tfrac{1}{2},\lambda) \frac{(-2\kappa)^j}{j!}, \tag{3.11a}$$

$$\kappa_0 = \kappa + \kappa \sum_{j=1}^{\infty} \sum_{m=0}^{j} \frac{\Gamma(2j+\frac{1}{2}) 2^j (\lambda-1)^{j-m} \lambda^m \kappa^j}{\Gamma(\frac{1}{2})(j+1)!j!T(j,m)}, \tag{3.11b}$$

where

$$T(j,m+1) = [(m+1)(m+\tfrac{1}{2})/(2j-m)(j-m)]T(j,m),$$

$$T(j,0) = 1$$

and

$$\kappa_0 = \frac{\kappa}{\sqrt{2\pi}} \sum_j \frac{\Gamma(3j/2+\frac{1}{4})\Gamma(3j/2+\frac{3}{4})}{\Gamma(j+\frac{1}{2})\Gamma(j+2)} \frac{(16\lambda\kappa)^j}{j!} \times {}_2F_1\left(\frac{3j}{2}+\frac{1}{4},\frac{3j}{2}+\frac{3}{4},j+2,-8\kappa\right). \tag{3.12}$$

Setting $\lambda = 1$ in Eqs. (3.9)–(3.12) gives results for the SDW case. It is interesting to note that according to the definition of Horn (see Erdelyi[13]), the double-hypergeometric series of Eq. (3.9) is of *order* 3, while that of Eq. (3.10) is of *order* 2. This is another example of the nonuniqueness of the order of multiple-hypergeometric series. Since one of the parameters of F in Eq. (3.11a) is $-j$, the hypergeometric series is simply a polynomial. Equation (3.11b) is convenient for numerical calculations, especially when $\lambda \geqslant 1$. It is seen that for $\lambda \geqslant 1$, the terms in Eq. (3.11b) are positive definite and hence the difficulty of adding large terms of alternating signs, as in Eq. (3.11a), is removed. For large κ, we obtain the following expression by analytically continuing the function F in Eq. (3.12) and collecting similar terms:

$$\kappa_0 = \frac{1}{2\pi}\left[(8\kappa)^{3/4} \sum_{j=0}^{\infty} \left(-\frac{1}{8}\right)^{j+1} \frac{\Gamma(j+\frac{1}{4})\Gamma(j-\frac{3}{4})}{\Gamma(j+\frac{1}{2})j!\kappa^j} \times {}_2F_1(-2j,j+\tfrac{1}{4},\tfrac{1}{2},\lambda) - (8\kappa)^{1/4} \sum_{j=0}^{\infty}\left(-\frac{1}{8}\right)^{j+1} \frac{\Gamma(j+\frac{3}{4})\Gamma(j-\frac{1}{4})}{\Gamma(j+\frac{3}{2})j!\kappa^j} \times {}_2F_1(-2j-1,j+\tfrac{3}{4},\tfrac{1}{2},\lambda)\right]. \tag{3.13}$$

We now invert Eq. (3.13) to obtain an expression valid for large κ_0:

$$\kappa = \frac{1}{z^2} \sum_i a_i z^i, \tag{3.14}$$

where

$$z = (\alpha\kappa_0)^{-2/3}, \quad \alpha = 3\Gamma(3/4)^2/\pi^{1/2}2^{5/4}$$

and the first few a_i's are given by

$$a_0 = 1, \quad a_1 = -\tfrac{4}{3}G_0, \quad a_2 = \tfrac{2}{3}G_0^2 - \tfrac{1}{16}F_1, \quad a_3 = -\tfrac{8}{81}G_0^3 + \tfrac{1}{48}(G_0F_1 - G_1),$$

$$a_4 = -\frac{5}{243}G_0^4 + \frac{1}{144}G_0(G_0F_1 - G_1) - \frac{1}{6144}(3F_1^2 - 5F_2), \quad a_5 = \frac{5}{6144}G_0F_2 - \frac{1}{1024}G_1F_1 + \frac{7}{10240}G_2,$$

$$a_6 = \frac{14}{6561}G_0^6 - \frac{5}{3888}G_0^3(G_0F_1 - G_1) + \frac{1}{55296}G_0(25G_0F_2 + 63G_2) + \frac{5}{18432}(G_0^2F_1^2 - 4G_0G_1F_1 - G_1^2) + \frac{5}{393216}(5F_1F_2 - 3F_3 - 2F_1^3),$$

$$a_7 = \frac{16}{19683}G_0^7 + \frac{7}{41472}G_0(G_0^2F_1^2 - 3G_0G_1F_1 - 3G_1^2) - \frac{7}{11664}G_0^4(G_0F_1 - G_1) - \frac{7}{294912}F_1(G_0F_1^2 + 3G_1F_1 - 5G_0F_2) + \frac{49}{46080}G_0^2G_2 + \frac{35}{248832}G_0^3F_2 - \frac{35}{1179648}(3G_0F_3 - G_1F_2) + \frac{49}{655360}G_2F_1 - \frac{77}{1966080}G_3,$$

$$a_8 = \frac{15}{131072}G_0^2(F_1F_2 - F_3) - \frac{1}{2048}G_0^2G_1^2 + \frac{1}{65536}G_1(5G_0F_2 - 9G_0F_1^2 - 3G_1F_1) + \frac{45}{8388608}F_1(F_1F_2 - F_3) + \frac{7}{655360}(64G_0^3G_2 - 11G_0G_3 + 18G_0G_2F_1 + 3G_1G_2) - \frac{1}{33554432}(54F_1^4 + 25F_2^2) + \frac{585}{234881024}F_4,$$

where

$$F_j \equiv {}_2F_1(-2j,j+\tfrac{1}{4},\tfrac{1}{2},\lambda), \quad G_j \equiv C_2F_1(-2j-1,j+\tfrac{3}{4},\tfrac{1}{2},\lambda), \quad C = 3\Gamma(\tfrac{3}{4})^4/\pi^2 2^{3/2}.$$

IV. NUMERICAL RESULTS OF THE ASDW OSCILLATOR

In this section we illustrate the practical usefulness of the BGNF series given in Eq. (3.5) and its analytic continuation (3.14) by presenting some numerical results for these series as well as their Padé approximants. The numerical results were obtained by setting $b = 0.075$ and $g = 0.005$ (corresponding to $\lambda = 9/8$). A plot of the potential with the coordinate origin at the bottom of the right well is given in Fig. 1. For these parameter values, the barrier height $V(x_2)$ and the depth of the left well $V(x_3)$ assume the values

5.04288998566897891 and −15.49210873566897891, respectively.

A. Classical action and energy

A set of numerical values of the classical energy $E = \kappa/g$ and the corresponding action $K_0 = \kappa_0/g$ are given in Tables I and II. Table I contains the results for $E < V(x_2)$, while the entries of Table II are for $E > V(x_2)$. Since the BGNF series converges slowly when E is close to the barrier height $V(x_2)$, we have considered Padé approximants of the series in Eqs. (3.5) and (3.14).

In Table I, the results of two types of Padé approximants are reported. Following our procedure developed for the SW and SDW oscillators, we have obtained the classical energies from a [14,13] Padé approximant of the cube of the series in Eq. (3.5) for the two asymmetric wells. The cube of the series in Eq. (3.5) is expressed as $\kappa^3 = \kappa_0^3 U$. Then U is approximated by a [14,13] Padé approximant ($U \approx P[14,13]$) which leads to $\kappa \approx \kappa_0 P[14,13]^{1/3}$. The different results for the two wells are obtained by implementing the scalings for b, g, and the Hamiltonian given in Sec. II. These types of Padé results are presented in columns 2 and 6 of Table I. The other type of approximation is obtained by using a [14,13] Padé approximant of the series itself; the results are shown in columns 3 and 7. The results in columns 4 and 8 are obtained by using 30 terms in the sum of the series, while the exact results in columns 5 and 9 are obtained by solving Eq. (3.11b) for given K_0. It may be mentioned that the results obtained from Eq. (3.11b) and those from the numerical integration of Eq. (3.6) agree to all figures within the numerical accuracy. It can be seen from Table I that the results from the two types of Padé approximations and those from the series compare well with the exact results. As E approaches the value $V(x_2)$, the disagreement among these results begins to show; it becomes clear that the Padé approximant of the cube of the series yields the best results for any given number of terms in the series.

TABLE II. The classical energy as a function of the action above the potential barrier for $b = 0.075$ and $g = 0.005$. The results under $E(PS)$ are obtained from a [13,13] Padé approximant of Eq. (3.14), while those under $E(SR)$ are obtained by using 28 terms in the series given in Eq. (3.14). The numbers for E(EXACT) are obtained by numerically integrating the action integral given in Eq. (3.6).

K_0	$E(PS)$	$E(SR)$	E(EXACT)
25.4	5.07465	5.06957	5.09037
25.5	5.10449	5.10012	5.11749
26.5	5.42899	5.42829	5.43206
27.5	5.79067	5.79087	5.79167
57.5	23.32457	23.32457	23.32457
58.5	24.05134	24.05134	24.05134
59.5	24.78519	24.78519	24.78519
100.5	59.98498	59.98498	59.98498
101.5	60.95061	60.95061	60.95061
102.5	61.92072	61.92072	61.92072
148.5	110.88764	110.88764	110.88764
149.5	112.03772	112.03772	112.03772
150.5	113.19113	113.19113	113.19113

In Table II, the entries in column 2 are obtained from a [13,13] Padé approximation of the series in Eq. (3.14), while those in column 3 are obtained by using 28 terms in the sum of this series. The exact results in column 4 are obtained by numerically integrating Eq. (3.6). Once again we notice

TABLE I. The classical energy as a function of the action in the asymmetric wells for $b = 0.075$ and $g = 0.005$. Here $E(LPC)$, $E(LPS)$, and $E(LSR)$ and $E(RPC)$, $E(RPS)$, and $E(RSR)$ represent energies for a given K_0 in the left and right wells, respectively. We obtain $E(LPC)$ and $E(RPC)$ from a [14,13] Padé approximant of the cube of the series in Eq. (3.5); $E(LPS)$ and $E(RPS)$ are obtained from a [14,13] Padé approximant of Eq. (3.5); and $E(LSR)$ and $E(RSR)$ are obtained from Eq. (3.5) using 30 terms in the sum. The E(EXACT) are obtained from Eq. (3.11b).

K_0	$E(LPC)$	$E(LPS)$	$E(LSR)$	E(EXACT)	$E(RPC)$	$E(RPS)$	$E(RSR)$	E(EXACT)
0.50	−14.84400	−14.84400	−14.84400	−14.84400	0.49558	0.49558	0.49558	0.49558
1.50	−13.56033	−13.56033	−13.56033	−13.56033	1.45859	1.45859	1.45859	1.45859
2.50	−12.29389	−12.29389	−12.29389	−12.29389	2.37950	2.37950	2.37950	2.37950
3.50	−11.04534	−11.04534	−11.04534	−11.04534	3.25052	3.25052	3.25052	3.25052
4.50	−9.81540	−9.81540	−9.81540	−9.81540	4.05845	4.05845	4.05845	4.05845
5.50	−8.60485	−8.60485	−8.60485	−8.60485	4.77345	4.77345	4.77356	4.77345
5.90	−8.12626	−8.12626	−8.12626	−8.12626	5.01000	5.01001	5.01320	5.00996
6.50	−7.41460	−7.41460	−7.41460	−7.41460				
7.50	−6.24566	−6.24566	−6.24566	−6.24566				
8.50	−5.09917	−5.09917	−5.09917	−5.09917				
9.50	−3.97645	−3.97645	−3.97645	−3.97645				
10.50	−2.87905	−2.87905	−2.87905	−2.87905				
11.50	−1.80881	−1.80881	−1.80881	−1.80881				
12.50	−0.76796	−0.76796	−0.76796	−0.76796				
13.50	0.24073	0.24073	0.24073	0.24073				
14.50	1.21368	1.21368	1.21368	1.21368				
15.50	2.14608	2.14608	2.14608	2.14608				
16.50	3.03095	3.03095	3.03100	3.03095				
17.50	3.85697	3.85697	3.85748	3.85697				
18.50	4.60115	4.60115	4.60595	4.60115				
19.13	4.99611	4.99617	5.02093	4.99532				

TABLE III. Comparison of the semiclassical and quantum energies of the ASDW oscillator for $b = 0.075$ and $g = 0.005$. The entries in columns 2, 5, and 8 are taken from Tables I and II. The quantum energies are obtained by diagonalizing a 1800×1800 Hamiltonian matrix in the harmonic oscillator basis set with $\beta^2 = 2.7482987390560$.

n_L	E^L_{BGNF}	E_{quant}	n_R	E^R_{BGNF}	E_{quant}	n	E_{BGNF}	E_{quant}
0	− 14.84400	− 14.84467	0	0.49558	0.49395	25	5.10449	5.16383
1	− 13.56033	− 13.56102	1	1.45859	1.45668	26	5.42899	5.42597
2	− 12.29389	− 12.29462	2	2.37950	2.37717	27	5.79067	5.80092
3	− 11.04534	− 11.04610	3	3.25052	3.24742	...	...	...
4	− 9.81540	− 9.81619	4	4.05845	4.05351	...	...	...
5	− 8.60485	− 8.60568	5	4.77345	4.76461	57	23.32457	23.32501
6	− 7.41460	− 7.41548				58	24.05134	24.05177
7	− 6.24566	− 6.24659				59	24.78519	24.78562
8	− 5.09917	− 5.10015				...	...	...
9	− 3.97645	− 3.97751				...	...	...
10	− 2.87905	− 2.88020				100	59.98498	59.98526
11	− 1.80881	− 1.81007				101	60.95061	60.95089
12	− 0.76796	− 0.76936				102	61.92072	61.92100
13	0.24073	0.23915				...	...	...
14	1.21368	1.21186				...	...	...
15	2.14608	2.14388				148	110.88764	110.88786
16	3.03095	3.02817				149	112.03772	112.03794
17	3.85697	3.85271				150	113.19113	113.19135
18	4.60115	4.58974						

that except for a small region near the top of the potential barrier, our analytic continuation of the BGNF series and its Padé approximation provides accurate classical results.

B. Comparison of semiclassical and quantum energies

Here we compare semiclassical and quantum energies of the ASDW oscillator for the sample parameter values mentioned above. The quantum energies E_{quant} were obtained by diagonalizing a 1800×1800 Hamiltonian matrix generated by the basis set

$$\psi_n = [\beta/\sqrt{\pi}2^n n!]^{1/2} H_n(\beta x) e^{-(\beta^2 x^2)/2},$$

where H_n is the Hermitian polynomial of order n and β is an adjustable parameter. The value of the parameter $\beta^2 = 2.7482987390560$ was obtained by minimizing the trace of the Hamiltonian matrix. The quantum energies obtained in this manner are given in Table III.

As mentioned in Sec. III, the semiclassical energies are easily obtained in the BGNF approach by setting the action $K_0 = m + \frac{1}{2}$ (where m is an integer) in the BGNF series (3.5) and its analytic continuation (3.14): We present in Table III only the Padé approximant results of these series. The numbers in columns 2 and 5 are taken from columns 2 and 6 of Table I and those in column 8 are taken from column 2 of Table II.

It can be seen from Table III that the BGNF approach provides the EBK energies of the ASDW oscillator for all values of the quantum number m in the interval $0 \leqslant m \leqslant \infty$. Except for a small region near the top of the barrier, the agreement between the semiclassical and quantum results is quite good. This agreement improves as m increases and the real benefit of the semiclassical approach is appreciated for the very large values of m for which the quantum calculations become prohibitive. It is interesting to notice how the semiclassical quantum numbers in the two wells are combined to provide the starting quantum number in the region above the barrier height. Here we have 19 levels in the left well and six levels in the right well, giving the total of 25 levels in the two wells. The first quantum number above the potential hump is 25, indicating the 26th level if the quantum numbers are counted from zero. This feature of the EBK quantum numbers is the same in the SDW and ASDW cases.

V. SUMMARY

In summary, we have presented the BGNF series for a general ASDW oscillator with a polynomial potential of degree 4. By identifying this BGNF series as the inversion of a double-series involving hypergeometric functions, we were able to obtain expressions that are valid beyond the region of convergence of the original series. In turn, these expressions allowed us to demonstrate the practical usefulness of the BGNF approach in determining semiclassical energies, which were compared with the corresponding quantum energies. An interesting feature of the EBK quantum numbers is discussed. A rather surprising observation concerning the classical period in the two wells is also reported.

ACKNOWLEDGMENT

This research was supported in part by a grant to MKA from the Natural Sciences and Engineering Research Council of Canada.

[1] Y. Onodera, Prog. Theor. Phys. **44**, 1477 (1970).
[2] J. P. Codaccioni and R. Caboz, J. Math. Phys. **25**, 2436 (1984).
[3] L. S. Luo and J. Nuttall, Phys. Rev. Lett. **57**, 2241 (1986); P. E. Shanley, Phys. Lett. A **117**, 161 (1986); see, also, the references in citation 11.
[4] C. M. Bender and T. T. Wu, Phys. Rev. **184**, 1231 (1969).
[5] S. Graffi, V. Grecchi, and B. Simon, Phys. Lett. B **32**, 631 (1970).
[6] B. Simon, Ann. Phys. **58**, 76 (1970).
[7] B. Simon, Int. J. Quant. Chem. **21**, 3 (1982).

[8]J. F. Loiseau, J. P. Codaccioni, and R. Caboz, Math. Comput. **50**, 501 (1988).
[9]B. Stec, Acta Phys. Pol. A **68**, 681 (1985).
[10]M. K. Ali, W. R. Wood, and J. S. Devitt, J. Math. Phys. **27**, 1806 (1986).
[11]M. K. Ali and W. R. Wood, Prog. Theor. Phys. **78**, 766 (1987).
[12]P. F. Byrd and M. D. Friedman, *Handbook of Elliptic Integrals for Engineers and Scientists* (Springer, Berlin, 1971).
[13]M. Abramowitz and I. A. Stegun, *Handbook of Mathematical Functions* (Dover, New York, 1968); A. Erdelyi, *Higher Transcendental Functions* (McGraw-Hill, New York, 1953), Vol. I.
[14]G. D. Birkhoff, *Dynamical Systems* (AMS) Colloquium, New York, 1927), Vol. IX; V. I. Arnold, *Mathematical Methods of Classical Mechanics* (Springer, New York, 1980); C. L. Siegel, Ann. Math. **42**, 806 (1941).
[15]S. Brajamani and P. S. Mazumdar, Int. J. Theor. Phys. **27**, 397 (1988); M. Robnik, J. Phys. A **17**, 109 (1984); T. Uzer and R. A. Marcus, J. Chem. Phys. **81**, 5013 (1984); C. Jaffé and W. P. Reinhardt, *ibid.* (1982); R. B. Shirts and W. P. Reinhardt, **77**, 5204 (1982); W. P. Reinhardt and D. Farelly, J. Phys. (Paris) Colloq. **43**, C2 (1982); R. T. Swimm and J. B. Delos, J. Chem. Phys. **71**, 1706 (1979); D. W. Noid and R. A. Marcus, *ibid.* **67**, 559 (1977); D. W. Noid, M. L. Koszykowski, M. Tabor, and R. A. Marcus, *ibid.* **73**, 6169 (1980); R. C. Churchill, G. Pecelli, and D. L. Rod, in *Lecture Notes in Physics* (Springer, Berlin, 1978), Vol. 93; R. C. Churchill, M. Kummer, and D. L. Rod, private communication, 1981); R. C. Churchill and D. Lee, Notes Pure Appl. Math. **92**, 239 (1984).
[16]F. G. Gustavson, Astron. J. **71**, 670 (1966).
[17]M. K. Ali, J. Math. Phys. **26**, 2565 (1985).
[18]A. Deprit, Celest. Mech. **1**, 12 (1969).
[19]A. A. Kamel, Celest. Mech. **3**, 90 (1970).
[20]W. R. Wood and M. K. Ali, J. Phys. A **20**, 351 (1987).
[21]A. Einstein, Verhandl, Deut. Ges. **19**, 82 (1917); L. Brillouin, J. Phys. Radium **7**, 353 (1926); J. B. Keller, Ann. Phys. **4**, 180 (1958).

Asymptotic velocities in classical mechanics

Eugene Gutkin
Department of Mathematics, University of Southern California, Los Angeles, California 90089-1113

(Received 16 September 1988; accepted for publication 2 February 1989)

The scattering of a system of classical particles on the line with repulsive interactions is considered. The scattering is called regular if for any motion of the system the asymptotic velocities are distinct. It was known that the scattering is regular if the interactions decay sufficiently fast with the distance. Here the dynamics of repelling particles are put into the framework of Hamiltonians with cone potentials. The notion of regularity of scattering is defined for Hamiltonian systems with cone potentials and conditions for regularity are established that do not depend on the rate of decay of potentials at infinity. Then these results are applied to the scattering of repelling particles on the line.

I. INTRODUCTION

Consider the classical motion of n point masses $m_1,...,m_n$ with repulsive interactions. Let the particles move in the space $\mathbb{R}^d$ of d dimensions and denote by $x_1,...,x_n$ their positions and by $\dot{x}_1,...,\dot{x}_n$ their velocities. It is known that the asymptotic velocities (at plus infinity, for concreteness)

$$\dot{x}_i(\infty) = \lim_{t\to\infty} \dot{x}_i(t), \quad i = 1,...,n, \tag{1}$$

exist for any trajectory of the motion.[1,2] It is intuitively clear that for a typical trajectory the asymptotic velocities are distinct:

$$\dot{x}_1(\infty) \neq \cdots \neq \dot{x}_n(\infty). \tag{2}$$

We call property (2) the regularity of asymptotic velocities.

To study the scattering in classical mechanics,[3] it is important to know that the asymptotic velocities are regular for all motions. Property (2) has been studied for one-dimensional particles that cannot pass each other.[4–6] In this situation it is known that if the interactions between particles decay fast with the separation, then the asymptotic velocities are regular for any motion. The fast decay condition seemed natural because it is also needed for the existence of asymptotic phases.[7]

A regularity property similar to (2) has been introduced for the Hamiltonians with cone potential.[8] Such a Hamiltonian in the space of N dimensions is characterized by a proper cone C in $\mathbb{R}^N$, and the asymptotic velocity $\dot{x}(\infty)$ of any motion $\{x(t)\}$ exists and belongs to the dual cone C^*. Let Int C denote the interior of C. The regularity property corresponding to (2) is

$$\dot{x}(\infty) \in \text{Int } C^*. \tag{3}$$

Under the fast decay assumption on the cone potential, (3) holds for all motions, and the fast decay is needed for the existence of scattering in the context of Hamiltonians with cone potentials.[4]

Thus, so far, the regularity of asymptotic velocities (2) and (3) has been associated with the fast decay of interactions, and therefore with the scattering. In Ref. 5, Hubacher has considered the systems of classical particles of equal mass m on the line interacting by a repulsive pair potential $v(x)$. The Hamiltonian of such a system is

$$H = \frac{m}{2}\sum_{i=1}^{n} \dot{x}_i^2 + \sum_{i<j} v(x_j - x_i). \tag{4}$$

Assume that the pair potential is repulsive, more precisely that it satisfies the following conditions.

Repulsivity Conditions: Potential v is a positive continuously differentiable function on (x_0, ∞), where $-\infty \leqslant x_0$. Besides $v(x) \to \infty$ as $x \to x_0$ from the right, and $v'(x) < 0$.

Under these conditions, the limit of $v(x)$, as $x \to \infty$, exists, and in what follows we assume without loss of generality that this limit is zero. No assumptions on fast decay of $v(x)$ as $x \to \infty$ are made.

Theorem[5]: For particle systems with the Hamiltonian (4), the repulsivity conditions are necessary and sufficient for the regularity (of asymptotic velocities)

$$\dot{x}_1(\infty) < \cdots < \dot{x}_n(\infty) \tag{5}$$

for all motions.

In this paper we extend the sufficiency part of Hubacher's theorem. Namely, we show that the regularity (5) still holds when the repulsive pair potentials $v_{ij}(x)$ are arbitrary, as long as they have, roughly speaking, the same rate of decay as $x \to \infty$. We also get rid of the assumption that the masses of particles are equal. To prove this result, we use the technique of cone potentials, and generalize the regularity (5) to a class of Hamiltonians with cone potentials. Hence our results separate the regularity of asymptotic velocities from the fast decay of interactions assumption.

II. REGULARITY OF ASYMPTOTIC VELOCITIES FOR A CLASS OF CONE POTENTIALS

Let $f_1(x)$, $f_2(x)$ be continuous positive functions defined on (x_1, ∞), (x_2, ∞). Assume that as $x \to \infty$ both functions go to zero or to infinity.

Definition 1: We say that f_1 and f_2 have the same rate (of decay or growth) at infinity if

$$f_1(x)/f_2(x) = O(1), \quad f_2(x)/f_1(x) = O(1), \tag{6}$$

as $x \to \infty$.

The following is a variation of a lemma from Ref. 5.

Lemma 1: Assume that the functions $f_1(x)$, $f_2(x)$ have the same rate of decay at infinity and that $\int_a^\infty f_1(x)dx < \infty$ for some $a > x_1$. Let $x_1(t) \geqslant x_1$ and $x_2(t) \geqslant x_2$ be continuously dif-

 0022-2488/89/061245-05$02.50

ferentiable for $t \geqslant 0$, with $\dot{x}_1(t) \to 0$ and $\dot{x}_2(t) \to b > 0$ as $t \to \infty$.

Suppose that

$$\int_0^\infty f_1(x_1(t))dt < \infty. \tag{7}$$

Then for any $\varepsilon > 0$ there is $T \geqslant 0$ such that, for $t > T$,

$$\int_t^\infty f_2(x_2(\tau))d\tau < \varepsilon \int_t^\infty f_1(x_1(\tau))d\tau < \infty. \tag{8}$$

Proof: Clearly, $x_2(t) \to \infty$ as $t \to \infty$. By (7), $x_1(t) \to \infty$, also. Indeed, if this is not the case, then, since $\dot{x}_1(t)$ is bounded, there is $L > 0$ such that $x_1(t) < L$ for all t. Then, for any T,

$$\int_0^T f_1(x_1(t))dt \geqslant \left[\min_{x_1 \leqslant x \leqslant L} f(x)\right] T, \tag{9}$$

contradicting (7).

For T large enough, the following inequalities hold for all $t > T$:

$$x_1(t) < x_2(t), \quad |\dot{x}_1(t)| < \varepsilon b/2, \quad \dot{x}_2(t) > b/2. \tag{10}$$

Then, for $t > T$,

$$\begin{aligned}
&\frac{\varepsilon b}{2} \int_t^\infty f_1(x_1(\tau))d\tau \\
&\quad \geqslant \left| \int_t^\infty f_1(x_1(\tau))dx_1(\tau) \right| \\
&\quad = \int_{x_1(t)}^\infty f_1(x)dx \geqslant \int_{x_2(t)}^\infty f_1(x)dx \\
&\quad = \int_t^\infty f_1(x_2(\tau))\dot{x}_2(\tau)d\tau \geqslant \frac{b}{2} \int_t^\infty f_1(x_2(\tau))d\tau.
\end{aligned}$$

Hence, for $t > T$ large enough,

$$\int_t^\infty f_1(x_2(\tau))d\tau \leqslant \varepsilon \int_t^\infty f_1(x_1(\tau))d\tau. \tag{11}$$

Since $x_2 \to \infty$, by (6), for $t > T$ large enough,

$$\int_t^\infty f_2(x_2(\tau))d\tau \leqslant c \int_t^\infty f_1(x_2(\tau))d\tau, \tag{12}$$

with some positive constant c. Combining (11) and (12), we obtain Lemma 1.

Denote by $\langle x, y \rangle$ the inner product in the Euclidean space $\mathbb{R}^N$ and set $\|x\|^2 = \langle x,x \rangle$. Recall that a cone C in $\mathbb{R}^N$ is called proper if C does not contain a straight line. Let $e_1,...,e_n$ be arbitrary nonzero vectors. The set

$$C = \left\{ \sum_{i=1}^n r_i e_i, \quad r_i \geqslant 0 \right\}$$

is the cone spanned by $e_1,...,e_n$. The dual cone C^* consists of vectors e^* such that $\langle e^*, e_i \rangle \geqslant 0$ for all e_i.

With any vectors $e_1,...,e_n$ in $\mathbb{R}^N$, any continuously differentiable functions v_i on (a_i, ∞), $a_i \geqslant -\infty$, and a positive definite matrix M (mass matrix), we associate the Hamiltonian

$$H = \frac{1}{2} \langle M\dot{x}, \dot{x} \rangle + \sum_{i=1}^n v_i(\langle x,e_i \rangle). \tag{13}$$

The corresponding dynamics $x(t)$ takes place in the region

$$\Omega = \{x: \ \langle x,e_i \rangle > a_i, \quad i = 1,...,n\}.$$

The limit $\dot{x}(\infty) = \lim \dot{x}(t)$, as $t \to \infty$, if it exists for a given trajectory $x(t)$, is called the asymptotic velocity at infinity.

In what follows we assume that the potentials $v_i(r)$ satisfy the repulsivity conditions of Sec. 1, and that the cone C spanned by $e_1,...,e_n$ is proper. Then (13) is a Hamiltonian with the cone potential[8]

$$V(x) = \sum_{i=1}^n v_i(\langle x,e_i \rangle). \tag{14}$$

For such Hamiltonians the asymptotic velocities $\dot{x}(\infty)$ exist for all motions and $\dot{x}(\infty) \in C^*$ (see Ref. 8).

Definition 2: We say that a motion $x(t)$ has a regular asymptotic velocity (at plus infinity) if $\dot{x}(\infty) \in \text{Int } C^*$.

Theorem 1: Let the derivatives $f_i(r) = -v_i'(r)$, $i = 1,...,n$, have the same rate of decay as $r \to \infty$. Then every motion of the Hamiltonian dynamics defined by (13) is regular.

Proof: Using a simple transformation,[4] we reduce the situation to the case $M = 1$, i.e., the Hamiltonian is

$$H = \frac{1}{2} \|\dot{x}\|^2 + \sum_{i=1}^n v_i(\langle x,e_i \rangle). \tag{15}$$

Denote $\dot{x}(\infty)$ by b. Then

$$\langle b,e_i \rangle \geqslant 0, \quad i = 1,...,n, \tag{16}$$

and $b \in \text{Int } C^*$ if and only if we have strict inequalities in (16). We assume that b does not belong to C^* and relabel the vectors e_i so that

$$\begin{aligned}
&\langle b,e_i \rangle = 0, \quad i = 1,...,m \geqslant 1, \\
&\text{and} \\
&\langle b,e_i \rangle > 0, \quad i = m+1,...,n.
\end{aligned} \tag{17}$$

Denote by C_0 the cone spanned by $e_1,...,e_m$. Since $C_0 \subset C$, the cone C_0 is proper, and the intersection $\text{Int } C_0 \cap \text{Int } C_0^*$ is not empty.

Denote by $F(x)$ the force vector at x, i.e.,

$$\begin{aligned}
F(x) = -\text{grad } V(x) &= \sum_{i=1}^n -v_i'(\langle x,e_i \rangle)e_i \\
&= \sum_{i=1}^n f_i(\langle x,e_i \rangle)e_i.
\end{aligned} \tag{18}$$

The equation of motion corresponding to (15) is

$$\frac{d}{dt}\dot{x}(t) = F(x(t)), \tag{19}$$

and, for any vector e,

$$\frac{d}{dt}\langle \dot{x}(t),e \rangle = \sum_{i=1}^n f_i(\langle x(t),e_i \rangle)\langle e_i,e \rangle. \tag{20}$$

Hence

$$\begin{aligned}
&\int_t^\infty \frac{d}{dt}\langle \dot{x}(\tau),e \rangle d\tau \\
&\quad = \langle b,e \rangle - \langle \dot{x}(t),e \rangle = \sum_{i=1}^n \left[\int_t^\infty f_i(\langle x(\tau),e \rangle) \right] \langle e_i,e \rangle.
\end{aligned} \tag{21}$$

Taking e from $\text{Int } C^*$, we obtain that

$$\int_t^\infty f_i(\langle x(\tau),e_i \rangle)d\tau < \infty, \tag{22}$$

for all i. By the proof of Lemma 1, $\langle x(t),e_i \rangle \to \infty$, as $t \to \infty$, for all i.

Let now $e \in \text{Int } C_0 \cap \text{Int } C_0^*$. Rewriting (20) as

$$\frac{d}{dt}\langle \dot{x}(t),e_i\rangle = \sum_{i=1}^{m} f_i(\langle x(t),e_i\rangle)\langle e_i,e\rangle + \sum_{i=m+1}^{n} f_i(\langle x(t),e_i\rangle)\langle e_i,e\rangle, \tag{23}$$

integrating (23) from t to infinity, and using that $\langle b,e\rangle = 0$, by assumption, we obtain

$$\begin{aligned}\langle \dot{x}(t),e\rangle &= -\sum_{i=1}^{m}\int_t^\infty f_i(\langle x(\tau),e_i\rangle)\langle e_i,e\rangle d\tau \\ &\quad -\sum_{i=m+1}^{n}\int_t^\infty f_i(\langle x(\tau),e_i\rangle)\langle e_i,e\rangle d\tau \\ &\leqslant -\sum_{i=1}^{m}\langle e_i,e\rangle\int_t^\infty f_i(\langle x(\tau),e_i\rangle)d\tau \\ &\quad +\sum_{i=m+1}^{n}|\langle e_i,e\rangle|\int_t^\infty f_i(\langle x(\tau),e_i\rangle)d\tau. \end{aligned}\tag{24}$$

For $i = 1,...,m$, $\langle e_i,e\rangle > 0$. By Lemma 1, for any $\varepsilon > 0$ we can find T such that, for $t > T$,

$$\begin{aligned}&\sum_{i=m+1}^{n}|\langle e_i,e\rangle|\int_t^\infty f_i(\langle x(\tau),e_i\rangle)d\tau \\ &\quad \leqslant \varepsilon\sum_{j=1}^{m}\langle e_j,e\rangle\int_t^\infty f_j(\langle x(\tau),e_j\rangle)d\tau. \end{aligned}\tag{25}$$

Taking $\varepsilon = \frac{1}{2}$, we obtain from (24) and (25) that, for t large enough,

$$\langle \dot{x}(t),e\rangle \leqslant -\frac{1}{2}\sum_{i=1}^{m}\langle e_i,e\rangle\int_t^\infty f_i(\langle x(\tau),e_i\rangle)d\tau < 0. \tag{26}$$

Therefore $\langle x(t),e\rangle$ is bounded above as $t \to \infty$, contradicting the assertion $\langle x(t),e_i\rangle \to \infty$, for all i, obtained earlier in the proof. This contradiction proves the theorem.

Discussion: Consider the dynamics $x(t)$ defined by the Hamiltonian (13) where the vectors $e_1,...,e_n$ span a proper cone C. Although the repulsivity of the potentials v_i is essential for the asymptotic velocities $x(\infty)$ to be regular, the other condition of Theorem 1 is not necessary, in general. Consider, for instance, the Hamiltonian (13) with an exponential cone potential[8]

$$V(x) = \sum_{i=1}^{n} c_i e^{-a_i\langle x,e_i\rangle}, \tag{27}$$

with $c_i,a_i > 0$, for $i = 1,...,n$. The functions $f_i(r) = a_i c_i e^{-a_i r}$ have the same rate of decay at infinity only if $a_1 = \cdots = a_n$. Nevertheless, the asymptotic velocity $\dot{x}(\infty)$ exists and is regular for any motion.[4]

Let $v(x)$ satisfy the repulsivity conditions of Sec. I, and consider the special class of Hamiltonians (13):

$$H = \frac{1}{2}\langle M\dot{x},\dot{x}\rangle + \sum_{i=1}^{n} c_i v(\langle x,e_i\rangle), \tag{28}$$

with $c_1,...,c_n > 0$. For the Hamiltonians (28) the rate of decay assumption of Theorem 1 is satisfied automatically. Hence Theorem 1 and the preceding discussion imply the following.

Corollary 1: For the Hamiltonians (28) the repulsivity of v is necessary and sufficient for the asymptotic velocity $\dot{x}(\infty)$ to exist and be regular for all motions.

Consider n repulsive potentials $v_1(r),...,v_n(r)$ satisfying the assumption of Theorem 1, i.e., the derivatives $v_i'(r)$ have the same rate of decay as $r \to \infty$. Recall that we have assumed throughout that $v_i(r) \to 0$ as $r \to \infty$. Since

$$v_i(r) = -\int_r^\infty v_i'(s)ds$$

and since, for any i,j,

$$|v_i'(r)| \leqslant \text{const}\,|v_j'(r)|,$$

for r large enough, potentials $v_i(r)$ satisfy Definition 1, i.e., all v_i have the same rate of decay as $r \to \infty$. The converse fails, in general. Assume now that the functions $v_i(r)$ satisfy a stronger-than-Definition-1 assumption:

$$\lim_{r\to\infty} v_i(r)/v_j(r) \neq 0, \quad \text{for all } i,j. \tag{29}$$

Then, by l'Hôpital's rule,

$$\lim_{r\to\infty} v_i'(r)/v_j'(r) = \lim_{r\to\infty} v_i(r)/v_j(r) \neq 0, \tag{30}$$

for all i,j, which implies the following.

Corollary 2: Let $v_1(r),...,v_n(r)$ be repulsive potentials satisfying Eq. (29). Then all the motions of the dynamics defined by the Hamiltonian (13) have regular asymptotic velocities, i.e., $\dot{x}(\infty) \in \text{Int } C^*$.

III. REGULARITY OF ASYMPTOTIC VELOCITIES FOR MANY-BODY PROBLEMS ON THE LINE

We apply the results of Sec. II to the scattering of classical particles on the line. First, we consider the particles interacting by directed repulsive forces. The Hamiltonian of such a system has the form

$$H = \sum_{i=1}^{N}\frac{m_i}{2}\dot{x}_i^2 + \sum_{i<j} v_{ij}(x_j - x_i), \tag{31}$$

with pair potentials $v_{ij}(r)$ satisfying the repulsivity conditions.

Theorem 2: If the forces $f_{ij}(r) = -v_{ij}'(r)$ have the same rate of decay as $r \to \infty$ or if the pair potentials satisfy

$$\lim_{r\to\infty} v_{ij}(r)/v_{kl}(r) \neq 0, \tag{32}$$

for all pairs $i<j$, $k<l$, then the asymptotic velocities $\dot{x}_i(\infty)$ for any motion satisfy the regularity condition (5).

Proof: Consider the space $\mathbb{R}^N$ with the standard basis $e_1,...,e_N$. The vectors $e_j - e_i$, $i<j$, span a proper cone C with the dual cone $C^* = \{x_1 \leqslant \cdots \leqslant x_N\}$. Setting $x = (x_1,...,x_N)$ and $M = (m_1,...,m_n)$ the diagonal mass matrix, we rewrite (31) as

$$H = \frac{1}{2}\langle M\dot{x},\dot{x}\rangle + \sum_{i<j} v_{ij}(\langle x,e_j - e_i\rangle), \tag{33}$$

which is a Hamiltonian of the form (13). We have

$$\text{Int } C^* = \{x_1 < \cdots < x_N\}.$$

Now Theorem 1 and Corollary 2 imply the assertion.

Let us consider a special class of the Hamiltonians (31):

$$H=\sum_{i=1}^{N}\frac{m_i}{2}\dot{x}_i^2+\sum_{i<j}c_{ij}v(x_j-x_i), \tag{34}$$

where $c_{ij}>0$ and the potential $v(r)$ is arbitrary.

Corollary 3: All of the motions $\{x_i(t),\ 1\leqslant i\leqslant N\}$ defined by the Hamiltonian (34) have regular asymptotic velocities

$$\dot{x}_1(\infty)<\cdots<\dot{x}_N(\infty)$$

if and only if $v(r)$ is a repulsive potential.

Proof: The assertion follows immediately from the proof of Theorem 2 and Corollary 1.

Remark: Setting in Corollary 3, $m_1=\cdots=m_N$ and taking the c_{ij} all equal, we recover Theorem 1 of Ref. 5.

Now we consider particles on the line interacting by central repulsive forces, and assume that the particles cannot pass each other. The Hamiltonian of such a system is

$$H=\sum_{i=1}^{N}\frac{m_i}{2}\dot{x}_i^2+\sum_{i<j}v_{ij}(|x_i-x_j|), \tag{35}$$

where $v_{ij}(r)$ are repulsive potentials on $(0,\infty)$.

Theorem 3: Assume that the forces $f_{ij}(r)=-v_{ij}'(r)$ have the same rate of decay as $r\to\infty$, or that the potentials $v_{ij}(r)$ satisfy condition (32). Then the asymptotic velocities $\dot{x}_i(\infty)$ are distinct:

$$\dot{x}_1(\infty)\neq\cdots\neq\dot{x}_N(\infty), \tag{36}$$

for any motion defined by the Hamiltonian (35).

Proof: Since the particles cannot pass each other, the configuration space is the disjoint union of regions of the form $\{x_{i_1}<\cdots<x_{i_N}\}$. In any such region we relabel the particles so that it becomes $\{x_1<\cdots<x_N\}$ and apply Theorem 2.

Corollary 4: Consider the dynamics of particles on the line with the Hamiltonian

$$H=\sum_{i=1}^{N}\frac{m_i}{2}\dot{x}_i^2+\sum_{i<j}c_{ij}v(|x_i-x_j|), \tag{37}$$

with $c_{ij}>0$.

The asymptotic velocities $\dot{x}_i(\infty)$ for any trajectory of the dynamics defined by the Hamiltonian (37) are distinct if and only if v is a repulsive potential on $(0,\infty)$.

Proof: Just as in the proof of Theorem 3, we reduce the situation to the one considered in Corollary 3.

In applications, one is often led to consider the Hamiltonians (31) and (35) with some potentials v_{ij} equal to zero, or to the Hamiltonians (34) and (37) with some $c_{ij}=0$. For instance, the Hamiltonian of the nearest neighbor interaction,

$$H=\sum_{i=1}^{N}\frac{m_i}{2}\dot{x}_i^2+\sum_{i=1}^{N-1}v_i(x_{i+1}-x_i), \tag{38}$$

is a special case of (31) with $v_{ij}=0$ for $j-i>1$.

More generally, let P be a set of pairs $(i<j)$, $1\leqslant i,j\leqslant N$, and consider the N-particle Hamiltonian

$$H=\sum_{i=1}^{N}\frac{m_i}{2}\dot{x}_i^2+\sum_{(i<j)\in P}v_{ij}(x_j-x_i), \tag{39}$$

where the pair potentials $v_{ij}(r)$, $(i<j)\in P$, satisfy the repulsivity requirements of Sec. I. In the system (39) the masses m_i and m_j interact by the directed repulsive force $f_{ij}=-v_{ij}'(x_j-x_i)$ if $(i<j)\in P$ and they do not interact if the pair (i,j) does not belong to P.

Theorem 4: Consider the system of particles with the Hamiltonian (39), and assume that P contains the set

$$P_0=\{(i,i+1),\quad 1\leqslant i\leqslant N-1\}.$$

If the forces $f_{ij}(r)$ have the same rate of decay for all $(i,j)\in P$, or if the potentials $v_{ij}(r)$ and $v_{kl}(r)$ satisfy (32) for all pairs (i,j) and (k,l) from P, then the asymptotic velocities $\dot{x}_j(\infty)$ for any motion satisfy the regularity condition (5).

Proof: As in the proof of Theorem 2, we consider the cone C_P spanned by the vectors e_j-e_i for $(i<j)\in P$. Obviously, C_P belongs to the cone C spanned by all vectors e_j-e_i, $i<j$. Since the cone corresponding to P_0 is equal to C, the inclusion $P_0\subset P$ implies $C_P=C$. The rest of the proof of Theorem 2 goes through and proves the assertion.

Corollary 5: Consider the nearest neighbor Hamiltonian

$$H=\sum_{i=1}^{N}\frac{m_i}{2}\dot{x}_i^2+\sum_{i=1}^{N-1}c_iv(x_{i+1}-x_i), \tag{40}$$

with $c_i>0$. The asymptotic velocities of all motions of the system (4) are regular $\{\dot{x}_1(\infty)<\cdots<\dot{x}_N(\infty)\}$ if and only if v is a repulsive potential.

Proof: Follows from Theorem 4 the same way Corollary 3 follows from Theorem 2.

The situation is more complicated for the central repulsive forces (35), if we allow $v_{ij}=0$ for some pairs $(i<j)$. Let P be a set of pairs $(i<j)$ and consider the Hamiltonian

$$H=\frac{1}{2}\sum_{i=1}^{N}m_i\dot{x}_i^2+\sum_{(i<j)\in P}v_{ij}(|x_i-x_j|). \tag{41}$$

We assume that all pair potentials $v_{ij}(r)$, $(i<j)\in P$, are repulsive with $v_{ij}(0)=\infty$. This means that the configuration space X is divided into $N!$ regions,

$$X_w=\{x_{i_1}<\cdots<x_{i_N}\},$$

where $w=(i_1,\ldots,i_N)$ are permutations of $(1,2,\ldots,N)$. For any permutation w, we set $P_w=\{(i_1,i_2),\ldots,(i_{N-1},i_N)\}$.

Theorem 5: Consider the system of particles with the Hamiltonian (41). Assume that all forces have the same rate of decay at infinity or that the pair potentials satisfy condition (32).

If P contains P_w then the region X_w is invariant under the dynamics, and for any motion in X_w the asymptotic velocities satisfy

$$\dot{x}_{i_1}(\infty)<\cdots<\dot{x}_{i_N}(\infty). \tag{42}$$

In particular, the asymptotic velocities for any motion in X_w are distinct.

Proof: For a domain X_w we relabel the particles in an obvious way, and obtain the Hamiltonian (39), while P_w becomes P_0. Now we apply Theorem 4 and obtain the assertion.

Discussion: If P contains the sets P_w for all w, then P contains all pairs $(i<j)$; hence Theorem 3 applies. On the other hand, if P does not contain P_w for some w, then we cannot guarantee that the motions starting in X_w have regular asymptotic velocities.

Example: Consider the system of three particles with the Hamiltonian

$$H = \tfrac{1}{2}(\dot{x}_1^2 + \dot{x}_2^2 + \dot{x}_3^2) + v(|x_1 - x_2|) + v(|x_2 - x_3|), \tag{43}$$

where $v(r)$ is a repulsive potential. By Theorem 5, in the region $\{x_1 < x_2 < x_3\}$ we have $\dot{x}_1(\infty) < \dot{x}_2(\infty) < \dot{x}_3(\infty)$. But in the region $\{x_1 < x_3 < x_2\}$ we can only assert that the asymptotic velocities satisfy $\dot{x}_1(\infty) \leqslant \dot{x}_2(\infty)$ and $\dot{x}_3(\infty) \leqslant \dot{x}_2(\infty)$.

ACKNOWLEDGMENT

The author was partially supported by NSF Grant No. DMS 86-00350.

[1]G. Galperin, "On systems of locally interacting and repelling particles moving in space," Trans. Mos. Math. Soc. **43**, 159 (1981).
[2]L. N. Vaserstein, "On systems of particles with finite range and/or repulsive interactions," Commun. Math. Phys. **69**, 31 (1979).
[3]J. Moser, "The scattering problem for some particle systems on the line," *Springer Lecture Notes in Mathematics*, Vol. 597 (Springer, Berlin, 1977), pp. 441–463.
[4]E. Gutkin, "Regularity of the scattering trajectories in classical mechanics," Commun. Math. Phys. **119**, 1 (1988).
[5]A. Hubacher, "Classical scattering theory in one dimension," preprint.
[6]J. Moser, "Three integrable Hamiltonian systems connected with isospectral deformations," Adv. Math. **16**, 197 (1975).
[7]G. Galperin, "Asymptotic behavior of particle motion under repulsive forces," Commun. Math. Phys. **84**, 547 (1982).
[8]E. Gutkin, "Asymptotics of trajectories for cone potentials," Physica D **17**, 235 (1985).

Explicitly time-dependent constants/symmetries of the higher-order KP equations

K. M. Case and A. Monge
The Rockefeller University, 1230 York Avenue, New York, New York 10021

(Received 8 November 1988; accepted for publication 18 January 1989)

For higher-order Kadomtsev–Petviashvili (KP) equations, the existence of explicitly time-dependent constants of motion is discussed. Generalizing the results for the ordinary KP equation, constants $T_{n,l}^{(m)}$ polynomial in t are constructed for each lth order KP equation, for all $n \geqslant m\ (l-1)-1$.

I. INTRODUCTION

A lot of attention has been recently paid to the Kadomtsev–Petviashvili (KP) equation[1] in the study of several different physical problems.[2]

It has been shown that the KP equation is a completely integrable Hamiltonian system,[3,4] i.e., possessing an infinite number of constants of motion in involution. According to a theorem by Case,[5] these constants are generators of symmetries of the equation. Later on, it was found that the KP equation has an infinite number of "new" constants that are polynomials in t. More precisely, Chen *et al.*[6,7] derived a set of explicitly time-dependent symmetries for both integrable nonlinear evolution equations with one spatial dimension (for example, the Korteveg–de Vries equation) and with two spatial dimensions (namely the KP equation). In one spatial dimension, these new symmetries are in general symmetries of the equation without being Hamiltonian symmetries, in the sense that no corresponding constants of motion can be found. Instead, for the KP equation, the new symmetries are related to explicitly time-dependent constants of motion, and in Ref. 8 it has been shown that for all $n \geqslant 0$, there are $n+2$ constants and one pseudoconstant (whose appearance is due to the existence of Casimir invariants, and which does not generate a symmetry).

The fact that the infinite set of the "old" (time-independent) constants are in involution means that the KP equation is but one of a hierarchy of equations, which is simply obtained by considering these constants as Hamiltonians.

In the last few years, several authors have investigated a KP hierarchy resulting from the compatibility conditions of a hierarchy of systems of linear equations.[9] This hierarchy consists of nonlinear partial differential equations in an increasing number of independent variables.

Here we want to show that for each member of the first hierarchy there is an infinite set of new symmetries polynomial in t, and how these can be constructed. We will proceed by induction, moving up from the already known result about symmetries linear in t of the higher-order KP equations.[10]

II. KP EQUATION AND ITS GENERAL PROPERTIES

We take the KP equation in the form

$$q_t = \partial_x^{-1}\partial_y^2 q - \partial_x[3q^2 + q_{xx}]. \tag{1}$$

This equation can be put in Hamiltonian form,

$$q_t = [q,H], \tag{2}$$

by taking as Hamiltonian

$$H = \int\left\{\frac{(\partial_x^{-1}\partial_y q)^2}{2} + \frac{q_x^2}{2} - q^3\right\}, \tag{3}$$

with the Poisson bracket

$$[F,G] = \int \frac{\partial F}{\partial q}\partial_x \frac{\partial G}{\partial q}. \tag{4}$$

Here, and in the following, we make the convention that

$$\int\left\{\cdots\right\} = \int_{-\infty}^{+\infty}\int_{-\infty}^{+\infty}\left\{\cdots\right\}dx\,dy. \tag{5}$$

If we write Eq. (1) in the form

$$q_t = K(q), \tag{6}$$

then a solution σ of the linearized equation

$$\sigma_t = K'(\sigma), \quad \left(K'(\sigma) = \frac{d}{d\epsilon}K(q+\epsilon\sigma)|_{\epsilon=0}\right), \tag{7}$$

is called a symmetry of the equation. We will say that this is a Hamiltonian symmetry if a functional C can be found such that σ is of the form

$$\sigma = [q,C] \equiv \partial_x \frac{\delta C}{\delta q}. \tag{8}$$

Then[11] C is a constant of motion.

The explicitly time-dependent symmetries found by Chen *et al.*,[7] together with the previously known ones, constitute an infinite-dimensional Lie algebra. As mentioned above, the new symmetries are Hamiltonian.[12] Therefore, we can consider the Lie algebra of the corresponding constants with the Poisson bracket as Lie product, and commutation relations[8]

$$[C_n^{(m)}, C_s^{(r)}] = \frac{r(n+1) - m(s+1)}{3} C_{n+s-2}^{(m+r-1)}, \tag{9}$$

where the general form of $C_n^{(m)}$ is

$$C_n^{(m)} = \sum_{k=0}^{m}\binom{m}{k} t^{m-k} J_n^{(k)}, \tag{10}$$

as given in Ref. 8, where also the $J_n^{(k)}$ are constructed and some of them found explicitly. For each $n \geqslant 0$ there are $n+2$ constants $C_n^{(m)}$, $m = 0,1,\ldots,n+1$, and one pseudoconstant $C_n^{(n+2)}$. The constants $C_n^{(m)}$ are obtained from $C_n^{(n+2)}$ by repeated differentiation, and the pseudoconstant is constructed from the Casimir invariant

 0022-2488/89/061250-04$02.50

$$J_n^{(n+2)} = -\frac{1}{18}\int y^{n+2} q, \tag{11}$$

by taking repeated Poisson brackets with the Hamiltonian. For $m=0$ we have the time-independent constants. The first few of them are

$$C_0^{(0)} \equiv J_0^{(0)} = \int \frac{q^2}{6}, \quad C_1^{(0)} \equiv J_1^{(0)} = \frac{1}{3}\int q\,\partial_x^{-1}\partial_y q,$$

$$C_2^{(0)} \equiv J_2^{(0)} = H,$$

$$C_3^{(0)} \equiv J_3^{(0)} = \int \left\{ 2q_x q_y + \frac{2}{3} q(\partial_x^{-1}\partial_y)^3 q - 4q^2 \partial_x^{-1}\partial_y q \right\}.$$

The fact that the Poisson bracket (4) in the case of the KP equation is singular, i.e., that there are Casimir invariants, means that formula (9) is not correct for arbitrary indices. In the following, we will assume that m, n, r, and s are chosen in such a way that $C_n^{(m)}$, $C_s^{(r)}$, and $C_{n+s-2}^{(m+r-1)}$ are constants, i.e., that for fixed n and s, m and r satisfy

$$\begin{aligned} &m \leqslant n+1, \quad r \leqslant s+1, \\ &m+r \leqslant n+s. \end{aligned} \tag{12}$$

Then formula (9) is correct (see Appendix A). Moreover, if we look at the terms in (9) which are not explicitly time dependent, we see that similar commutation relations hold for the $J_n^{(m)}$'s:

$$[J_n^{(m)}, J_s^{(r)}] = \frac{r(n+1) - m(s+1)}{3} J_{n+s-2}^{(m+r-1)}. \tag{13}$$

III. THE HIGHER-ORDER KP EQUATIONS

The infinite hierarchy of KP equations is given by

$$q_t = K_l(q), \quad l = 0,1,..., \tag{14}$$

with

$$K_l = \partial_x \frac{\delta C_l^{(0)}}{\delta q} \equiv \partial_x \frac{\delta J_l^{(0)}}{\delta q}. \tag{15}$$

These equations are Hamiltonian with Hamiltonians $H_l = J_l^{(0)}$, and they all are completely integrable.

We now want to see whether there are constants polynomial in t for Eq. (14) for arbitrary l. For $l=0$ and $l=1$, we have two degenerate cases (corresponding to the equations $q_t = q_x/3$ and $q_t = 2q_y/3$, respectively) that will be treated separately in Appendix B. Here we will assume $l \geqslant 2$. Write

$$T_{n,l}^{(m)} = \sum_{k=0}^{m} \binom{m}{k} t^{m-k} L_{n,l}^{(k)}, \tag{16}$$

with

$$L_{n,l}^{(0)} = J_n^{(0)}. \tag{17}$$

For fixed l, $T_{n,l}^{(m)}$ will be constant of the lth order KP equation if and only if

$$L_{n,l}^{(k-1)} = -(1/k)[L_{n,l}^{(k)}, J_l^{(0)}], \quad k = 1,...,m. \tag{18}$$

This is equivalent to the requirement that

$$\delta q = [q, T_{n,l}^{(m)}] \tag{19}$$

be a solution of

$$\delta q_t = K_l'(\delta q). \tag{20}$$

With δq chosen as in Eq. (19) we have

$$\frac{d}{dt}\delta q = \sum_{k=1}^{m} (m-k+1)\binom{m}{k-1} t^{m-k}[q, L_{n,l}^{(k-1)}] + [[q, T_{n,l}^{(m)}], J_l^{(0)}]. \tag{21}$$

By the Jacobi identity, the second term on the right-hand side becomes

$$[[q, T_{n,l}^{(m)}], J_l^{(0)}] = -[[T_{n,l}^{(m)}, J_l^{(0)}], q] - [[J_l^{(0)}, q], T_{n,l}^{(m)}]. \tag{22}$$

But, using the fact that $[L_{n,l}^{(0)}, J_l^{(0)}] = 0$ and Eq. (18), we have

$$[[T_{n,l}^{(m)}, J_l^{(0)}], q] = -\sum_{k=1}^{m} \binom{m}{k} k t^{m-k}[L_{n,l}^{(k-1)}, q]. \tag{23}$$

Equation (21) finally reads

$$\frac{d}{dt}\delta q = [[q, J_l^{(0)}], T_{n,l}^{(m)}]. \tag{24}$$

Since δq is of the form (8), from the theorem proved in Ref. 12 we know that

$$K_l'(\delta q) = [[q, J_l^{(0)}], T_{n,l}^{(m)}], \tag{25}$$

and by comparing Eqs. (24) and (25) we conclude that δq is a solution of the linearized lth order KP Eq. (20).

We now construct $T_{n,l}^{(m)}$. For $m=0$, we have [see Eq. (17)]

$$T_{n,l}^{(0)} \equiv L_{n,l}^{(0)} = J_n^{(0)}. \tag{26}$$

For $m=1$, the construction is given in Ref. 10, where it was found

$$T_{n,l}^{(1)} = tL_{n,l}^{(0)} + L_{n,l}^{(1)}, \tag{27}$$

with

$$L_{n,l}^{(1)} = (3/(l+1))J_{n-l+2}^{(1)}, \tag{28}$$

for fixed l and for all $n \geqslant -2$. We will discuss this last condition on n later on.

For the generic $m=k$, we have to look for $L_{n,l}^{(k)}$. Assume that

$$L_{n,l}^{(k-1)} = \frac{3^{k-1}}{(l+1)^{k-1}} J_{n-(k-1)(l-2)}^{(k-1)}. \tag{29}$$

Then, using Eq. (18) we obtain

$$J_{n-(k-1)(l-2)}^{(k-1)} = \left[J_l^{(0)}, \frac{(l+1)^{k-1}}{k\,3^{k-1}} L_{n,l}^{(k)} \right]. \tag{30}$$

From the commutation relations (13) we have

$$[J_l^{(0)}, J_n^{(k)}] = (k(l+1)/3) J_{l+n-2}^{(k-1)}. \tag{31}$$

Let $n \to n - k(l-2)$. Equation (31) becomes

$$J_{n-(k-1)(l-2)}^{(k-1)} = [J_l^{(0)}, (3/k(l+1)) J_{n-k(l-2)}^{(k)}]. \tag{32}$$

Comparing Eqs. (30) and (32), we find

$$L_{n,l}^{(k)} = \frac{3^k}{(l+1)^k} J_{n-k(l-2)}^{(k)}, \tag{33}$$

which is exactly Eq. (29) with $k-1$ replaced by k.

In this way, we can construct for each lth order KP

equation ($l \geqslant 2$) the polynomial constants $T_{n,l}^{(m)}$. We must notice though that our procedure is correct only if no Casimir invariants appear in Eq. (31) [or, that is the same, in Eq. (32)]. From Eq. (32) with $k = m$, and recalling that $J_s^{(r)}$ is a Casimir invariant when $r = s + 2$ (while it is not for $r \leqslant s + 1$), we see that for $L_{n,l}^{(m)}$ to be well defined we have to require the condition $m \leqslant n - m(l-2) + 1$, i.e.,

$$n \geqslant m(l-1) - 1. \tag{34}$$

For fixed n and l, the largest integer m for which (34) is satisfied is

$$\left[\frac{n+1}{l-1}\right], \tag{35}$$

where $[\cdots]$ indicates the integer part.

We therefore conclude that for each l th order KP equation ($l \geqslant 2$), for each $n \geqslant 0$ there are $[(n+1)/(l-1)] + 1$ constants, for $m = 0,1,\ldots, [(n+1)/(l-1)]$. As l increases, for fixed n, we find fewer and fewer constants.

The possibility of constructing pseudoconstants for l th order KP equation (and then obtaining the real constants by repeated differentiation) is discussed in Appendix C.

APPENDIX A: COMMUTATION RELATIONS FOR THE KP EQUATION

Let us prove that the commutation relations (9) hold when $C_n^{(m)}$, $C_s^{(r)}$, and $C_{n+s-2}^{(m+r-1)}$ are all three constants. Using (10) we have

$$\begin{aligned}[C_n^{(m)}, C_s^{(r)}] &= t^{m+r}[J_n^{(0)}, J_s^{(0)}] \\ &\quad + t^{m+r-1}\{m[J_n^{(1)}, J_s^{(0)}] + r[J_n^{(0)}, J_s^{(1)}]\} \\ &\quad + \text{terms of lower order in } t.\end{aligned} \tag{A1}$$

We now apply the commutation relations found by Chen *et al.*,[7]

$$[J_n^{(0)}, J_s^{(0)}] = 0, \tag{A2}$$

$$[J_n^{(0)}, J_s^{(1)}] = [(n+1)/3]\, J_{n+s-2}^{(0)}, \tag{A3}$$

and obtain

$$[C_n^{(m)}, C_s^{(r)}] = t^{m+r-1}\left\{\frac{r(n+1) - m(s+1)}{3}\right\} J_{n+s-2}^{(0)} + \text{terms of lower order in } t. \tag{A4}$$

Assuming that there exists a unique constant $C_{n+s-2}^{(m+r-1)}$, whose leading term is $t^{m+r-1} J_{n+s-2}^{(0)}$, we conclude that the right-hand side of (A4) must be equal to $\{r(n+1) - m(s+1)\}/3$ times $C_{n+s-2}^{(m+r-1)}$, that is

$$[C_n^{(m)}, C_s^{(r)}] = \left\{\frac{r(n+1) - m(s+1)}{3}\right\} C_{n+s-2}^{(m+r-1)}. \tag{A5}$$

An interesting question is now to see whether Eq. (A5) is correct when either $C_n^{(m)}$ or $C_{n+s-2}^{(m+r-1)}$ is a pseudoconstant.

If $C_{n+s-2}^{(m+r-1)}$ is to be a pseudoconstant, then we can show with an example that (A5) cannot be valid in general. From Ref. 8 we have

$$\begin{aligned} C_0^{(1)} &= t J_0^{(0)} + J_0^{(1)}, \\ J_0^{(0)} &= \int \frac{q^2}{6}, \quad J_0^{(1)} = -\frac{1}{18}\int xq, \end{aligned} \tag{A6}$$

and

$$\begin{aligned} C_2^{(2)} &= t^2 J_2^{(0)} + 2t J_2^{(1)} + J_2^{(2)}, \\ J_2^{(0)} &= H, \quad J_2^{(1)} = \int \left\{\frac{q^2 x}{6} + \frac{q\, y(\partial_x^{-1}\partial_y q)}{3}\right\}, \\ J_2^{(2)} &= -\frac{1}{18}\int x^2 q + \frac{1}{6}\int y^2 q^2. \end{aligned} \tag{A7}$$

Explicit calculation shows that

$$[C_0^{(1)}, C_2^{(2)}] = -\tfrac{1}{3}\{t^2 J_0^{(0)} + 2t J_0^{(1)} + A\} \tag{A8}$$

where A (the time-independent term) is formally

$$A = -\frac{1}{18}\int y^2 q + \frac{1}{108}\int x^2. \tag{A9}$$

We recognize the first term to be the Casimir invariant $J_0^{(2)}$, but we also have a meaningless term. Meaningless terms will arise whenever we take the Poisson bracket $[F, G]$ of two functionals of the form

$$F = \int f(x,y)q, \quad G = \int g(x,y)q. \tag{A10}$$

It is easily seen that functionals of this form are present among the $J_n^{(m)}$'s beside the Casimir invariants $J_n^{(n+2)}$. Let us recall that[8]

$$\begin{aligned} J_n^{(n+1)} &= -(1/(n+2))[J_n^{(n+2)}, H], \\ J_n^{(n)} &= -(1/(n+1))[J_n^{(n+1)}, H]. \end{aligned} \tag{A11}$$

Then we have

$$J_n^{(n+1)} = -\frac{n+1}{18}\int y^n xq, \tag{A12}$$

and $J_n^{(n)}$ contains a term proportional to $\int y^{n-2}x^2 q$.

A problem similar to the one just discussed above is encountered if, in Eq. (A5), $C_n^{(m)}$ is a pseudoconstant. Of interest for the question of the existence of pseudoconstants for higher-order KP equations (see Appendix C) is whether the relation

$$[C_n^{(n+2)}, C_s^{(r)}] = \frac{r(n+1) - (n+2)(s+1)}{3} C_{n+s-2}^{(n+r+1)}, \quad r \leqslant s+1, \tag{A13}$$

holds, and in particular, whether it is true that

$$[J_n^{(n+2)}, J_s^{(0)}] = -\frac{(n+2)(s+1)}{3} J_{n+s-2}^{(n+1)}. \tag{A14}$$

We have not been able to prove (A14), and we can only say that $[J_n^{(n+2)}, J_s^{(0)}]$ does not produce meaningless terms. In fact, from (A11) we see that $J_n^{(n+2-k)}$ contains y^{n+2-2k}, and therefore $J_s^{(0)}$ does not contain terms of the form (A10).

APPENDIX B: DEGENERATE CASES $l = 0$ AND $l = 1$

Let us consider the two degenerate cases $l = 0$ and $l = 1$. Equation (14) becomes

$$q_t = \tfrac{1}{3} q_x \tag{B1}$$

and

$$q_t = \tfrac{2}{3} q_y, \tag{B2}$$

respectively.

Let us define for $l = 0$,

$$L^{(k)}_{n,0} = 3^k J^{(k)}_{n+2k}, \tag{B3}$$

and for $l = 1$,

$$L^{(k)}_{n,1} = (\tfrac{3}{2})^k L^{(k)}_{n+k}. \tag{B4}$$

For each n, $L^{(k)}_{n,0}$ and $L^{(k)}_{n,1}$ are never Casimir invariants for arbitrary $k \geqslant 0$. Therefore, in these cases we find for each $n \geqslant 0$ an infinite number of constants $T^{(m)}_{n,l}$, $m = 0,1,\ldots$.

As a consequence of this result, we can remove the previously imposed condition $l \geqslant 2$, and apply Eq. (33) for all $l \geqslant 0$, and for $n \geqslant k(l-1) - 1$.

APPENDIX C: PSEUDOCONSTANTS FOR HIGHER-ORDER KP EQUATIONS

We discuss here the possibility of constructing pseudoconstants for higher-order KP equations.

Suppose that $l(\geqslant 2)$ and n satisfy the condition

$$l - \left\{\frac{n+1}{l-1}\right\}(l-1) = 2, \tag{C1}$$

where $\{\cdots\}$ indicates the fractional part. Then for $m = [(n+1)/(l-1)] + 1$, define

$$L^{(m)}_{n,l} = (3^m/(l+1)^m) J^{(m)}_{n-m(l-2)}. \tag{C2}$$

$L^{(m)}_{n,l}$ is a Casimir invariant since from (C1) it follows that

$$m = n - m(l-2) + 2, \tag{C3}$$

and so $J^{(m)}_{n-m(l-2)}$ is a Casimir invariant. For $m = [(n+1)/(l-1)] + 1$, $T^{(m)}_{n,l}$ contains a Casimir invariant and is therefore a pseudoconstant.

Let us now suppose that Eq. (A14) holds. Then

$$[L^{(m)}_{n,l}, H_l] = -mL^{(m-1)}_{n,l}, \tag{C4}$$

and $L^{(k)}_{n,l}$, for $k \leqslant m-1$, can be found from $L^{(m)}_{n,l}$ by taking repeated Poisson brackets with the Hamiltonian H_l.

Explicit formulae can be given. Set $\bar{m} = [(n+1)/(l-1)] + 1$, and define $[L^{(\bar{m})}_{n,l}, H_l]_k$ by $[L^{(\bar{m})}_{n,l}, H_l]_0 \equiv L^{(\bar{m})}_{n,l}$, and the recursion relation

$$[L^{(\bar{m})}_{n,l}, H_l]_k = [[L^{(\bar{m})}_{n,l}, H_l]_{k-1}, H_l], \tag{C5}$$

with $k = 1,2,\ldots,\bar{m}$. Then

$$T^{(\bar{m})}_{n,l} = \sum_{s=0}^{\bar{m}} \frac{(-1)^s}{s!} t^s [L^{(\bar{m})}_{n,l}, H_l]_s, \tag{C6}$$

and

$$T^{(\bar{m}-r)}_{n,l} = \frac{(-1)^r}{\bar{m}(\bar{m}+1)\cdots(\bar{m}+1-r)} \times \sum_{s=0}^{\bar{m}-r} \frac{(-1)^s}{s!} t^s [L^{(\bar{m})}_{n,l}, H_l]_{s+r} \tag{C7}$$

for $r = 1,2,\ldots,\bar{m}$.

[1] B. B. Kadomtsev and V. I. Petviashvili, Sov. Phys. Dokl. **15**, 539 (1970).
[2] For example, a new radical approach to string theory involving the KP equation has been discussed by J. H. Schwarz, "Some Hot Topics in Superstring Theory," Caltech Preprint CALT-68-1481, 1988.
[3] V. E. Zakharov and E. I. Schulman, Physica D **1**, 192 (1980).
[4] J. E. Lin and H. H. Chen, Phys. Lett. A **89**, 163 (1982).
[5] K. M. Case, Phys. Rev. Lett. **40**, 351 (1978).
[6] H. H. Chen, Y. C. Lee, and J. E. Liu, in *Advances in Nonlinear Waves*, edited by L. Debnath (Pitman, London, 1985), Vol. II.
[7] H. H. Chen, Y. C. Lee, and J. E. Liu, Physica D **9**, 439 (1983).
[8] K. M. Case, Proc. Natl. Acad. Sci. USA **82**, 5242 (1985).
[9] E. Date, M. Jimbo, M. Kashiwara, and T. Miwa, in *Proceedings of RIMS Symposium on Non-linear Integrable Systems—Classical Theory and Quantum Theory*, edited by M. Jimbo and T. Miwa (World Scientific, Singapore, 1983). For a discussion of the conservation laws of this KP hierarchy and its connections to string theory, see, for example: Yu. I. Manin and A. O. Radul, Commun. Math. Phys. **8**, 65 (1985); N. Kawamoto, Y. Namikawa, A. Tsuchiya, and Y. Yamada, *ibid.* **116**, 247 (1988); G. Gilbert, *ibid.* **117**, 331 (1988).
[10] K. M. Case, J. Math. Phys. **26**, 1158 (1985).
[11] K. M. Case, Phys. Rev. Lett. **55**, 445 (1985).
[12] K. M. Case, Proc. Natl. Acad. Sci. USA **81**, 5093 (1984).

A bidirectional traveling plane wave representation of exact solutions of the scalar wave equation

Ioannis M. Besieris and Amr M. Shaarawi
Department of Electrical Engineering, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

Richard W. Ziolkowski
Engineering Research Division, Lawrence Livermore National Laboratory, P.O. Box 5504, L-156, Livermore, California 94550

(Received 6 June 1988; accepted for publication 28 December 1988)

A new decomposition of exact solutions to the scalar wave equation into bidirectional, forward and backward, traveling plane wave solutions is described. The resulting representation is a natural basis for synthesizing pulse solutions that can be tailored to give directed energy transfer in space. The development of known free-space solutions, such as the focus wave modes, the electromagnetic directed energy pulse trains, the spinor splash pulses, and the Bessel beams, in terms of this decomposition will be given. The efficacy of this representation in geometries with boundaries, such as a propagation in a circular waveguide, will also be demonstrated.

I. INTRODUCTION

The possibility of solutions of the wave equation that describe localized, slowly decaying transmission of energy in space-time has been suggested by several groups in recent years. These include efforts on "focus wave modes",[1–5] "EDEPT's,"[6–9] "splash modes",[10,11] "EM missiles",[12–16] "Bessel beams",[17–22] "EM bullets"[23,24] and "transient beams."[8,25–29] Much of this work was actually motivated by the pioneering work of Brittingham.[1] It has been recently discovered that these original focus wave modes represent Gaussian beams that translate through space with only local deformations and are the fundamental modes of a class of solutions that describe fields that originate from moving complex sources.[2] In particular, the scalar wave equation in real space, viz.,

$$[\partial_t^2 - \nabla^2]\Psi(\mathbf{r},t) = 0\,, \tag{1.1}$$

with a wave speed normalized to unity, has as an exact solution, the moving, modified Gaussian pulse

$$\Psi_\beta(\mathbf{r},t) = e^{i\beta(z+t)}(e^{-\beta\rho^2/V}/4\pi V)\,. \tag{1.2}$$

The complex variance $1/V = 1/A - i/R$ yields the beam spread $A = a_1 + \zeta^2/a_1$, the phase front curvature $R = \zeta + a_1^2/\zeta$, and beam waist $w = (A/\beta)^{1/2}$. Here, $\zeta = z - t$ and ρ denotes the radial cylindrical coordinate. The fundamental pulse (1.2) describes a Gaussian beam that translates through space-time with only local variations. It represents a generalization of earlier work by Deschamps[30] and Felsen[31] describing Gaussian beams as fields radiated from stationary complex source points.

As discussed in Ref. 9, the fundamental Gaussian pulse has either a plane wave or a particlelike character depending on whether β is small or large. Moreover, for all β it shares with the plane wave the property of having infinite energy. However, as with the plane waves, this is not to be considered as a drawback *per se*. The above solution procedure has introduced an added degree of freedom into the solution through the variable β that can be exploited. As shown in Refs. 2 and 5–9, fundamental Gaussian pulse fields, corresponding to different values of β, can be used as basis functions to represent new transient solutions of Eq. (1.1). In particular, the general electromagnetic directed energy pulse train (EDEPT) solution

$$\Psi(\mathbf{r},t) = \int_0^\infty d\beta\,\Psi_\beta(\mathbf{r},t)F(\beta) = \frac{1}{4\pi[a_1 + i(z-t)]}\int_0^\infty d\beta\,F(\beta)e^{-\beta s(\rho,z,t)}\,, \tag{1.3}$$

where

$$s(\rho,z,t) = \rho^2/[a_1 + i(z-t)] - i(z+t)\,, \tag{1.4}$$

is an exact source-free solution of the wave equation. This representation, in contrast to a plane wave decomposition, utilizes basis functions that are more localized in space and hence, by their very nature, are better suited to describe the directed transfer of electromagnetic energy in space. The resulting pulses have finite energy if the function $F(\beta)\beta^{-1/2}$ is square integrable.[9]

As reported in Ref. 2, the superposition (1.3), with the "complex traveling center wave" basis functions, has an inverse. The functions

$$\Phi_\beta(\rho,\zeta,\eta) = 8\sqrt{\pi}\,e^{-(\zeta/4\beta a_1)^2}\Psi_\beta(\rho,\zeta,\eta)\,, \tag{1.5}$$

with $\eta = z + t$, are orthogonal to the Ψ_β. This means these basis functions satisfy the completeness relation

$$\int_{-\infty}^{+\infty} d\eta \int_{-\infty}^{+\infty} d\zeta \int_0^\infty d\rho\,\rho\Phi_\beta^*(\rho,\zeta,\eta)\Psi_{\beta'}(\rho,\zeta,\eta) = \delta(\beta - \beta')\,, \tag{1.6}$$

where Φ_β^* is the complex conjugate of Φ_β. Hence an inversion of the superposition (1.3) exists.

Clearly, different spectra $F(\beta)$ in Eq. (1.3) lead to different wave equation solutions, and hence, to different solutions of Maxwell's equations. Many interesting solutions of the wave equation can be created by simply referring to a

 0022-2488/89/061254-16$02.50

Laplace transform table. One particular interesting spectrum selection, recognized by Ziolkowski,[6] is the "modified power spectrum" (MPS)

$$F(\beta) = [p/\Gamma(q)]\,(p\beta - b)^{q-1} e^{-[(p\beta - b)a_2]}, \quad \beta > b/p,$$
$$= 0, \quad b/p > \beta \geqslant 0. \tag{1.7}$$

It is so named because it is derived from the power spectrum $F(\beta) = \beta^{q-1}\exp(-\beta a_2)$ by a scaling and a truncation. This choice of spectrum leads to the MPS pulse

$$\Psi(\mathbf{r},t) = \frac{1}{4\pi(a_1 + i\zeta)} \frac{e^{-bs/p}}{[a_2 + s/p]^q}. \tag{1.8}$$

Solutions to Maxwell's equations follow naturally from these scalar wave equation solutions using a Hertz potential formulation.

The MPS pulse, for example, can be optimized so that it is localized near the direction of propagation and its original amplitude can be recovered out to extremely large distances from its initial location. This is demonstrated in Fig. 1, which shows surface plots and the corresponding contours plots of the electromagnetic energy density U of a TE electromagnetic MPS pulse relative to the pulse center locations at $z = 0.0$ km and $z = 9.42\times 10^9$ km. The MPS parameters are $a_2 = 1.0$ m, $q = 1.0$, $b = 1.0\times 10^{14}$ m^{-1}, $p = 6.0\times 10^{15}$, and $a_1 = 1.0\times 10^{-2}$ m. The energy density U is normalized to its maximum value at $t = 0$. The transverse space coordinate ρ is measured in meters; the longitudinal space coordinate $\zeta = z - t$ is the distance in meters along the direction of propagation away from the pulse center $z = ct$. These results definitively show the localization of the field near the direction of propagation over very large distances.

The MPS pulses are being characterized further and potential launching mechanisms are under investigation. However, it was recognized by Besieris and Shaarawi[32] that the representation (1.3) and its inverse has a generalization that can be exploited to explain these and other localized, slowly decaying solutions in a single framework. This new representation is the main purpose of this paper. It is based on a decomposition of exact solutions of the scalar wave equation into bidirectional, forward and backward, traveling plane wave solutions. The resulting representation is a natural basis for synthesizing pulse solutions. The derivation of this representation from a general operator embedding scheme will be described in Sec. II. The connections between this decomposition and various localized, slowly decaying solutions will be made explicit in Sec. III. In Sec. IV, the bidirectional representation will be extended to other classes of equations, e.g., the Klein–Gordon and the dissipative scalar wave equations that model wave propagation in dispersive and dissipative media, respectively. A specific demonstration of the efficacy of the new representation will be given in connection with an initial-boundary value modeling an infi-

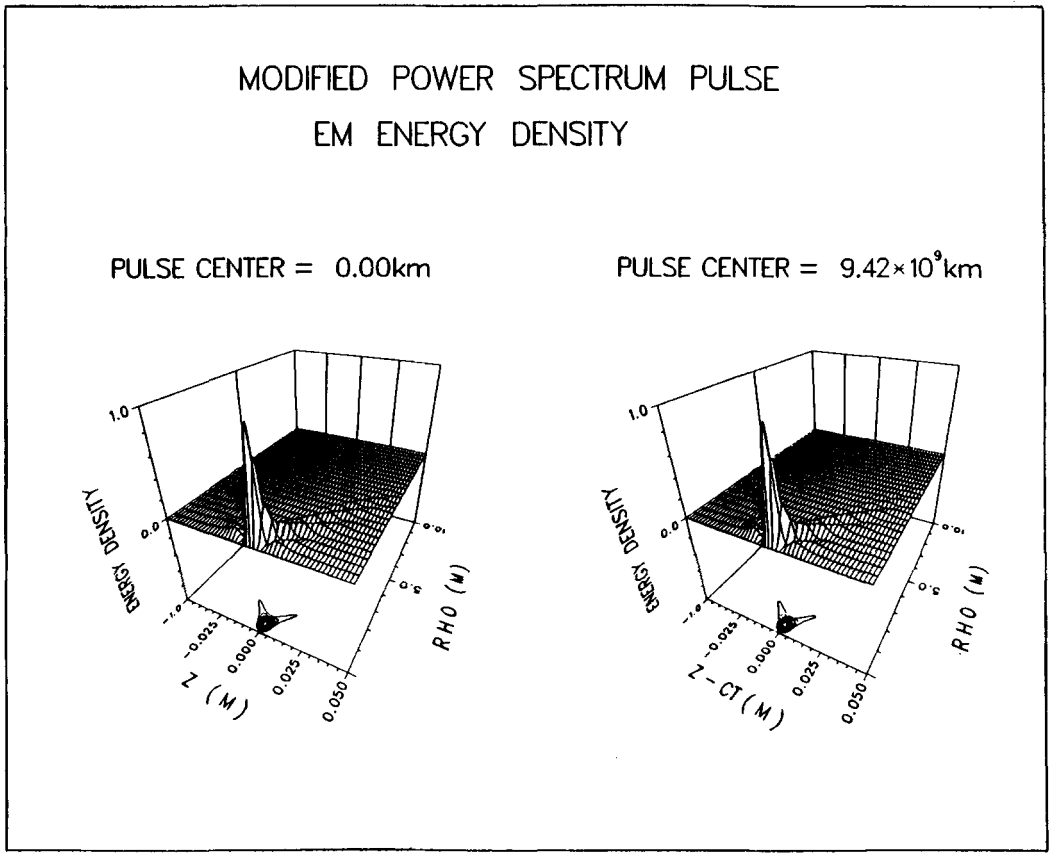


FIG. 1. The field energy of the electromagnetic MPS pulse is shown for the parameters: $a_1 = 1.0$ m, $q = 1.0$, $b = 1.0\times 10^{14}$ m^{-1}, $p = 6.0\times 10^{15}$, and $a_2 = 0.01$ m.

nite waveguide excited by a localized initial pulse. A summary of the results in this paper will be provided in Sec. V.

II. BIDIRECTIONAL PLANE WAVE DECOMPOSITION

The Cauchy problem

$$[\partial_t^2 + \hat{\Omega}(-i\nabla)]u(\mathbf{r},t) = 0, \quad \mathbf{r}\in R^3, \quad t>0, \tag{2.1a}$$

$$u(\mathbf{r},0) = u_0(\mathbf{r}), \quad u_t(\mathbf{r},0) = u_1(\mathbf{r}), \tag{2.1b}$$

where u is a real scalar-valued function and $\hat{\Omega}$ is a positive, self-adjoint, possibly pseudodifferential operator, can be used as a mathematical model for a large number of physical situations.

A Fourier synthesis of the solution to the Cauchy problem (2.1) can be effected as follows[33]:

$$u(\mathbf{r},t) = 2\,\mathrm{Re}\{\Psi(\mathbf{r},t)\}; \tag{2.2a}$$

$$\Psi(\mathbf{r},t) = \frac{1}{(2\pi)^3}\int_{R^3} d\mathbf{k}\, F(\mathbf{k}) e^{-i(\mathbf{k}\cdot\mathbf{r} - \Omega^{1/2}(\mathbf{k})t)}; \tag{2.2b}$$

$$F(\mathbf{k}) = \frac{1}{2}\left[\tilde{u}_0(\mathbf{k}) - i\frac{\tilde{u}_1(\mathbf{k})}{\Omega^{1/2}(\mathbf{k})}\right]. \tag{2.2c}$$

The complex value signal Ψ is generated via a linear superposition of plane waves propagating in the $\mathbf{k}$ direction with phase speeds $\Omega^{1/2}(\mathbf{k})/|\mathbf{k}|$. These plane waves are characterized by wave vectors $\mathbf{k}$ and they are weighted by the Fourier spectrum $F(\mathbf{k})$.

Equation (2.2) constitutes a mathematical solution to the Cauchy problem (2.1). However, for purposes of later comparison, the superposition (2.2b) can be recast in a more general form as follows:

$$\Psi(\mathbf{r},t) = \frac{1}{(2\pi)^4}\int_{R^3} d\mathbf{k}\int_{R^1} d\omega\, \hat{F}(\mathbf{k},\omega) \times e^{-i(\mathbf{k}\cdot\mathbf{r}-\omega t)}\delta[-\omega^2 + \Omega(\mathbf{k})]. \tag{2.3}$$

The spectra entering into Eqs. (2.2b) and (2.3) are linked through the relationship

$$F(\mathbf{k}) = \frac{\hat{F}(\mathbf{k},\Omega(\mathbf{k}))}{4\pi|\Omega^{1/2}(\mathbf{k})|}. \tag{2.4}$$

Conditions can also be specified under which Ψ is square integrable, or, even further, under which the solution $u(\mathbf{r},t)$ of (2.1) is a finite energy signal. There is, however, a basic drawback associated with the Fourier method; namely, that in most cases the integral for Ψ can be computed only approximately by a variety of asymptotic approaches, such as the method of stationary phase/saddle point,[34,35] ray-theoretic techniques[36,37] and phase space methods,[38] or can be carried out numerically. Very few exact analytical solutions to (2.2b) are available, even for the simple, single mode dispersion relationship $\omega\equiv\Omega^{1/2}(\mathbf{k}) = (k^2+\mu^2)^{1/2}$ corresponding to the Klein–Gordon equation.

The Cauchy problem (2.1) will be used in the sequel as a vehicle for presenting a new principle of superposition that provides more freedom and flexibility when dealing with certain classes of solutions, e.g., the EDEPT solutions to the scalar wave equation.

Different types of superpositions are obtained by dividing the operator $L = [\partial_t^2 + \hat{\Omega}(-i\nabla)]$ into parts, each having its own eigenfunctions. A general solution can be constructed from the product of such eigenfunctions, together with a constraint relationship between their eigenvalues. The manner in which the operator L is partitioned determines the form of the final superposition. For example, the Fourier decomposition follows from partitioning L into two parts: $L_1 = \partial_t^2$ and $L_2 = \hat{\Omega}(-i\nabla)$. The superposition (2.3) contains the constraint $\omega = \Omega^{1/2}(\mathbf{k})$ relating the eigenvalues of L_1 and L_2 corresponding to the eigenfunctions $\exp(+i\omega t)$ and $\exp(-i\mathbf{k}\cdot\mathbf{r})$, respectively.

In general, the operator L can be partitioned in many different ways. Consider, for example, the preliminary splitting of the operator $\hat{\Omega}(-i\nabla)$ as follows:

$$\hat{\Omega}(-i\nabla) = \hat{A}(-i\partial_z) + [\hat{\Omega}(-i\nabla) - \hat{A}(-i\partial_z)] \equiv \hat{A}(-i\partial_z) + \hat{B}(-i\nabla_T, -i\partial_z). \tag{2.5}$$

The operator $\hat{A}(-i\partial_z)$, which may or may not be a natural part of $\hat{\Omega}(-i\nabla)$, is assumed to be positive, self-adjoint and the choice of the preferred variable z is arbitrary. By taking the Fourier transform with respect to the transverse components, the complex wave function $\Psi(\mathbf{r},t)$ can be expressed as

$$\Psi(\mathbf{r},t) = \frac{1}{(2\pi)^2}\int_{R^2} d\boldsymbol{\kappa}\, \tilde{\psi}(\boldsymbol{\kappa},z,t) e^{-i\boldsymbol{\kappa}\cdot\boldsymbol{\rho}}, \tag{2.6}$$

with $\tilde{\psi}(\boldsymbol{\kappa},z,t)$ governed by the equation

$$[\partial_t^2 + \hat{A}(-i\partial_z) + \hat{B}(\boldsymbol{\kappa}, -i\partial_z)]\tilde{\psi}(\boldsymbol{\kappa},z,t) = 0. \tag{2.7}$$

The operator $L\equiv\partial_t^2 + \hat{\Omega}(-\boldsymbol{\kappa}, -i\partial_z)$ can now be partitioned as follows:

$$L_1 = \partial_t^2 + \hat{A}(-i\partial_z), \tag{2.8a}$$

$$L_2 = \hat{B}(-\boldsymbol{\kappa}, -i\partial_z). \tag{2.8b}$$

The most natural eigenfunctions of the operator L_1 are given by

$$\psi_e(z,t) = e^{-i\alpha\zeta}e^{i\beta\eta}, \tag{2.9}$$

where ζ and η are defined as follows:

$$\zeta = z - t\,\mathrm{sgn}(\alpha)\alpha^{-1}A^{1/2}(\alpha), \tag{2.10a}$$

$$\eta = z + t\,\mathrm{sgn}(\beta)\beta^{-1}A^{1/2}(\beta). \tag{2.10b}$$

The corresponding eigenvalues, denoted by $\lambda(\alpha,\beta)$, are given explicitly as follows:

$$\lambda(\alpha,\beta) = A(\beta-\alpha) - [A(\alpha) + A(\beta) + 2\,\mathrm{sgn}(\alpha)A^{1/2}(\alpha)\mathrm{sgn}(\beta)A^{1/2}(\beta)]. \tag{2.11}$$

The elementary functions (2.9) consist of products of two plane waves traveling in opposite directions, with wavenumber-dependent phase speeds equal to $\mathrm{sgn}(\alpha)\alpha^{-1}A^{1/2}(\alpha)$ and $\mathrm{sgn}(\beta)\beta^{-1}A^{1/2}(\beta)$, respectively.

The bilinear functions (2.9) are also eigenfunctions of L_2, with corresponding eigenvalues equal to $B(-\boldsymbol{\kappa},\beta-\alpha)$. As a consequence, a linear superposition of the bidirectional elementary solutions ψ_e results in a solution of Eq. (2.7), viz.,

$$\tilde{\psi}(\boldsymbol{\kappa},z,t) = \int_{R^1} d\alpha\int_{R^1} d\beta\, C(\boldsymbol{\kappa},\alpha,\beta) \times e^{-i\alpha\zeta}e^{i\beta\eta}\delta[\lambda(\alpha,\beta) + B(-\boldsymbol{\kappa},\beta-\alpha)], \tag{2.12}$$

where the constraint $\lambda(\alpha,\beta) + B(-\boldsymbol{\kappa},\beta-\alpha) = 0$ is included in the integration. A general solution to Eq. (2.1) can

be obtained by resorting to a transverse Fourier inversion [cf. Eq. (2.6)]; specifically,

$$\Psi(\mathbf{r},t)=\frac{1}{(2\pi)^2}\int_{R^2}d\boldsymbol{\kappa}\,e^{-i\boldsymbol{\kappa}\cdot\boldsymbol{\rho}}\int_{R^1}d\alpha\int_{R^1}d\beta\,C(\boldsymbol{\kappa},\alpha,\beta)\times e^{-i\alpha\zeta}e^{i\beta\eta}\delta[\lambda(\alpha,\beta)+B(-\boldsymbol{\kappa},\beta-\alpha)]\,. \quad (2.13)$$

This representation constitutes a generalization of the three-dimensional Fourier synthesis [cf. Eq. (2.3)]; in the latter, the operator $\hat{A}(-i\partial_z)$ was chosen to be a constant given by the relations

$$\mathrm{sgn}(\alpha)A^{1/2}(\alpha)+\mathrm{sgn}(\beta)A^{1/2}(\beta)=\omega \quad (2.14a)$$

and

$$\alpha-\beta=k_z\,. \quad (2.14b)$$

The main advantage of this decomposition is the introduction of the embedded operator $\hat{A}(-i\partial_z)$. This provides a fresh approach for addressing different classes of problems. At the same time, the flexibility that one can enjoy through a clever choice of $\hat{A}(-i\partial_z)$ may open the way to approach some of the more impenetrable problems.

To clarify these ideas, consider specifically the case of the three-dimensional scalar wave equation for which $\hat{\Omega}(-i\nabla)=-\nabla^2$. The operator L, in this case, assumes the form $L=\partial_t^2-\nabla^2$ and Eq. (2.1a) simplifies to

$$[\partial_t^2-\nabla^2]u(\mathbf{r},t)=0\,. \quad (2.15)$$

In cylindrical coordinates, the Laplacian ∇^2 can be written as follows:

$$\nabla^2=\partial_z^2+\partial_\rho^2+\rho^{-1}\partial_\rho+\rho^{-2}\partial_\phi^2\,.$$

In the usual Fourier decomposition, the operator L is divided into two parts:

$$L_1=-[\partial_z^2+\partial_\rho^2+\rho^{-1}\partial_\rho+\rho^{-2}\partial_\phi^2]\,, \quad (2.16a)$$

$$L_2=\partial_t^2\,. \quad (2.16b)$$

The eigenfunctions of L_1 are $J_n(\kappa\rho)e^{\pm in\phi}e^{\pm ik_z z}$ and $N_n(\kappa\rho)e^{\pm in\phi}e^{\pm ik_z z}$, where $J_n(\kappa\rho)$ and $N_n(\kappa\rho)$ are Bessel functions of the first and second kind, respectively, and the eigenvalues equal $\kappa^2+k_z^2$. The operator L_2 has eigenfunctions $e^{\pm i\omega t}$ with eigenvalues $-\omega^2$. An elementary solution to the scalar wave equation (2.15) can be written as

$$\Psi_e(\mathbf{r},t)=[A_nJ_n(\kappa\rho)+B_nN_n(\kappa\rho)]e^{\pm in\phi}e^{-i(k_z z\pm\omega t)}\,, \quad (2.17a)$$

with the constraint

$$\kappa^2+k_z^2-\omega^2=0\,. \quad (2.17b)$$

Neglecting the terms $N_n(\kappa\rho)$ because of their infinite values at $\rho=0$, one obtains a special case of the superposition (2.3) that gives the general Fourier synthesis solution to the scalar wave equation:

$$\Psi(\mathbf{r},t)=\frac{1}{(2\pi)^2}\sum_{n=0}^{\infty}\int_0^\infty d\kappa\int_{-\infty}^{+\infty}d\omega\int_{-\infty}^{+\infty}\times dk_z\,A_n(\omega,k_z,\kappa)\kappa J_n(\kappa\rho)e^{\pm in\phi}\times e^{-ik_z z}e^{+i\omega t}\delta(\omega^2-\kappa^2-k_z^2)\,. \quad (2.18)$$

Next consider the choice $\hat{A}(-i\partial_z)=-\partial_z^2$ which reduces Eqs. (2.10) to

$$\zeta=z-t \quad \text{and} \quad \eta=z+t\,. \quad (2.19)$$

The operator L can be written, in this case, as

$$L=-[4\partial_{\zeta\eta}^2+\partial_\rho^2+\rho^{-1}\partial_\rho+\rho^{-2}\partial_\phi^2]\,,$$

and it can be partitioned as follows:

$$L_1=-[\partial_\rho^2+\rho^{-1}\partial_\rho+\rho^{-2}\partial_\phi^2]\,, \quad (2.20a)$$

$$L_2=-4\partial_{\zeta\eta}^2\,. \quad (2.20b)$$

The eigenfunctions of L_1 are given now by $J_n(\kappa\rho)e^{\pm in\phi}$ and $N_n(\kappa\rho)e^{\pm in\phi}$, and its eigenvalues equal $+\kappa^2$. The operator L_2 has eigenfunctions $e^{-i\alpha\zeta}e^{i\beta\eta}$ with eigenvalues $-4\alpha\beta$. An elementary solution to the scalar wave equation (2.15) can be written as

$$\Psi_e(\mathbf{r},t)\equiv\Psi_e(\rho,\zeta,\eta)=[C_nJ_n(\kappa\rho)+D_nN_n(\kappa\rho)]e^{\pm in\phi}e^{-i\alpha\zeta}e^{i\beta\eta}\,, \quad (2.21a)$$

with the constraint

$$\alpha\beta=\kappa^2/4\,. \quad (2.21b)$$

This constraint limits the value of α and β either to be both negative or both positive. A general solution to the scalar wave equation can be written in the nonconventional form

$$\Psi(\rho,\zeta,\eta)=\frac{1}{(2\pi)^2}\sum_{\substack{l=-1\\ l\neq 0}}^{+1}\sum_{n=0}^{\infty}\int_0^\infty d\kappa\int_0^\infty d(l\alpha)\times\int_0^\infty d(l\beta)\,C_n(l\alpha,l\beta,\kappa)\kappa J_n(\kappa\rho)\times e^{\pm in\phi}e^{-il\alpha\zeta}e^{il\beta\eta}\delta(\alpha\beta-\kappa^2/4)\,. \quad (2.22)$$

The two representations [cf. Eqs. (2.18) and (2.22)] may appear to be very different. There exists, however, a one to one correspondence between these two superpositions through the change of variables

$$k_z=\alpha-\beta\,,\quad \omega=\alpha+\beta\,. \quad (2.23)$$

By using these relationships, the new representation (2.22) can be transformed into the Fourier synthesis given in Eq. (2.18), with the following connection between their spectra:

$$A_n(\omega,k_z,\kappa)=2C_n[\tfrac{1}{2}(\omega+k_z),\tfrac{1}{2}(\omega-k_z),\kappa]\,. \quad (2.24)$$

It should be noted that this transformation requires a careful handling of the limits of integration. A complete discussion of this point will be given later when dealing with specific examples.

The representation (2.22) provides a fresh path through which exact solutions to the scalar wave equation can be obtained. Although such a representation is not a familiar one, solutions obtained using (2.22) can still be easily transformed into the more popular Fourier superposition, and one can link the bidirectional results to the more conventional Fourier interpretation. To emphasize these ideas, one can remove the constraint in (2.18) by integrating over ω, hence reducing (2.18) to a form similar to (2.2), with

$$\Psi(\mathbf{r},t)=\frac{1}{(2\pi)^2}\sum_{n=0}^{\infty}\int_0^\infty d\kappa\int_{-\infty}^{+\infty}\times dk_z\left[A_n\left[\sqrt{\kappa^2+k_z^2},k_z,\kappa\right]e^{-i(k_z z-\sqrt{\kappa^2+k_z^2}\,t)}+A_n\left[-\sqrt{\kappa^2+k_z^2},k_z,\kappa\right]\right.$$

$$\times e^{-i(k_z z+\sqrt{\kappa^2+k_z}\,t)}]\frac{\kappa J_n(\kappa\rho)}{2\sqrt{\kappa^2+k_z}}e^{\pm in\phi}. \quad (2.25)$$

This is a special case of the superposition (2.2) and consists of the sum of two components, one traveling in the positive z direction and the other in the negative z direction. A common problem that arises when dealing with such integrals is associated with the branch-cut type singularities. These can pose significant difficulties even when the integrals are solved either asymptotically or computed numerically. On the other hand, when the constraint is integrated out of Eq. (2.22), one obtains

$$\Psi(\rho,\zeta,\eta)=\frac{1}{(2\pi)^2}\sum_{n=0}^{\infty}\int_0^{\infty}d\kappa\int_0^{\infty} \times d\beta\left[C_n\left(\frac{\kappa^2}{4\beta},\beta,\kappa\right)e^{(-i\kappa^2/4\beta)\zeta}e^{i\beta\eta} - C_n\left(-\frac{\kappa^2}{4\beta},-\beta,\kappa\right)e^{(i\kappa^2/4\beta)\zeta}e^{-i\beta\eta}\right] \times\frac{\kappa}{\beta}J_n(\kappa\rho)e^{\pm in\phi}. \quad (2.26)$$

Similar to the Fourier synthesis where one can choose either the positive or negative ω branch, we can choose to work with either the positive or the negative branch of α and β. In what follows, for convenience only, we choose the positive branch. Notice that, unlike Eq. (2.25), the terms in the above integral consist of products of two plane waves traveling in opposite directions. An important characteristic of the representation (2.26) is that the branch-cut singularities in Eq. (2.25) have been converted into algebraic singularities. This provides a novel approach to finding solutions to the scalar wave equation. New exact solutions can be obtained by choosing appropriate spectra C_n, for which the corresponding Fourier spectra A_n might be very complicated and could not have been guessed. Moreover, because of the nature of the branch-cut singularities in the Fourier synthesis, problems arise because of their multivaluedness and because large oscillations accompany any attempt to evaluate them either numerically or asymptotically. We have found that one can circumvent such problems by dealing with the bidirectional synthesis and its tame algebraic singularities.

A number of important mathematical issues dealing with the new bidirectional synthesis will be considered at this point. These will include the completeness of the expansions (2.22) and (2.26), the inversion properties for $C_n(\alpha,\beta,\kappa)$, and conditions that must be imposed on the spectrum in order to ensure square integrability. The feasibility of solving Cauchy initial value problems on the basis of the new representation will be addressed in Sec. IV.

Completeness follows directly from the fact that the superimposed functions are either exponential or Bessel functions, which are both orthogonal functions and form their own complete sets. However, the inversion of $C_n(\alpha,\beta,\kappa)$ is not obvious. A generalization of Ziolkowski's formula (1.6) had to be used. Using the positive β branch in (2.26), $\Psi(\rho,\zeta,\eta)$ can be constructed as follows:

$$\Psi(\rho,\zeta,\eta)=\frac{1}{(2\pi)^2}\sum_{n=-\infty}^{+\infty}\int_0^{\infty}d\kappa\int_0^{\infty}d\beta\frac{1}{\beta} \times C_n\left(\frac{\kappa^2}{4\beta},\beta,\kappa\right)\kappa J_n(\kappa\rho)e^{in\phi}e^{-(i\kappa^2/4\beta)\zeta}e^{i\beta\eta}. \quad (2.27)$$

The inversion formula corresponding to this superposition is given by

$$C_m\left(\frac{\kappa^2}{4\beta},\beta,\kappa\right)=\frac{1}{4\sqrt{\pi}}\int_{-\pi}^{+\pi}d\phi\int_{-\infty}^{+\infty}d\zeta \times e^{-\zeta^2/16\beta^2}\int_{-\infty}^{+\infty}d\eta\int_0^{\infty}d\rho\,\rho\,J_m(\kappa\rho) \times\Psi(\rho,\zeta,\eta)e^{-im\phi}e^{(i\kappa^2/4\beta)\zeta}e^{-i\beta\eta}. \quad (2.28)$$

Note the appearance of the Gaussian measure over ζ. A similar measure occurred in Ziolkowski's inversion (1.6). However, in the more general inversion (2.28) the additional parameter a_1 has disappeared. The validity of the inversion will be demonstrated below in connection to specific examples.

To investigate the possible restrictions on the spectrum C_0 that would ensure square integrability of the solution, one can consider the integral over β in Eq. (2.27), namely,

$$\psi(z,t)=\int_0^{\infty}d\beta\frac{1}{\beta}C_0\left(\frac{\kappa^2}{4\beta},\beta,\kappa\right)e^{-i\kappa^2(z-t)/4\beta}e^{i\beta(z+t)}. \quad (2.29)$$

By rearranging the variables in the exponentials, one obtains

$$\psi(z,t)=\int_0^{\infty}d\beta\frac{1}{\beta}C_0\left(\frac{\kappa^2}{4\beta},\beta,\kappa\right) \times\exp\left[-iz\left(\frac{\kappa^2}{4\beta}-\beta\right)\right]\exp\left[it\left(\frac{\kappa^2}{4\beta}+\beta\right)\right],$$

or

$$\psi(z,t)=\int_0^{\infty}d\beta\frac{1}{\beta}C_0\left(\frac{\kappa^2}{4\beta},\beta,\kappa\right) \times\exp\left[\frac{i\kappa z}{2}\left(\frac{2\beta}{\kappa}-\frac{\kappa}{2\beta}\right)\right] \times\exp\left[\frac{\kappa t}{2}\left(\frac{i2\beta}{\kappa}-\frac{\kappa}{i2\beta}\right)\right]. \quad (2.30)$$

By using the Laurent expansion of the Bessel generating function, viz.,

$$\exp\left[\frac{x}{2}\left(t-\frac{1}{t}\right)\right]=\sum_{n=-\infty}^{+\infty}J_n(x)t^n, \quad (2.31)$$

the exponentials in (2.30) can be rewritten as

$$\exp\left[\frac{i\kappa z}{2}\left(\frac{2\beta}{\kappa}-\frac{\kappa}{2\beta}\right)\right]=\sum_{n=-\infty}^{+\infty}(i)^nI_n(\kappa z)\left[\frac{2\beta}{\kappa}\right]^n,$$

$$\exp\left[\frac{\kappa t}{2}\left(\frac{i2\beta}{\kappa}-\frac{\kappa}{i2\beta}\right)\right]=\sum_{m=-\infty}^{+\infty}(i)^mJ_m(\kappa t)\left[\frac{2\beta}{\kappa}\right]^m.$$

By using these expansions, Eq. (2.30) can be rewritten as

$$\psi(z,t)=\sum_{n=-\infty}^{+\infty}\sum_{m=-\infty}^{+\infty}C_{mn}I_n(\kappa z)J_m(\kappa t), \quad (2.32a)$$

where

$$C_{mn} = \int_0^\infty d\beta \left[\frac{2\beta}{\kappa}\right]^{m+n} i^{m+n} C_0\left(\frac{\kappa^2}{2\beta}, \beta, \kappa\right). \quad (2.32b)$$

A necessary condition for the convergence of (2.32a) is that $C_{mn} < \infty$ for all values of m and n ranging from $-\infty$ to $+\infty$. By considering the integral (2.32b), it is then obvious that $C_0(\kappa^2/\beta,\beta,\kappa)$ should obey the conditions

$$\lim_{\beta \to 0} \frac{1}{\beta^r} C_0\left(\frac{\kappa^2}{4\beta}, \beta, \kappa\right) < \infty, \quad (2.33a)$$

$$\lim_{\beta \to \infty} \beta^r C_0\left(\frac{\kappa^2}{4\beta}, \beta, \kappa\right) < \infty, \quad (2.33b)$$

for arbitrary β, and $r = m + n$, for any integer values of m and n. A good candidate is a spectrum of the form

$$C_0\left(\frac{\kappa^2}{4\beta}, \beta, \kappa\right) = \beta^r \exp\left[-\beta a_1 + \frac{\kappa^2}{4\beta} a_2\right]. \quad (2.34)$$

This is similar to the one used for the splash pulses and EDEPT solutions, as will be demonstrated in Sec. III.

In summary, the procedure described in this section provides an alternate way of synthesizing solutions to different partial differential equations. Such representations are characterized by different types of singularities that may facilitate their asymptotic or numerical evaluation. This is a flexible procedure that changes with the types of equations considered. Moreover, solutions to the same equation may have different representations depending on how the operator L is partitioned. In Sec. III, the bidirectional representation (2.22) will be used as a natural superposition for the synthesis of Brittingham-like solutions, e.g., focus wave modes, splash pulses, Bessel beams, and EDEPT solutions. This will enable us to gain a better understanding of these unusual solutions, and by using the transformation (2.23), to obtain more information about their Fourier spectral content. Other types of equations, dealing with dispersive and dissipative problems, will be discussed in Sec. IV, where it will be demonstrated that the bidirectional representation can reduce the complexity level of such equations to that of the three-dimensional scalar wave equation.

III. BIDIRECTIONAL PLANE WAVE DECOMPOSITION OF KNOWN SOLUTIONS

It was demonstrated in the previous section that the main achievement of the embedding technique is to introduce a time-symmetric bidirectional representation. For the scalar wave equation, such a representation is given in Eq. (2.22). It turns out that such a superposition provides the most natural approach for synthesizing Brittingham-like solutions. This section is devoted, mainly, to substantiating this claim. Starting with the scalar analog of Brittingham's FWM's, it will be shown that by choosing very simple spectra $C_n(\alpha,\beta,\kappa)$, mostly of the type given in Eq. (2.34), all known Brittingham-like solutions can be synthesized. Because of the simple transformation (2.23), it will be easy to transform such solutions to their Fourier picture, from which a basic understanding of their spectral content can be achieved. These examples can also provide a vehicle through which the inversion formula can be checked. The following discussion will be restricted to the zeroth order mode ($n = 0$). This is a matter of convenience and does not affect the generality of the procedure.

A. Focus wave modes

The focus wave modes (FWM's) were originally stimulated by the work of Brittingham,[1] who derived their vector form in connection with Maxwell's equations. Their scalar form was derived by Belanger,[3] Sezinger,[4] and Ziolkowski.[2] These modes, the zeroth order of which is given in Eq. (1.2), are characterized by an infinite energy content. Motivated by the bidirectional character of the solution (1.2), it will be shown below that the representation (2.22) can be used to synthesize the FWM's associated with the scalar wave equation.

Consider the spectrum

$$C_0(\alpha,\beta,\kappa) = (\sqrt{\pi}/2)\sigma e^{-\sigma^2(\beta-\beta')^2} e^{-\alpha a_1}. \quad (3.1)$$

Substituting it into Eq. (2.22) results in the expression

$$\Psi(\rho,\zeta,\eta) = \frac{1}{(2\pi)^2}\int_0^\infty d\kappa \int_0^\infty d\beta \int_0^\infty d\alpha \times \frac{\sqrt{\pi}\sigma}{2} e^{-\sigma^2(\beta-\beta')^2} e^{-\alpha a_1} \kappa J_0(\kappa\rho) \times e^{-i\alpha\zeta} e^{i\beta\eta} \delta[\alpha\beta - \kappa^2/4]. \quad (3.2)$$

An integration over α reduces Eq. (3.2) to

$$\Psi(\rho,\zeta,\eta) = \frac{\sigma}{8\pi^{3/2}}\int_0^\infty d\kappa \int_0^\infty d\beta \frac{\kappa}{\beta} \times J_0(\kappa\rho) e^{-\sigma^2(\beta-\beta')^2} e^{-\kappa^2(a_1+i\zeta)/4\beta} e^{i\beta\eta}.$$

By using equation (6.631.4) in Gradshteyn and Ryzhik,[39] viz.,

$$\int_0^\infty dx\, x^{\nu+1} e^{-ax^2} J_\nu(\alpha x) = \frac{\alpha^\nu}{(2a)^{\nu+1}} e^{-\alpha^2/4a}, \quad (3.3)$$

the integration over κ can be carried out explicitly, yielding

$$\Psi(\rho,\zeta,\eta) = \frac{\sigma}{4\pi^{3/2}}\int_0^\infty d\beta \frac{1}{(a_1+i\zeta)} \times e^{-\sigma^2(\beta-\beta')^2} e^{-\beta\rho^2/(a_1+i\zeta)} e^{i\beta\eta}.$$

To carry out the final integration over β, Eq. (3.4621.1) in Gradshteyn and Ryzhik[39] is used, viz.,

$$\int_0^\infty dx\, x^{\nu-1} e^{-\gamma x} e^{-\beta x^2} = \frac{\Gamma(\nu)}{(2\beta)^{\nu/2}} e^{\gamma^2/8\beta} D_{-\nu}\left(\frac{\gamma}{\sqrt{2\beta}}\right),$$

to give the solution

$$\Psi(\rho,\zeta,\eta) = \frac{1}{4\pi(a_1+i\zeta)} \frac{e^{-\sigma^2\beta'^2}}{\sqrt{2\pi}} \times e^{(\Lambda - i\eta - 2\sigma^2\beta')^2/8\sigma^2} D_{-1}\left(\frac{\Lambda - i\eta - 2\sigma^2\beta'}{\sqrt{2}\sigma}\right), \quad (3.4a)$$

where D_{-1} is the parabolic cylinder function of order -1 and

$$\Lambda = \rho^2/(a_1 + i\zeta).$$

The solution (3.4a) is a generalization of the scalar FWM's; the latter can be recovered by taking the limit

$\sigma \to \infty$, for which the spectrum in (3.1) reduces to

$$C_0(\alpha,\beta,\kappa) = (\pi/2)\delta(\beta-\beta')e^{-\alpha a_1}.$$

In what follows, we shall use $\hat{\delta}$ to denote the part of the spectrum in (3.1) that reduces to the Dirac δ function as $\sigma \to \infty$. This will yield less cumbersome expressions and will make our discussion more transparent. As for the solution in (3.4a), it is more convenient to compute the limit $\sigma \to \infty$ after rewriting the parabolic cylinder function in an alternate form using the identity (9.254.1) in Gradshteyn and Ryzhik,[39] namely,

$$D_{-1}(z) = e^{z^2/4}\sqrt{\pi/2}\,[1-\Phi(z/\sqrt{2})]\,,$$

where $\Phi(z)$ is the probability integral defined as follows:

$$\Phi(z) = \frac{2}{\sqrt{\pi}}\int_0^z e^{-t^2}\,dt\,.$$

Rewriting the wave function (3.4a) in terms of the probability integral, viz.,

$$\Psi(\rho,\zeta,\eta) = \frac{e^{-\beta'(\Lambda - i\eta)}}{8\pi(a_1+i\zeta)}\,e^{(\Lambda-i\eta)^2/4\sigma^2} \times\left[1-\Phi\left(\frac{\Lambda-i\eta}{2\sigma}-\sigma\beta'\right)\right],$$

facilitates taking the limit $\sigma \to \infty$ since $\Phi(-\infty) = -1$. Hence as σ goes to ∞, the above expression reduces to the scalar FWM solution

$$\Psi(\rho,\zeta,\eta) = [1/4\pi(a_1+i\zeta)]e^{-\beta'(\Lambda - i\eta)}\,. \tag{3.4b}$$

Felsen and Heyman[25,29] have established the acausal nature of the FWM's, in the limit where $\beta' a_1 \gg 1$, using their approximate STT theory. The causality issue can be handled in a more direct way by transforming (3.2) into the Fourier picture using the relationships

$$\beta = \tfrac{1}{2}(\omega - k_z)\,,\quad \alpha = \tfrac{1}{2}(\omega + k_z)\,. \tag{3.5}$$

If $\Psi(\rho,\zeta,\eta)$ in Eq. (3.2) is rewritten as

$$\Psi(\rho,\zeta,\eta) = \frac{1}{(2\pi)^2}\int_0^\infty d\kappa\,\kappa J_0(\kappa\rho)\tilde{\psi}(\kappa,\zeta,\eta)\,, \tag{3.6a}$$

with

$$\tilde{\psi}(\kappa,\zeta,\eta) = \int_0^\infty d\beta\int_0^\infty d\alpha\,\frac{\pi}{2}\,\hat{\delta}(\beta-\beta') \times e^{-\alpha a_1}e^{-i\alpha\zeta}e^{i\beta\eta}\,\delta\left[\alpha\beta-\frac{\kappa^2}{4}\right], \tag{3.6b}$$

one can use Eq. (3.5) to express $\tilde{\psi}(\kappa,z,t)$ as follows:

$$\tilde{\psi}(\kappa,z,t) = \int_0^\infty dk_z\int_{k_z}^\infty d\omega\,\frac{\pi}{2}\,\hat{\delta}\left[\frac{(\omega-k_z)}{2}-\beta'\right]e^{-(a_1/2)(\omega+k_z)}\,\delta\left[\frac{\omega^2}{4}-\frac{k_z^2}{4}-\frac{\kappa^2}{4}\right]e^{-i(k_z z-\omega t)}$$
$$+\int_{-\infty}^0 dk_z\int_{-k_z}^\infty d\omega\,\frac{\pi}{2}\,\hat{\delta}\left[\frac{(\omega-k_z)}{2}-\beta'\right]e^{-(a_1/2)(\omega+k_z)}\,\delta\left[\frac{\omega^2}{4}-\frac{k_z^2}{4}-\frac{\kappa^2}{4}\right]e^{-i(k_z z-\omega t)}\,. \tag{3.7}$$

An integration over ω simplifies (3.7) to

$$\tilde{\psi}(\kappa,z,t) = \int_0^\infty dk_z\,\frac{2\pi}{\sqrt{k_z^2+\kappa^2}}\,\hat{\delta}\left[\sqrt{k_z^2+\kappa^2}-k_z-2\beta'\right]e^{-(a_1/2)\left[\sqrt{k_z^2+\kappa^2}+k_z\right]}e^{-i(k_z z-\overline{\omega}t)}$$
$$+\int_0^\infty dk_z\,\frac{2\pi}{\sqrt{k_z^2+\kappa^2}}\,\hat{\delta}\left[\sqrt{k_z^2+\kappa^2}+k_z-2\beta'\right]e^{-(a_1/2)\left[\sqrt{k_z^2+\kappa^2}-k_z\right]}e^{i(k_z z+\overline{\omega}t)}\,, \tag{3.8}$$

where $\overline{\omega} = \sqrt{k_z^2+\kappa^2}$. Referring to Fig. 2, it is clear that the first integral in (3.8) vanishes for $\beta' < \kappa/2$ while the second one vanishes for $\beta' > \kappa/2$. As a result, $\Psi(\mathbf{r},t)$ can be divided into two parts, one traveling in the positive z direction and the other in the negative z direction, viz.,

$$\Psi(\mathbf{r},t) = \Psi^+(\mathbf{r},t) + \Psi^-(\mathbf{r},t)\,,$$

where

$$\Psi^+(\mathbf{r},t) = \frac{1}{(2\pi)^2}\int_{2\beta'}^\infty d\kappa\,\kappa J_0(\kappa\rho) \times\int_0^\infty dk_z\,F(k_z,\kappa)e^{-i(k_z z-\overline{\omega}t)}\,, \tag{3.9a}$$

$$\Psi^-(\mathbf{r},t) = \frac{1}{(2\pi)^2}\int_0^{2\beta'} d\kappa\,\kappa J_0(\kappa\rho) \times\int_0^{\beta'} dk_z\,F(-k_z,\kappa)e^{i(k_z z+\overline{\omega}t)}\,, \tag{3.9b}$$

and

$$F(k_z,\kappa) = (2\pi/\sqrt{k_z^2+\kappa^2})\hat{\delta}\left[\sqrt{k_z^2+\kappa^2}-k_z-2\beta'\right] \times e^{-\left[\sqrt{k_z^2+\kappa^2}+k_z\right]a_1/2}\,. \tag{3.10}$$

If the parameter a_1 is large, the spectrum $F(k_z,\kappa)$ in (3.10) has a very narrow bandwidth, while $F(-k_z,\kappa)$ can maintain a balance between $\sqrt{k_z^2+\kappa^2}$ and k_z in the exponential and, consequently, can have a much larger bandwidth bounded by the upper limits of integration over κ and k_z in (3.9b). In this case, the predominant contribution to $\Psi(\mathbf{r},t)$ comes from $\Psi^-(\mathbf{r},t)$. This contribution is primarily a nonlocalized plane wave moving in the negative z direction. If, on the other hand, a_1 is very small, both $F(k_z,\kappa)$ and $F(-k_z,\kappa)$ have large bandwidths and because of the limited range of integration in the expression for $\Psi^-(\mathbf{r},t)$ compared to the infinite range for $\Psi^+(\mathbf{r},t)$, one expects that $\Psi^+(\mathbf{r},t)$ becomes much larger than $\Psi^-(\mathbf{r},t)$. In this case, the solution $\Psi(\mathbf{r},t)$ behaves like a localized pulse moving in the positive z direction.

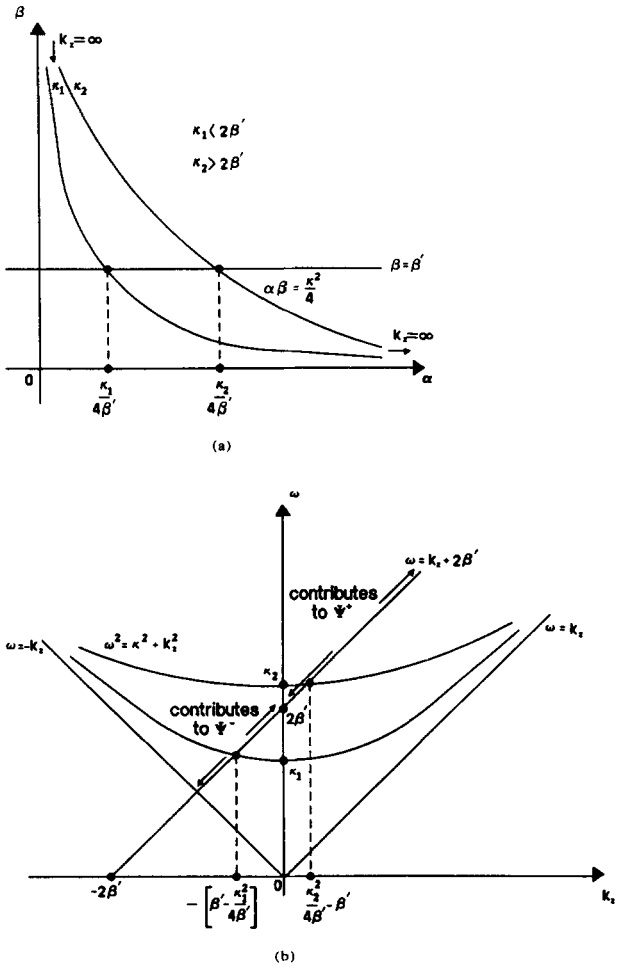


FIG. 2. The constraints $\alpha\beta = \kappa^2/4$ and $\beta = \beta'$ are shown in: (a) the $\alpha\beta$ plane and (b) the $k_z\omega$ plane. The contributions to Ψ^+ and Ψ^- are indicated by arrows.

In closing this subsection, we shall check the validity of the inversion given in Eq. (2.28). A substitution of (3.4b) into Eq. (2.28), leads to the expression

$$C_0\left(\frac{\kappa^2}{4\beta}, \beta, \kappa\right) = \frac{\sqrt{\pi}}{2}\int_{-\infty}^{+\infty} d\zeta\, e^{-\zeta^2/16\beta^2}\int_{-\infty}^{+\infty} d\eta \times \int_0^{+\infty} d\rho\, \rho J_0(\kappa\rho)\frac{e^{i\beta'\eta}}{4\pi(a_1 + i\zeta)} \times e^{-\beta'\rho^2/(a_1+i\zeta)} e^{(i\kappa^2/4\beta)\zeta} e^{-i\beta\eta}. \quad (3.11)$$

Integrating over η and making use of Eq. (3.3) in order to carry out the integration over ρ, it follows that

$$C_0\left(\frac{\kappa^2}{4\beta}, \beta, \kappa\right) = \frac{\sqrt{\pi}}{8}\int_{-\infty}^{+\infty} d\zeta\, e^{-\zeta^2/16\beta^2}\delta(\beta - \beta')\frac{1}{\beta'} e^{-\kappa^2 a_1/4\beta'}.$$

The relation

$$\int_{-\infty}^{+\infty} d\zeta\, e^{-\zeta^2/16\beta^2} = \sqrt{16\pi}\,\beta$$

yields, finally, the result

$$C_0(\kappa^2/4\beta, \beta, \kappa) = (\pi/2) e^{-\kappa^2 a_1/4\beta}\delta(\beta - \beta'), \quad (3.12)$$

which is identical to the spectrum given in Eq. (3.1) provided that $\alpha = \kappa^2/4\beta$. The latter follows from the constraint embodied in Eq. (2.21b).

B. Splash modes

The original "splash mode" was introduced by Ziolkowski[2] as the first example of the class of finite energy solutions constructed from superpositions of the original FWM's. Hillion[10,11] has extended the FWM and the splash mode concepts to the realm of spinors. Ziolkowski's splash pulse can be derived within the framework of the bidirectional representation by choosing the spectrum $C_0(\alpha,\beta,\kappa)$ as follows:

$$C_0(\alpha,\beta,\kappa) = (\pi/2)\beta^{q-1} e^{-(\alpha a_1 + \beta a_2)}. \quad (3.13)$$

It should be noted that this choice is a specific example of the general class of spectra given in Eq. (2.34). Substituting (3.13) into Eq. (2.22) yields

$$\Psi(\rho,\zeta,\eta) = \frac{1}{(2\pi)^2}\int_0^\infty d\kappa \int_0^\infty d\beta \int_0^\infty d\alpha \frac{\pi}{2}\beta^{q-1} \times e^{-(\alpha a_1 + \beta a_2)}\kappa J_0(\kappa\rho) e^{-i\alpha\zeta} e^{i\beta\eta}\delta\left[\alpha\beta - \frac{\kappa^2}{4}\right]. \quad (3.14)$$

The integration over α can be carried out explicitly, viz.,

$$\Psi(\rho,\zeta,\eta) = \frac{1}{8\pi}\int_0^\infty d\kappa \int_0^\infty d\beta \times \kappa J_0(\kappa\rho)\beta^{q-2} e^{-(\kappa^2/4\beta)(a_1 + i\zeta)} e^{-\beta(a_2 - i\eta)}.$$

Equation (3.471.9) in Gradshteyn and Ryzhik,[39] viz.,

$$\int_0^\infty d\beta\, \beta^{q-1}\exp\left[-\frac{a}{\beta} - b\beta\right] = 2\left[\frac{a}{b}\right]^{q/2} K_q\left[2\sqrt{ab}\right], \quad (3.15)$$

facilitates the integration over β; specifically,

$$\Psi(\rho,\zeta,\eta) = \frac{1}{8\pi}\int_0^\infty d\kappa\, \kappa^q J_0(\kappa\rho) 2\left(\frac{(a_1 + i\zeta)}{4(a_2 - i\eta)}\right)^{(q-1)/2} \times K_{q-1}\left[\kappa\sqrt{(a_1 + i\zeta)(a_2 - i\eta)}\right], \quad (3.16)$$

where K_q is the modified Bessel function of the second kind. To carry out the final integration over κ, formula (6.576.3) in Gradshteyn and Ryzhik[39] is used, viz.,

$$\int_0^\infty dx\, x^{-\lambda} K_\mu(ax) J_\nu(bx) = \frac{b^\nu \Gamma((\nu - \lambda + \mu + 1)/2)\Gamma((\nu - \lambda - \mu + 1)/2)}{2^{\lambda+1} a^{\nu-\lambda+1}\Gamma(\nu+1)} \times F\left(\frac{\nu - \lambda + \mu + 1}{2}, \frac{\nu - \lambda - \mu + 1}{2}, 1, -\frac{b^2}{a^2}\right). \quad (3.17)$$

This leads to the result

$$\Psi(\rho,\zeta,\eta)=\frac{\Gamma(q)}{4\pi}\left(\frac{(a_1+i\zeta)}{(a_2-i\eta)}\right)^{(q-1)/2}\times\frac{F(q,1,1,-\rho^2/[(a_1+i\zeta)(a_2-i\eta)])}{[(a_1+i\zeta)(a_2-i\eta)]^{(q+1)/2}},\tag{3.18}$$

where $F(q,1,1,-p)$ is the hypergeometric function. The latter has the property that

$$F(q,1,1,-p)=1/(1+p)^q.$$

Hence (3.18) takes the form

$$\Psi(\rho,\zeta,\eta)=\frac{\Gamma(q)}{4\pi(a_1+i\zeta)}\left[(a_2-i\eta)+\frac{\rho^2}{(a_1+i\zeta)}\right]^{-q},\tag{3.19}$$

which is identical to that for the splash pulse introduced in Ref. 2.

It is interesting to note in connection with Eq. (3.16) that the transverse components are separated from ζ and η. The portion of $\Psi(\rho,\zeta,\eta)$ depending on ζ and η only, viz.,

$$\tilde{\psi}(\kappa,\zeta,\eta)\equiv\pi\left(\frac{(a_1+i\zeta)}{4(a_2-i\eta)}\right)^{(q-1)/2}\times\kappa^{q-1}K_{q-1}[\kappa\sqrt{(a_1+i\zeta)(a_2-i\eta)}]\tag{3.20}$$

is a solution to the one-dimensional Klein–Gordon equation.

The scalar wave equation analog to Hillion's splash modes can easily be derived by choosing the spectrum

$$C_0(\alpha,\beta,\kappa)=(\pi/2)J_\nu(\beta b)e^{-\alpha a_1}.\tag{3.21}$$

In this case,

$$\Psi(\rho,\zeta,\eta)=\frac{1}{(2\pi)^2}\int_0^\infty d\kappa\int_0^\infty d\beta\int_0^\infty d\alpha\,\frac{\pi}{2}\times J_\nu(\beta b)e^{-\alpha a_1}\kappa J_0(\kappa\rho)e^{-i\alpha\zeta}e^{i\beta\eta}\delta\left[\alpha\beta-\frac{\kappa^2}{4}\right],\tag{3.22}$$

or

$$\Psi(\rho,\zeta,\eta)=\frac{1}{8\pi}\int_0^\infty d\kappa\int_0^\infty d\beta\,\frac{\kappa}{\beta}\times J_0(\kappa\rho)J_\nu(\beta b)e^{-\kappa^2(a_1+i\zeta)/4\beta}e^{i\beta\eta}$$

upon integrating over α. The integration over κ can be carried out using Eq. (3.3). One finds

$$\Psi(\rho,\zeta,\eta)=\frac{1}{4\pi}\int_0^\infty d\beta\,\frac{J_\nu(\beta b)}{(a_1+i\zeta)}e^{-\beta s},$$

where

$$s=\rho^2/(a_1+i\zeta)-i\eta.\tag{3.23}$$

Using relation (6.611.1) in Gradshteyn and Ryzhik,[39] viz.,

$$\int_0^\infty dx\,e^{-\alpha x}J_\nu(\beta x)=\frac{\beta^{-\nu}[\sqrt{\alpha^2+\beta^2}-\alpha]^\nu}{\sqrt{\alpha^2+\beta^2}},\tag{3.24}$$

$\Psi(\rho,\zeta,\eta)$ assumes, finally, the form

$$\Psi(\rho,\zeta,\eta)=\frac{1}{4\pi(a_1+i\zeta)}\frac{b^{-\nu}[\sqrt{s^2+b^2}-s]^\nu}{\sqrt{s^2+b^2}},\tag{3.25}$$

which is a solution to the three-dimensional scalar wave equation analogous to Hillion's spinors. The Bessel function $J_\nu(\beta b)$ entering into the (3.21) forms a complete orthogonal set. This means that any spectrum expressed as

$$C_0(\alpha,\beta,\kappa)=(\pi/2)F(\beta)e^{-\alpha a_1},$$

with

$$F(\beta)=\int_0^\infty db\,B(b)J_\nu(\beta b),$$

can result in the solution

$$\Psi(\rho,\zeta,\eta)=\frac{1}{4\pi(a_1+i\zeta)}\int_0^\infty dbB(b)\,\frac{b^{-\nu}[\sqrt{s^2+b^2}-s]^\nu}{\sqrt{s^2+b^2}},\tag{3.26}$$

which is a generalization of Hillion's result.

The Fourier spectral content of Ziolkowski's splash pulse will be discussed in the next section in conjunction with the "modified power spectrum" (MPS) pulse. The Fourier picture corresponding to Hillion's solution can be obtained using the same procedure as in Sec. III A. Starting with the function

$$\tilde{\psi}(\kappa,\zeta,\eta)=\int_0^\infty d\beta\int_0^\infty d\alpha\,\frac{\pi}{2}J_\nu(\beta b)\times e^{-\alpha a_1}e^{-i\alpha\zeta}e^{i\beta\eta}\delta\left[\alpha\beta-\frac{\kappa^2}{4}\right],\tag{3.27}$$

the relationships given in Eq. (3.5) can be used to find the corresponding Fourier representation; specifically,

$$\tilde{\psi}(\kappa,z,t)=\int_0^\infty dk_z\int_{k_z}^\infty d\omega\,\frac{\pi}{2}J_\nu\left[\frac{b}{2}(\omega-k_z)\right]e^{-(a_1/2)(\omega+k_z)}\delta\left[\frac{\omega^2}{4}-\frac{k_z^2}{4}-\frac{\kappa^2}{4}\right]e^{-i(k_zz-\omega t)}$$
$$+\int_{-\infty}^0 dk_z\int_{-k_z}^\infty d\omega\,\frac{\pi}{2}J_\nu\left[\frac{b}{2}(\omega-k_z)\right]e^{-(a_1/2)(\omega+k_z)}\delta\left[\frac{\omega^2}{4}-\frac{k_z^2}{4}-\frac{\kappa^2}{4}\right]e^{-i(k_zz-\omega t)}.\tag{3.28}$$

By integrating over ω, it follows that

$$\tilde{\psi}(\kappa,z,t)=\int_0^\infty dk_z\,\frac{\pi}{\sqrt{k_z^2+\kappa^2}}J_\nu\left[\frac{b}{2}\{\sqrt{k_z^2+\kappa^2}-k_z\}\right]e^{-(a_1/2)[\sqrt{k_z^2+\kappa^2}+k_z]}e^{-i(k_zz-\bar{\omega}t)}$$
$$+\int_0^\infty dk_z\,\frac{\pi}{\sqrt{k_z^2+\kappa^2}}J_\nu\left[\frac{b}{2}\{\sqrt{k_z^2+\kappa^2}+k_z\}\right]e^{-(a_1/2)[\sqrt{k_z^2+\kappa^2}-k_z]}e^{+i(k_zz-\bar{\omega}t)}.\tag{3.29}$$

It is seen, then, that $\Psi(\mathbf{r},t)$ can be divided into two portions, $\Psi^+(\mathbf{r},t)$ and $\Psi^-(\mathbf{r},t)$, given by

$$\Psi^+(\mathbf{r},t) = \frac{1}{(2\pi)^2}\int_0^\infty d\kappa\,\kappa J_0(\kappa\rho)\int_0^\infty dk_z \times F(k_z,\kappa)e^{-i(k_z z-\bar{\omega}t)} \tag{3.30a}$$

and

$$\Psi^-(\mathbf{r},t) = \frac{1}{(2\pi)^2}\int_0^\infty d\kappa\,\kappa J_0(\kappa\rho)\int_0^\infty dk_z \times F(-k_z,\kappa)e^{i(k_z z+\bar{\omega}t)}, \tag{3.30b}$$

where

$$F(k_z,\kappa) = \frac{\pi}{\sqrt{k_z^2+\kappa^2}} J_\nu\left[\frac{b}{2}\left\{\sqrt{k_z^2+\kappa^2}+k_z\right\}\right] e^{-(a_1/2)\left[\sqrt{k_z^2+\kappa^2}+k_z\right]}. \tag{3.31}$$

Unlike the FWM's, the spectrum in this case is not singular. As in the case of the FWM's, however, the $\Psi^-(\mathbf{r},t)$ part will predominate for large values of the parameter a_1. On the other hand, the contributions from both parts of the spectrum are almost equal for small values of a_1. This can be seen from the ratio

$$\frac{F(k_z,\kappa)}{F(-k_z,\kappa)} = \frac{J_\nu\left[\frac{b}{2}\left\{\sqrt{k_z^2+\kappa^2}-k_z\right\}\right]}{J_\nu\left[\frac{b}{2}\left\{\sqrt{k_z^2+\kappa^2}+k_z\right\}\right]} e^{-a_1 k_z}.$$

As indicated earlier,

$$F(k_z,\kappa) \ll F(-k_z,\kappa)$$

for large values of a_1. This is true for most of the frequency range contributing to the integrations (3.30a) and (3.30b). On the other hand, for a_1 very small,

$$F(k_z,\kappa) \simeq F(-k_z,\kappa)$$

for the most significant components of this spectrum.

C. EDEPT's

These solutions, which were first introduced by Ziolkowski,[8,9] have finite energy, are extremely localized and they are highly directive. Another important feature of these solutions is that they contain certain parameters that can be "tweaked up" so that a pulse is predominantly propagating in one direction. An interesting example of the EDEPT solutions is the MPS pulse which can be synthesized in the context of the bidirectional representation using the shifted spectrum

$$C_0(\alpha,\beta,\kappa) = [\pi p/2\Gamma(q)](p\beta-b)^{q-1} \times e^{-[\alpha a_1+(p\beta-b)a_2]}, \quad \beta > b/p, \qquad = 0, \quad b/p > \beta \geqslant 0. \tag{3.32}$$

A substitution of this spectrum into (2.22) leads to the following solution:

$$\Psi(\rho,\zeta,\eta) = \frac{1}{(2\pi)^2}\int_0^\infty d\kappa \int_{b/p}^\infty d\beta \int_0^\infty d\alpha \times \frac{\pi p}{2\Gamma(q)}(p\beta-b)^{q-1}e^{-[\alpha a_1+(p\beta-b)a_2]} \times \kappa J_0(\kappa\rho)e^{-i\alpha\zeta}e^{i\beta\eta}\delta\left[\alpha\beta-\frac{\kappa^2}{4}\right]. \tag{3.33}$$

By integrating over α, it follows that

$$\Psi(\rho,\zeta,\eta) = \frac{1}{8\pi}\int_0^\infty d\kappa \int_{b/p}^\infty d\beta\,\kappa J_0(\kappa\rho)\frac{p(p\beta-b)^{q-1}}{\beta\Gamma(q)} \times e^{-\kappa^2(a_1+i\zeta)/4\beta}e^{-(p\beta-b)a_2}e^{i\beta\eta}.$$

The integration over κ can be performed by resorting to the change of variables $\beta' = \beta - b/p$, and making use of Eq. (3.3):

$$\Psi(\rho,\zeta,\eta) = \frac{1}{4\pi}\int_0^\infty d\beta' \frac{p^q\beta'^{q-1}}{\Gamma(q)} \times e^{-\beta'(s+pa_2)}\frac{e^{-bs/p}}{(a_1+i\zeta)}.$$

The integration over β' can be carried out explicitly, resulting in the wave function

$$\Psi(\rho,\zeta,\eta) = \frac{1}{4\pi(a_1+i\zeta)}\frac{e^{-bs/p}}{[a_2+s/p]^q}, \tag{3.34}$$

which is identical to the MPS pulse introduced by Ziolkowski.[8,9]

A detailed analysis of the behavior of the MPS has been presented elsewhere.[8,9] Our main interest, at this point, is to transfer (3.33) into the corresponding Fourier representation in order to study the contributions from the positive- and negative-going components of the solution. A procedure identical to that introduced earlier yields, in this case,

$$\tilde{\psi}(\kappa,z,t) = \int_0^\infty d\omega \int_{-\infty}^{+\infty} dk_z \frac{\pi p}{2\Gamma(q)}\left[\frac{p}{2}(\omega-k_z)-b\right]^{q-1} e^{-a_1(\omega+k_z)/2}e^{-a_2(p(\omega-k_z)/2-b)} \times \delta\left[\frac{\omega^2}{4}-\frac{k_z^2}{4}-\frac{\kappa^2}{4}\right]e^{-i(k_z z-\omega t)} \tag{3.35a}$$

for $\frac{1}{2}(\omega-k_z) > b/p$, and

$$\tilde{\psi}(\kappa,z,t) = 0 \tag{3.35b}$$

for $\frac{1}{2}(\omega+k_z) < b/p$. The indicated ranges in the ω,k_z plane can be seen clearly by referring to Fig. (3b). Carrying out the integration over ω changes (3.35) to

$$\tilde{\psi}(\kappa,z,t) = \int_{-\infty}^{+\infty} dk_z \frac{\pi p}{\Gamma(q)}\left[\frac{p}{2}\left\{\sqrt{k_z^2+\kappa^2}-k_z\right\}-b\right]^{q-1} e^{ba_2}e^{-a_1\left\{\sqrt{k_z^2+\kappa^2}+k_z\right\}/2} \times e^{-a_2 p\left\{\sqrt{k_z^2+\kappa^2}-k_z\right\}/2}e^{-i(k_z z-\bar{\omega}t)}$$

for $\sqrt{k_z^2+\kappa^2}-k_z > 2b/p$, and

$$\tilde{\psi}(\kappa,z,t) = 0$$

for $\sqrt{k_z^2+\kappa^2}-k_z < 2b/p$. Solving for k_z and splitting $\tilde{\psi}(\kappa,z,t)$ into positive- and negative-going parts, results, finally, in the components

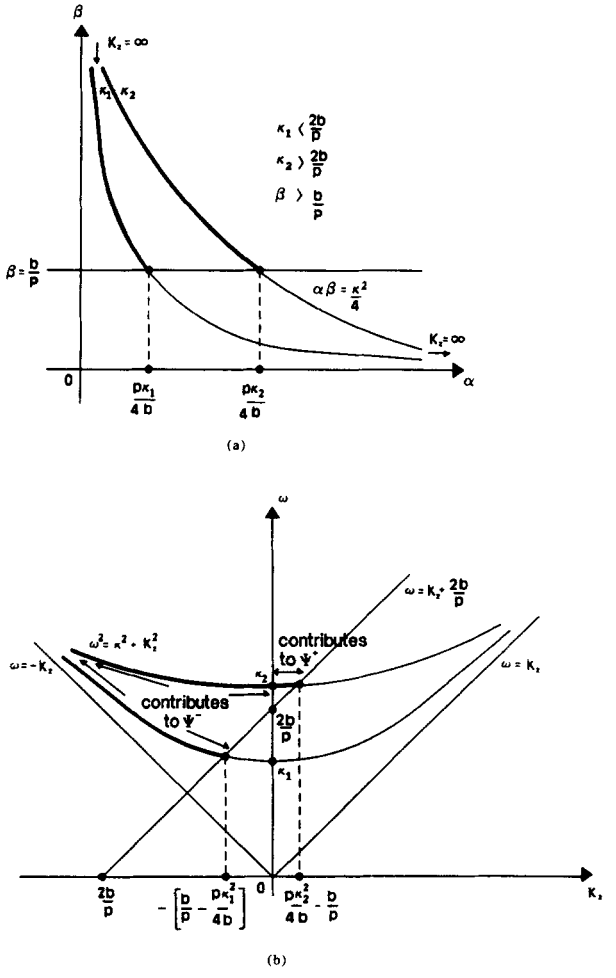


FIG. 3. The constraint $\alpha\beta=\kappa^2/4$ and the lower bound $\beta=b/p$ of the MPS pulse are shown in: (a) the $\alpha\beta$ plane and (b) the $k_z\omega$ plane. The contributions to Ψ^+ and Ψ^- are indicated by arrows.

$$\Psi^+(\mathbf{r},t)=\frac{1}{(2\pi)^2}\int_{2b/p}^{\infty}d\kappa\,\kappa J_0(\kappa\rho)\times\int_0^{\kappa^2p/4b-b/p}dk_z\,F(k_z,\kappa)e^{-i(k_zz-\bar{\omega}t)},\quad(3.36a)$$

$$\Psi^-(\mathbf{r},t)=\frac{1}{(2\pi)^2}\int_0^{2b/p}d\kappa\,\kappa J_0(\kappa\rho)\times\int_{b/p-\kappa^2p/4b}^{\infty}dk_z\,F(-k_z,\kappa)e^{i(k_zz+\bar{\omega}t)}+\frac{1}{(2\pi)^2}\int_{2b/p}^{\infty}d\kappa\,\kappa J_0(\kappa\rho)\times\int_0^{\infty}dk_z\,F(-k_z,\kappa)e^{i(k_zz+\bar{\omega}t)},\quad(3.36b)$$

where

$$F(k_z,\kappa)=\frac{\pi p}{\Gamma(q)}\left[\frac{p}{2}\left\{\sqrt{k_z^2+\kappa^2}-k_z\right\}-b\right]^{q-1}\times e^{ba_2}e^{-\sqrt{k_z^2+\kappa^2}(a_1+a_2p)/2}e^{-k_z(a_1-a_2p)/2}.\quad(3.37)$$

The strength of the MPS pulse arises from the introduced asymmetry in the positive- and negative-going components. This can be easily demonstrated by examining the ratio

$$F(k_z,\kappa)/F(-k_z,\kappa)=e^{(pa_2-a_1)k_z}$$

for $q=1$. By an appropriate choice of the parameters p, a_1, and a_2, the positive-going frequency components can be made much larger than the negative-going ones. This can be achieved by using large values of the product pa_2. It is also straightforward to demonstrate that in the limb $b\to0$ and $p\to1$ the MPS given in Eq. (3.34) is reduced to the splash pulse [cf. Eq. (3.19)]. Therefore, the Fourier spectral content of the splash pulse can be obtained directly from Eq. (3.36) by setting $b=0$ and $p=1$; specifically,

$$\Psi^+(\mathbf{r},t)=\frac{1}{(2\pi)^2}\int_0^{\infty}d\kappa\,\kappa J_0(\kappa\rho)\times\int_0^{\infty}dk_z\,F(k_z,\kappa)e^{-i(k_zz-\bar{\omega}t)},\quad(3.38a)$$

$$\Psi^-(\mathbf{r},t)=\frac{1}{(2\pi)^2}\int_0^{\infty}d\kappa\,\kappa J_0(\kappa\rho)\times\int_0^{\infty}dk_z\,F(-k_z,\kappa)e^{i(k_zz+\bar{\omega}t)},\quad(3.38b)$$

with

$$F(k_z,\kappa)=\pi\left[\tfrac{1}{2}\sqrt{k_z^2+\kappa^2}-k_z/2\right]^{q-1}\times e^{-\sqrt{k_z^2+\kappa^2}(a_1+a_2)/2}e^{-k_z(a_1-a_2)/2}.\quad(3.39)$$

To compare $\Psi^+(\mathbf{r},t)$ to $\Psi^-(\mathbf{r},t)$, consider the following ratio for $q=1$:

$$F(k_z,\kappa)/F(-k_z,\kappa)=e^{(a_2-a_1)k_z}.$$

It is clear from this expression that one can have a predominantly positive component if a_2 is chosen to be large and a_1 very small. However, unlike the MPS pulse, the splash pulse is not localized in the transverse directions. This is due to the absence of the parameters b and p that provide some control over the transverse localization through the factor $\exp(-bs/p)$ in Eq. (3.34).

D. Bessel beams

The "Bessel beams" were introduced by Durnin[20] and, like Brittingham's FWM's, they are characterized by an infinite energy content. It is of interest that such beams have been realized experimentally,[21] primarily because of the manner in which the behavior of an infinite energy beam can be realized approximately. It is possible to show that these beams can be represented by the time-symmetric bidirectional superposition (2.22). One can choose, in this case,

$$C_0(\alpha,\beta,\kappa)=4\pi\sigma\tau e^{-\sigma^2(\alpha+\beta-\omega_0)^2}e^{-\tau^2(\alpha-\beta-\lambda)^2},\quad(3.40)$$

for which Eq. (2.22) specializes to

$$\Psi(\rho,\zeta,\eta)=\frac{1}{(2\pi)^2}\int_0^{\infty}d\kappa\int_0^{\infty}d\beta\int_0^{\infty}d\alpha\times4\pi\sigma\tau e^{-\sigma^2(\alpha+\beta-\omega_0)^2}e^{-\tau^2(\alpha-\beta-\lambda)^2}\times\kappa J_0(\kappa\rho)e^{-i\alpha\zeta}e^{i\beta\eta}\delta\left[\alpha\beta-\frac{\kappa^2}{4}\right].\quad(3.41)$$

Integrating first over κ, one has

$$\Psi(\rho,\zeta,\eta)=\frac{2\sigma\tau}{\pi}\int_0^\infty d\beta\int_0^\infty d\alpha\, e^{-\sigma^2(\alpha+\beta-\omega_0)^2}\times e^{-\tau^2(\alpha-\beta-\lambda)^2}J_0(2\sqrt{\alpha\beta}\,\rho)e^{-i\alpha\zeta}e^{i\beta\eta}. \quad (3.42)$$

This integration is very hard to evaluate exactly; nevertheless, an asymptotic solution can be obtained for large values of $\sigma\tau$. Without any loss of generality we can take $\sigma=\tau$ and Eq. (3.42) can be rewritten as follows:

$$\Psi(\rho,\zeta,\eta)=\frac{2\sigma^2}{\pi}\int_0^\infty d\beta\int_0^\infty d\alpha\times e^{-\sigma^2[\omega_0^2+\lambda^2-2(\lambda+\omega_0)\alpha-2(\omega_0-\lambda)\beta+2\alpha^2+2\beta^2]}\times J_0(2\sqrt{\alpha\beta}\,\rho)e^{-i\alpha\zeta}e^{i\beta\eta}. \quad (3.43)$$

This is a double integral of the Laplace type and can be evaluated asymptotically for large σ^2. Following Bleistein and Handelsman,[40] the function

$$\phi(\alpha,\beta)=-[\omega_0^2+\lambda^2-2(\lambda+\omega_0)\alpha-2(\omega_0-\lambda)\beta+2\alpha^2+2\beta^2] \quad (3.44)$$

has critical points at $\phi_\alpha=\phi_\beta=0$, or at

$$\alpha_0=(\omega_0+\lambda)/2,\quad \beta_0=(\omega_0-\lambda)/2. \quad (3.45)$$

Since $\phi_{\alpha\alpha}=\phi_{\beta\beta}=-4$ and $\phi_{\alpha\beta}=0$, it follows that

$$\phi_{\alpha\alpha}(\alpha_0,\beta_0)<0,\quad \phi_{\beta\beta}(\alpha_0,\beta_0)<0,$$

$$\phi_{\alpha\alpha}(\alpha_0,\beta_0)\phi_{\beta\beta}(\alpha_0,\beta_0)-\phi_{\alpha\beta}^2(\alpha_0,\beta_0)=16>0,$$

and the critical point given by (3.45) is a maximum. Hence the integration (3.43) can be approximated by

$$\Psi(\rho,\zeta,\eta)=\frac{2\pi\sigma^{-2}e^{\sigma^2\phi(\alpha_0,\beta_0)}}{\sqrt{\phi_{\alpha\alpha}(\alpha_0,\beta_0)\phi_{\beta\beta}(\alpha_0,\beta_0)-\phi_{\alpha\beta}^2(\alpha_0,\beta_0)}}\times\frac{2\sigma^2}{\pi}J_0(2\sqrt{\alpha_0\beta_0}\rho)e^{-i\alpha_0\zeta}e^{i\beta_0\eta}+O(\sigma^{-2}).$$

Rearranging the terms and using Eq. (3.45), one gets

$$\Psi(\rho,\zeta,\eta)=J_0[\sqrt{\omega_0^2-\lambda^2}\,\rho]\times e^{-i\lambda(\eta+\zeta)/2}e^{i\omega_0(\eta-\zeta)/2}+O(\sigma^{-2}), \quad (3.46)$$

which in the limit $\sigma\to\infty$ reduces to

$$\Psi(\rho,\zeta,\eta)=J_0[\sqrt{\omega_0^2-\lambda^2}\,\rho]\,e^{-i\lambda(\eta+\zeta)/2}e^{i\omega_0(\eta-\zeta)/2}. \quad (3.47a)$$

Although the wave function given in (3.47a) was obtained from the asymptotic evaluation of the double integration (3.43), it turns out to be an exact solution to the scalar wave equation. In fact, it is same as Durnin's Bessel beam, which can be obtained by substituting

$$\zeta=z-t,\quad \eta=z+t$$

into Eq. (3.47a) and rewriting it as follows:

$$\Psi(\rho,\zeta,\eta)=J_0[\sqrt{\omega_0^2-\lambda^2}\,\rho]\,e^{-i(\lambda z-\omega_0 t)}. \quad (3.47b)$$

As in the case of Brittingham's FWM's, the spectrum associated with a Bessel beam is singular; specifically, the spectrum given in (3.40) reduces to a product of two Dirac delta functions as σ goes to ∞. On the other hand, the conversion to a Fourier picture is trivial in this case since (3.47b) is totally traveling in the positive z direction.

IV. EXTENSIONS OF THE BIDIRECTIONAL SYNTHESIS

In this section, we shall extend the ideas discussed in Sec. II to other classes of problems. The most natural extension is an application involving the three-dimensional Klein–Gordon equation which describes the propagation of waves in a dispersive medium. Another one deals with the use of the bidirectional representation in connection with dissipative problems modeled, for example, by the three-dimensional dissipative scalar wave equation and the telegraph equation; in these cases the operator $\hat{\Omega}(-i\nabla)$ is nonpositive. Using these two classes of problems, we shall show that solutions obtained via the bidirectional representation will be as easy to evaluate asymptotically or numerically as those for the three-dimensional wave equation. By virtue of this observation, new exact solutions can be obtained trivially using spectra similar to those in Sec. III.

The spectral analysis in Sec. II was carried out over the β-dependent part of the integral in Eq. (2.27). We were led to this procedure because the three-dimensional wave equation has the same structure as the one-dimensional Klein–Gordon equation. In particular, a Fourier transformation with respect to the transverse coordinates x and y reduces the three-dimensional wave equation to a one-dimensional Klein–Gordon equation of the following form:

$$[\partial_t^2-\partial_z^2+\kappa^2]\tilde{u}(\kappa,z,t)=0. \quad (4.1)$$

It should be observed that the functions $[I_n(\kappa z)J_m(\kappa t)]$ in the expression (2.32a) are not solutions to the one-dimensional Klein–Gordon equation. Only their sums over integer values of m and n constitute a solution to (4.1) and a delicate balance between the coefficients of $[I_n(\kappa z)J_m(\kappa t)]$ must be maintained in order to give finite solutions.

A natural extension is the three-dimensional Klein–Gordon equation describing the evolution of a signal propagating in dispersive media. For this case, the operator $\hat{\Omega}(-i\nabla)$ equals $-\nabla^2+\mu^2$ and Eq. (2.1a) takes the form

$$[\partial_t^2-\nabla^2+\mu^2]u(\mathbf{r},t)=0. \quad (4.2)$$

A general solution to this equation is analogous to that given by (2.18), namely,

$$u(\mathbf{r},t)=2\,\mathrm{Re}\{\Psi(\mathbf{r},t)\},$$

where $\Psi(\mathbf{r},t)$ can be represented by the following bidirectional superposition:

$$\Psi(\rho,\zeta,\eta)=\frac{1}{(2\pi)^2}\sum_{\substack{l=-1\\ l\neq 0}}^{+1}\sum_{n=0}^{\infty}\int_0^\infty d\kappa\int_0^\infty d(l\alpha)\int_0^\infty \times d(l\beta)C_n(l\alpha,l\beta,\kappa)\kappa J_n(\kappa\rho)\times e^{\pm in\phi}e^{-il\alpha\zeta}e^{il\beta\eta}\delta\left[\alpha\beta-\frac{1}{4}(\kappa^2+\mu^2)\right]. \quad (4.3)$$

In this case, a partitioning of $\hat{\Omega}(-i\nabla)$ was induced through the operators $\hat{A}(-i\partial_z)=-\partial_z^2$, $\hat{B}(-\kappa-i\partial_z)=\kappa^2+\mu^2$, and the new constraint relation is given by

$$\alpha\beta = \tfrac{1}{4}(\kappa^2 + \mu^2)\,. \tag{4.4}$$

Using the relationship (2.23), the representation (4.3) can be transformed into the conventional Fourier picture; specifically,

$$\Psi(\mathbf{r},t) = \frac{1}{(2\pi)^2}\sum_{n=0}^{\infty}\int_0^{\infty} d\kappa \int_{-\infty}^{+\infty} d\omega \times \int_{-\infty}^{+\infty} dk_z\, A_n(\omega,k_z,\kappa)\kappa J_n(\kappa\rho)e^{\pm in\phi} \times e^{-ik_z z}e^{+i\omega t}\delta(\omega^2 - \kappa^2 - k_z^2 - \mu^2)\,. \tag{4.5}$$

The only difference between (4.5) and (2.18) is the more complicated constraint relationship. For the problem under consideration, the constraint requires that

$$\omega^2 - \kappa^2 - k_z^2 = \mu^2\,, \tag{4.6}$$

which recovers the well known energy relation $E^2 = p^2 + \mu^2$. Recall that very few exact solutions to the three-dimensional Klein–Gordon equation are available. In this sense, the representation (4.3) is very valuable because it is characterized by the same algebraic singularities as (2.22). As a consequence, (4.3) allows the analytical computation of a rich class of novel exact and approximate solutions with as much facility as shown in Sec. III for the three-dimensional scalar wave equation. For example, all the spectra used in Sec. III can be used trivially to reproduce new solutions to the three-dimensional Klein–Gordon equation.

For physical situations requiring a nonpositive operator $\hat{\Omega}(-i\nabla)$, e.g., those modeled by the dissipative scalar wave equation and the telegraph equation, one can still obtain novel, exact solutions using the bidirectional synthesis procedure. Along these lines, consider the three-dimensional dissipative scalar wave equation

$$[\partial_t^2 - \nabla^2 + (c_1 + c_2)\partial_t + c_1c_2]\Psi(\mathbf{r},t) = 0\,, \tag{4.7}$$

which describes a wave traveling in a dissipative medium. Although Eq. (4.7) has a different structure than Eq. (2.1), an exponential transformation of the form

$$\Psi(\mathbf{r},t) = \exp[-\tfrac{1}{2}(c_1 + c_2)t]\,\hat{\Psi}(\mathbf{r},t) \tag{4.8}$$

reduces it to

$$[\partial_t^2 - \nabla^2 - \tfrac{1}{2}(c_1 - c_2)^2]\hat{\Psi}(\mathbf{r},t) = 0\,, \tag{4.9}$$

which is a special case of Eq. (2.1) with $\hat{\Omega}(-i\nabla) = -\nabla^2 - (c_1 - c_2)^2/2$. Notice that the above equation is similar to the Klein–Gordon equation (4.2) with an imaginary μ [i.e., $\mu^2 = -(c_1 - c_2)^2/2$]. The bidirectional representation can be written directly as

$$\hat{\Psi}(\rho,\zeta,\eta) = \frac{1}{(2\pi)}\sum_{\substack{l=-1\\ l\neq 0}}^{+1}\sum_{n=0}^{\infty}\int_0^{\infty} d\kappa \int_0^{\infty} d(l\alpha) \times \int_{-\infty}^{0} d(l\beta)\, C_n(l\alpha,l\beta,\kappa)\kappa J_n(\kappa\rho) \times e^{\pm in\phi}e^{-il\alpha\zeta}e^{il\beta\eta}\delta\left[\alpha\beta + \frac{1}{8}(c_1 - c_2)^2 - \frac{\kappa^2}{4}\right], \tag{4.10}$$

with the constraint

$$\alpha\beta = -\tfrac{1}{8}(c_1 - c_2)^2 + \kappa^2/4\,. \tag{4.11}$$

The same discussion concerning the nature of the singularities of this solution follows automatically, except for the fact that the hyperbolic constraint (4.11) can lie in the second or the fourth quadrants of the $\alpha\beta$ plane for $(c_1 - c_2)^2/2 > \kappa^2$, and in the first or third quadrants for $(c_1 - c_2)^2/2 < \kappa^2$. This also explains the difference in the α and β limits of the integration in Eq. (4.10).

The dissipative wave equation can be reduced to the telegraph equation by removing the dependence of Eq. (4.7) on the transverse coordinates x and y. The telegraph equation, which can be written as

$$[\partial_t^2 - \partial_z^2 + (c_1 + c_2)\partial_t + c_1c_2]\psi(z,t) = 0\,, \tag{4.12}$$

models the transmission of electromagnetic signals through wire cables. Using an exponential transformation of the form

$$\psi(z,t) = \exp[-\tfrac{1}{2}(c_1 + c_2)t]\,\hat{\psi}(z,t)\,, \tag{4.13}$$

reduces the telegraph equation to

$$[\partial_t^2 - \partial_z^2 - \tfrac{1}{2}(c_1 - c_2)^2]\,\hat{\psi}(z,t) = 0\,. \tag{4.14}$$

A celebrated solution due to Lord Kelvin involves the choice $c_1 = c_2$. This restriction reduces Eq. (4.14) to a one-dimensional scalar wave equation that has the distortion-free solutions $\hat{\psi}(z-t)$ and $\hat{\psi}(z+t)$. In an attempt to find solutions to Eq. (4.14) in the general case where $c_1 \neq c_2$, one runs into the same complications as those discussed earlier in connection to the Fourier representation of the one-dimensional Klein–Gordon equation, or the three-dimensional scalar wave equation. An alternative is to use the bidirectional representation

$$\hat{\psi}(\zeta,\eta) = \frac{1}{(2\pi)}\sum_{\substack{l=-1\\ l\neq 0}}^{+1}\int_0^{\infty} d(l\alpha)\int_{-\infty}^{0} d(l\beta)\, c_0(\alpha,\beta) \times e^{-il\alpha\zeta}e^{il\beta\eta}\delta\left[\alpha\beta + \frac{1}{8}(c_1 - c_2)^2\right], \tag{4.15}$$

with the constraint

$$\alpha\beta = -\tfrac{1}{8}(c_1 - c_2)^2\,. \tag{4.16}$$

(Only the second or the fourth quadrants of the $\alpha\beta$ plane need be used in this case since α and β must have different signs.) The Fourier synthesis corresponding to (4.15) can be obtained by using the transformation (2.23). This leads to

$$\hat{\psi}(z,t) = \frac{1}{2\pi}\int_{-\infty}^{+\infty} d\omega \int_{-\infty}^{+\infty} dk_z\, A_n(\omega,k_z) \times e^{-ik_z z}e^{i\omega t}\delta\left(\omega^2 - k_z^2 + \frac{1}{2}(c_1 - c_2)^2\right), \tag{4.17}$$

with the constraint

$$\omega^2 - k_z^2 = -\tfrac{1}{4}(c_1 - c_2)^2\,. \tag{4.18}$$

One should not get the wrong impression that the method introduced in this paper will replace the Fourier synthesis; on the contrary, the bidirectional synthesis complements it. As it was shown in Eq. (2.17), a Fourier decomposition is simply a special case of a general partitioning of the operator L. In many instances dealing with single frequency phenomena, the Fourier synthesis is the most intuitive one; however, this does not rule out all other representations, particularly

when they can lead to new exact solutions. Consider, for instance, initial value problems. Even though the bidirectional representation is characterized by an implicit dependence on time through the variables ζ and η, an initial value problem can still be handled successfully. As an example, consider the[AV: specific problem of pulse propagation through an infinitely long cylindrical waveguide. This problem is modeled by the three-dimensional scalar wave equation

$$(\nabla^2 - \partial_t^2)u(\mathbf{r},t) = 0,$$

with the initial conditions

$$u(\mathbf{r},0) = F(\rho,z), \tag{4.19a}$$

$$u_t(\mathbf{r},0) = G(\rho,z), \tag{4.19b}$$

and the boundary condition

$$u(R,z,t) = 0, \tag{4.19c}$$

where $\rho = R$ is the radius of the cross section of the waveguide. The functions $F(\rho,z)$ and $G(\rho,z)$ are assumed to be real. For this problem, it is advantageous to begin with the expression (2.27). A typical solution can then be written as

$$u(\mathbf{r},t) \equiv u(\rho,\zeta,\eta) = c_0(\kappa,\beta)J_0(\kappa\rho)e^{-(i\kappa^2/4\beta)\zeta}e^{i\beta\eta}. \tag{4.20}$$

Applying the boundary condition (4.19c), one obtains $J_0(\kappa R) = 0$. It immediately follows that $\kappa R = \kappa_{0m}$, where κ_{0m} are the zeros of the zeroth-order Bessel function. By summing over all modes and integrating over β, the general waveguide solution can be given as

$$u(\rho,\zeta,\eta) = \mathrm{Re}\sum_{m=1}^{\infty}\int_0^{\infty} d\beta\, c_0(\kappa_{0m},\beta) \times J_0\left(\frac{\kappa_{0m}\rho}{R}\right)e^{-i(\kappa_{0m}^2/4\beta R^2)\zeta}e^{i\beta\eta}. \tag{4.21}$$

The initial condition (4.19a) is satisfied if

$$F(\rho,z) = \mathrm{Re}\sum_{m=1}^{\infty}\int_0^{\infty} d\beta\, c_0(\kappa_{0m},\beta) \times J_0\left(\frac{\kappa_{0m}\rho}{R}\right)e^{-i(\kappa_{0m}^2/4\beta R^2 - \beta)z}. \tag{4.22}$$

The spectrum $c_0(\kappa_{0m},\beta)$, which is, in general, a complex function of κ_{0m} and β, can be determined by taking first the Fourier transform with respect to z and then the Hankel transform with respect to ρ in Eq. (4.22). This gives

$$f(\kappa_{0m},k_z) = \int_0^{\infty} d\beta \frac{R^2}{2}[J_1(\kappa_{0m})]^2 \int_{-\infty}^{+\infty}\frac{dz}{2} \times \left[c_0(\kappa_{0m},\beta)e^{-i(\kappa_{0m}^2/4\beta R^2 - \beta - k_z)z} + c_0^*(\kappa_{0m},\beta)e^{i(\kappa_{0m}^2/4\beta R^2 - \beta + k_z)z}\right], \tag{4.23}$$

where $c_0^*(\kappa_{0m},\beta)$ is the Hermitian conjugate of $c_0(\kappa_{0m},\beta)$, and $f(\kappa_{0m},k_z)$ is defined as

$$f(\kappa_{0m},k_z) = \int_{-\infty}^{+\infty} dz \int_{\infty}^{R} d\rho\, \rho J_0\left(\frac{\kappa_{0m}\rho}{R}\right)F(\rho,z)e^{+ik_z z}. \tag{4.24}$$

By integrating the right-hand side of Eq. (4.23) over z, it follows that

$$f(\kappa_{0m},k_z) = \frac{\pi}{2}R^2[J_1(\kappa_{0m})]^2\int_0^{\infty} d\beta \times\left[c_0(\kappa_{0m},\beta)\delta\left(\frac{\kappa_{0m}^2}{4\beta R^2} - \beta - k_z\right) + c_0^*(\kappa_{0m},\beta)\delta\left(\frac{\kappa_{0m}^2}{4\beta R^2} - \beta + k_z\right)\right]. \tag{4.25}$$

By performing, finally, the integration over β, the following relation is obtained:

$$\frac{c_0(\kappa_{0m},\beta_1)}{\sqrt{k_z^2 + (\kappa_{0m}/R)^2}}\beta_1 + \frac{c_0^*(\kappa_{0m},\beta_2)}{\sqrt{k_z^2 + (\kappa_{0m}/R)^2}}\beta_2 = \frac{2f(\kappa_{0m},k_z)}{\pi R^2[J_1(\kappa_{0m})]^2}. \tag{4.26}$$

Here,

$$\beta_1 = \tfrac{1}{2}\left[-k_z + \sqrt{(k_z^2 + (\kappa_{0m}/R)^2)}\right],$$

$$\beta_2 = \tfrac{1}{2}\left[+k_z + \sqrt{(k_z^2 + (\kappa_{0m}/R)^2)}\right].$$

It turns out that the initial condition (4.19b) is satisfied if

$$c_0(\kappa_{0m},\beta_1)\beta_1 - c_0^*(\kappa_{0m},\beta_2)\beta_2 = 2g(\kappa_{0m},k_z)/\pi R^2[J_1(\kappa_{0m})]^2, \tag{4.27}$$

where

$$g(\kappa_{0m},k_z) = \int_{-\infty}^{+\infty} dz \int_{\infty}^{R} d\rho\, \rho J_0\left(\frac{\kappa_{0m}\rho}{R}\right)G(\rho,z)e^{+ik_z z}. \tag{4.28}$$

A combination of Eqs. (4.26) and (4.27) results in the spectrum

$$c_0(\kappa_{0m},\beta_1) = \frac{1}{\beta_1\pi R^2[J_1(\kappa_{0m})]^2}\left(f(\kappa_{0m},k_z) \times\sqrt{k_z^2 + \left(\frac{\kappa_{0m}}{R}\right)^2} + g(\kappa_{0m},k_z)\right). \tag{4.29}$$

The relation $\beta_1 \equiv \beta(k_z)$ must be inverted in order to obtain $k_z \equiv k_z(\beta)$. Eq. (4.29) can be written, then, as

$$c_0(\kappa_{0m},\beta) = \frac{1}{\beta\pi R^2[J_1(\kappa_{0m})]^2}\left(f(\kappa_{0m},k_z(\beta)) \times\sqrt{k_z^2(\beta) + \left(\frac{\kappa_{0m}}{R}\right)^2} + g(\kappa_{0m},k_z(\beta))\right)$$

and the solution to the original problem can be expressed as

$$u(\rho,\zeta,\eta) = \mathrm{Re}\sum_{m=1}^{\infty}\frac{1}{\pi R^2[J_1(\kappa_{0m})]^2} \times\int_0^{\infty} d\beta\, e^{-i(\kappa_{0m}^2/4\beta R^2)\zeta}e^{i\beta\eta}J_0\left(\frac{\kappa_{0m}\rho}{R}\right) \times\frac{1}{\beta}\left(f(\kappa_{0m},k_z(\beta))\sqrt{k_z^2(\beta) + \left(\frac{\kappa_{0m}}{R}\right)^2} + g(\kappa_{0m},k_z(\beta))\right), \tag{4.30}$$

in terms of a superposition of the elementary blocks $e^{-i\alpha\zeta}e^{i\beta\eta}$. Obviously, the shape of the field $u(\rho,\zeta,\eta)$ depends on the choice of $F(\rho,z)$ and $G(\rho,z)$. If, for example, $F(\rho,z)$ is chosen in the separable form

$$F(\rho,z) = F_1(\rho)F_2(z)\,,$$

and

$$G(\rho,z) = 0\,,$$

Eq. (4.30) can be rewritten as

$$u(\rho,\zeta,\eta) = \mathrm{Re}\sum_{m=1}^{\infty}\frac{1}{\pi R^2[J_1(\kappa_{0m})]^2} \times\int_0^{\infty}d\beta\frac{\hat{F}_2(k_z(\beta))}{\beta} \times\left[k_z^2(\beta)+\left(\frac{\kappa_{0m}}{R}\right)^2\right]^{1/2}e^{-i(\kappa_{0m}^2/4\beta R^2)\zeta} \times e^{i\beta\eta}J_0\left(\frac{\kappa_{0m}\rho}{R}\right)\int_0^R d\rho'\,\rho' F_1(\rho')J_0\left(\frac{\kappa_{0m}}{R}\rho'\right), \tag{4.31}$$

where $\hat{F}_2(k_z)$ is the Fourier transform of $F_2(z)$. To be more specific, let

$$F_1(\rho) = (1/4\pi)J_0(\kappa_{0m}\rho/R)\,,$$

$$F_2(z) = K_0\left[(\kappa_{0m}/R)\sqrt{a^2+z^2}\,\right],$$

where K_0 is the zeroth-order modified Bessel function of the second kind. The initial conditions in this case have the form

$$u(\mathbf{r},0) = \frac{1}{4\pi}J_0\left(\frac{\kappa_{0m}\rho}{R}\right)K_0\left[\frac{\kappa_{0m}}{R}\sqrt{a^2+z^2}\,\right], \tag{4.32a}$$

$$u_t(\mathbf{r},0) = 0\,. \tag{4.32b}$$

The Fourier transform of the function $F_2(z)$, required in Eq. (4.31), is given in this case by

$$\hat{F}_2(k_z) = \frac{\pi}{\sqrt{k_z^2+\left(\frac{\kappa_{0m}}{R}\right)^2}}e^{-a\sqrt{k_z^2+(\kappa_{0m}/R)^2}}. \tag{4.33}$$

The expression for the root β_1, viz.,

$$(2\beta - k_z)^2 = k_z^2 + \kappa_{0m}^2/R^2\,,$$

can be used to invert $k_z = k_z(\beta)$. This leads to the relations

$$k_z = +\kappa_{0m}^2/4\beta R^2 - \beta\,, \tag{4.34}$$

$$k_z^2 + \kappa_{0m}^2/R^2 = (\kappa_{0m}^2/4\beta R^2+\beta)^2\,. \tag{4.35}$$

Equations (4.33) and (4.35) can be used in conjunction with (4.31) to obtain

$$u(\rho,\zeta,\eta) = \mathrm{Re}\sum_{m=1}^{\infty}\frac{J_0(\kappa_{0m}\rho/R)}{\pi R^2[J_1(\kappa_{0m})]^2}\int_0^{\infty}d\beta\frac{\pi}{\beta} \times e^{-(\kappa_{0m}^2/4\beta R^2+\beta)a}e^{-i(\kappa_{0m}^2/4\beta R^2)\zeta}e^{i\beta\eta} \times\int_0^R d\rho'\frac{\rho'}{4\pi}J_0\left(\frac{\kappa_{0m}}{R}\rho'\right)J_0\left(\frac{\kappa_{0m}}{R}\rho'\right). \tag{4.36}$$

By integrating over ρ', Eq. (4.36) simplifies to

$$u(\rho,\zeta,\eta) = \mathrm{Re}\left\{\int_0^{\infty}\frac{d\beta}{8\pi\beta} \times e^{-(\kappa_{0m}^2/4\beta R^2)(a+i\zeta)}e^{-\beta(a-i\eta)}J_0\left(\frac{\kappa_{0m}\rho}{R}\right)\right\}. \tag{4.37}$$

The remaining integration over β can be carried out using equation (3.478.4) in Gradshteyn and Ryzhik.[39] The solution to the initial boundary value problem under consideration then assumes the form

$$u(\mathbf{r},t) = \mathrm{Re}\left\{\frac{1}{4\pi}K_0\left[\frac{\kappa_{0m}}{R} \times\sqrt{(a+i\zeta)(a-i\eta)}\,\right]J_0\left(\frac{\kappa_{0m}\rho}{R}\right)\right\}. \tag{4.38}$$

An interesting variation of the solution given in (4.38) can be obtained by simply replacing the single parameter a by two parameters a_1 and a_2, namely,

$$u(\mathbf{r},t) = \mathrm{Re}\left\{\frac{1}{4\pi}K_0\left[\frac{\kappa_{0m}}{R} \times\sqrt{(a_1+i\zeta)(a_2-i\eta)}\,\right]J_0\left(\frac{\kappa_{0m}\rho}{R}\right)\right\}. \tag{4.39}$$

Because of the asymmetric dependence on the values of a_1 and a_2, the pulse given in Eq. (4.39) can be made to travel mainly in one direction. On the other hand, the solution (4.38) represents a pulse that will split into two halves propagating in opposite directions. Such claims can be verified by referring to Eqs. (3.38) and (3.39) which give the positive-going and the negative-going components of the splash pulse. The only difference entails the replacement of the integration over κ by a summation over κ_{0m}. The backward and forward spectra have the following ratio:

$$F(k_2,\kappa_{0m})/F(-k_2,\kappa_{0m}) = e^{(a_2-a_1)k_z}.$$

It is seen that for $a_2 = a_1$, the positive and the negative components have the same strength. On the other hand, for $a_2 \gg 1$ and $a_1 \ll 1$, $F(k_z,\ \kappa_{0m}) \gg F(-k_z,\ \kappa_{0m})$ and the pulse moves predominantly in the positive z direction. Moreover, in contradistinction to the splash pulse (3.19), the solution (4.39) is localized by the walls of the waveguide and one does not have to worry about its localization in the transverse direction.

V. CONCLUDING REMARKS

A novel bidirectional decomposition of solutions to partial differential equations into backward and forward traveling plane waves was introduced in this paper. This technique is distinct from other factorization methods available in the literature (cf., e.g., Ref. 41). The main difference stems from the fact that it involves a product of plane waves propagating in opposite directions, while usual factorization techniques decompose the solutions into a sum of forward and backward traveling plane waves. The bidirectional decomposition, which was developed within the framework of a more general embedding procedure, allows the construction of general solutions by means of a superposition of elementary bidirectional blocks. Such a novel superposition differs sig-

nificantly from the more conventional ones, e.g., the Fourier synthesis. In particular, it is characterized by algebraic singularities that can be much easier to handle than the branch-cut singularities arising usually in the Fourier synthesis. In spite of these differences, there is a one-to-one correspondence between the new synthesis and the Fourier method and these two methods complement each other.

Several mathematical aspects of the new synthesis were addressed. It was shown that the elementary blocks entering into this superposition are composed of exponential and Bessel functions which form complete sets of orthogonal functions. This led to an inversion formula, from which different spectra can be calculated from the knowledge of exact solutions. Necessary conditions for the choice of the spectra leading to convergent solutions were discussed.

The bidirectional decomposition was applied to the three-dimensional scalar wave equation, the three-dimensional Klein–Gordon equation, the three-dimensional dissipative wave equation, and the telegraph equation. For all these equations, it was demonstrated that new, exact solutions can be easily obtained. It was noted that the new synthesis provides the most natural basis for the construction of the unusual Brittingham-like solutions and that it can be used as a vehicle to find the Fourier spectral content of these solutions in order to gain a better understanding of their properties. Finally, it was shown that initial-boundary value problems can be solved using the bidirectional decomposition. A specific example was worked out in connection to an infinite waveguide and new solutions [cf. Eqs. (4.38) and (4.39)] were derived. These pulse solutions, especially (4.39), exhibit unusual decay patterns as they propagate down the waveguide. A detailed analysis of their properties has been published elsewhere.[42] A natural extension of this problem is the case of the open waveguide whose aperture is illuminated by the pulse given in Eq. (4.39). The solution to this problem is incorporated in Ref. 42.

ACKNOWLEDGMENTS

This work was performed in part by the Lawrence Livermore National Laboratory under the auspices of the U.S. Department of Energy under Contract No. W-7405-ENG-48.

[1]J. N. Brittingham, J. Appl. Phys. **54**, 1179 (1983).
[2]R. W. Ziolkowski, J. Math. Phys. **26**, 861 (1985).
[3]P. A. Belanger, J. Opt. Soc. Am. A **1**, 723 (1984).
[4]A. Sezginer, J. Appl. Phys. **57**, 678 (1985).
[5]P. A. Belanger, J. Opt. Soc. Am. A **3**, 541 (1986).
[6]R. W. Ziolkowski, Proc. Joint IEEE AP-S/URSI Symposium, Blacksburg, VA, June 1987.
[7]R. W. Ziolkowski and E. Heyman, Proc. URSI General Assembly, Tel Aviv, Israel, August 1987.
[8]R. W. Ziolkowski, Proc. SPIE Conference 873, Los Angeles, CA, January 1988.
[9]R. W. Ziolkowski, Phys. Rev. A **39**, 2005 (1989).
[10]P. Hillion, J. Appl. Phys. **60**, 2981 (1986).
[11]P. Hillion, J. Math. Phys. **28**, 1743 (1987).
[12]T. T. Wu, J. Appl. Phys. **57**, 2370 (1985).
[13]T. T. Wu, R. W. P. King, and H. M. Shen, J. Appl. Phys. **62**, 4036 (1987).
[14]H. M. Lee, Radio Sci. **22**, 1102 (1987).
[15]H. M. Shen, Proc. SPIE Conference 873, Los Angeles, CA, January 1988.
[16]H. M. Lee, Proc. Joint IEEE AP-S/URSI Symposium, Blacksburg, VA, June 1987.
[17]J. A. Stratton, *Electromagnetic Theory* (McGraw-Hill, New York, 1941), pp. 360 and 361.
[18]A. G. van Nie, Philips Res. Rep. **19**, 378 (1964).
[19]P. M. Morse and H. Feshbach, *Methods of Theoretical Physics* (McGraw-Hill, New York, 1953), Vol. I, Chap. 7.
[20]J. Durnin, J. Opt. Soc. Am. A **4**, 651 (1987).
[21]J. Durnin, J. J. Miceli, Jr., and J. H. Eberly, Phys. Rev. Lett. **58**, 1499 (1987).
[22]D. DeBeer, S. R. Hartmann, and Friedberg, Phys. Rev. Lett. **59**, 2611 (1987).
[23]H. E. Moses, J. Math. Phys. **25**, 1905 (1984).
[24]H. E. Moses and R. T. Prosser, IEEE Trans. Antennas Propag. **AP-34**(2), 188 (1986).
[25]E. Heyman and L. B. Felsen, IEEE Trans. Antennas Propag. **AP-34**(8), 1062 (1986).
[26]E. Heyman and B. Z. Steinberg, J. Opt. Soc. Am. A **4**, 473 (1987).
[27]P. D. Einziger and S. Raz, J. Opt. Soc. Am. A **4**, 3 (1987).
[28]E. Heyman, B. Z. Steinberg, and L. P. Felsen, J. Opt. Soc. Am. A **4**, 2081 (1987).
[29]L. P. Felsen and E. Heyman, Proc. SPIE Conference 873, Los Angeles, CA, January 1988.
[30]G. A. Deschamps, Electron. Lett. **7**, 684 (1971).
[31]L. P. Felsen, Symp. Math. **18**, 39 (1976).
[32]I. M. Besieris and A. M. Shaarawi, Proc. Joint IEEE AP-S/URSI Symposium, Blacksburg, VA, June 1987.
[33]L. Hormander, *The Analysis of Linear Partial Differential Operators Vol. III* (Springer, New York, 1985).
[34]L. Brillouin, *Wave Propagation and Group Velocity* (Academic, New York, 1960).
[35]L. B. Felsen and N. Marcuvitz, *Radiation and Scattering of Waves* (Prentice–Hall, Englewood Cliffs, NJ, 1973).
[36]R. M. Lewis, Arch. Ratl. Mech. Anal. **20**, 191 (1965).
[37]L. B. Felsen, in *Transient Electromagnetic Fields*, edited by L. B. Felsen (Springer, New York, 1976).
[38]N. D. Hoc, I. M. Besieris, and M. E. Sockell, IEEE Trans. Antennas Propag. **AP-33**, 1237 (1985).
[39]I. S. Gradshteyn and I. M. Ryzhik, *Tables of Integrals, Series and Products* (Academic, New York, 1965).
[40]N. Bleistein and R. A. Handelsman, *Asymptotic Expansions of Integrals* (Dover, New York, 1986), Chap. 8.
[41]V. H. Weston, J. Math. Phys. **27**, 1722 (1986).
[42]A. M. Shaarawi, I. M. Besieris, and R. W. Ziolkowski, J. Appl. Phys. **65**, 805 (1989).

On the uniqueness of prescriptions for spurious poles in Feynman integrals

Harald Skarke
Institut für Theoretische Physik, Technische Universität Wien, Karlsplatz 13, A-1040 Vienna, Austria

(Received 13 September 1988; accepted for publication 25 January 1989)

It is shown that one loop Feynman integrals containing spurious poles $(kn)^{-1}$ are completely determined by consistency properties, if they are allowed to depend either on one nonlightlike or on two lightlike vectors. Although no ε prescription is given to the spurious poles, the results are identical either with the principal value or with the Leibbrandt/Mandelstam results.

I. INTRODUCTION

Since axial-type gauges $f(nA)=0$ have been used for the first time,[1] the amount of interest in these gauges has grown considerably. They are in some way more "physical" than covariant gauges, ghosts decouple in homogeneous axial-type gauges, and the light-cone gauge has had some fundamental applications in supersymmetry[2] and string theory.[3] The major disadvantages are the loss of manifest Lorentz covariance and the occurrence of so-called "spurious poles" $(kn)^{-1}$ in the propagators. In handling these poles in Feynman graph calculations it is vital to distinguish between the cases $n^2\neq 0$ (axial gauge) and $n^2=0$ (light-cone gauge). For the axial gauge the principal value (PV) prescription

$$\left(\frac{1}{kn}\right)_{\mathrm{PV}}=\frac{1}{2}\left(\frac{1}{kn+i\varepsilon}+\frac{1}{kn-i\varepsilon}\right) \tag{1}$$

has been used successfully in one loop integrals,[4] but it is unsuitable for $n^2=0$. For this case a second lightlike vector has been introduced in the two equivalent prescriptions

$$\left(\frac{1}{kn}\right)_{\mathrm{L}}=\frac{kn^*}{knkn^*+i\varepsilon},\qquad \left(\frac{1}{kn}\right)_{\mathrm{M}}=\frac{1}{kn+i\varepsilon\ \mathrm{sgn}\ kn^*}, \tag{2}$$

suggested by Leibbrandt[5] and Mandelstam[2] (LM). The aim of this paper is to show that, by allowing the values of Feynman integrals to depend either on one vector n_μ with $n^2\neq 0$ or on two vectors n,n^* with $n^2=(n^*)^2=0$, certain basic consistency properties of integrals (without assuming any sort of ε-type prescription for spurious poles) imply unique results identical with the PV or LM results.

II. THEORY

Our considerations take place in a 2ω-dimensional Minkowski space with a $+\ -\ -\cdots$ metric. A Feynman integrand will always consist of a Lorentz-invariant "Feynman-type" part and of spurious poles. If we restrict ourselves at first to a single simple spurious pole and combine all Feynman denominators with the usual Feynman trick, the most general one loop integral will be

$$I_{\mu_1\cdots\mu_r}(p,L;n,\alpha)=\int d^{2\omega}k\,\frac{k_{\mu_1}\cdots k_{\mu_r}}{(k^2-2pk-L)^\alpha}\frac{1}{kn}. \tag{3}$$

The ω dependence will be suppressed in our notation throughout this paper. We make the following natural assumptions on the properties of $I_{\mu_1\cdots\mu_r}$:

(a) correct mass dimension $2\omega-2\alpha+r-1$;

(b) compatibility with differentiation with respect to p and L:

$$\frac{\partial I_{\mu_1\cdots\mu_r}(p,L;n,\alpha)}{\partial p_{\mu_{r+1}}}=2\alpha I_{\mu_1\cdots\mu_{r+1}}(p,L;n,\alpha+1), \tag{4}$$

$$\frac{\partial I_{\mu_1\cdots\mu_r}(p,L;n,\alpha)}{\partial L}=\alpha I_{\mu_1\cdots\mu_r}(p,L;n,\alpha+1); \tag{5}$$

(c) homogeneity in n and, if another vector n^* is introduced, in n^*:

$$I_{\mu_1\cdots\mu_r}(p,L;\lambda n,\alpha)=(1/\lambda)I_{\mu_1\cdots\mu_r}(p,L;n,\alpha); \tag{6}$$

(d) compatibility with nonspurious Feynman integrals:

$$n_{\mu_r}I_{\mu_1\cdots\mu_r}(p,L;n,\alpha)=\int d^{2\omega}k\,\frac{k_{\mu_1}\cdots k_{\mu_{r-1}}}{(k^2-2pk-L)^\alpha}; \tag{7}$$

and (e) invariance under shifts of integration variables.

Because of (b) we can restrict ourselves to

$$I(p,L;n,\alpha)=\int d^{2\omega}k\,\frac{1}{(k^2-2pk-L)^\alpha}\frac{1}{kn} \tag{8}$$

and determine $I_{\mu_1\cdots\mu_r}$ by differentiation with respect to p_μ. For $r\geqslant\alpha$ one has to assume analytic continuation in α. Thus (d) becomes

$$n^\mu\frac{\partial I(p,L;n,\alpha)}{\partial p^\mu}=2\alpha\int d^{2\omega}k\,\frac{1}{(k^2-2pk-L)^{\alpha+1}}. \tag{9}$$

Assumption (e) implies

$$I(p,L;n,\alpha)=\int d^{2\omega}k\,\frac{1}{(k^2-p^2-L)^\alpha}\frac{1}{kn+pn}, \tag{10}$$

showing that $I(p,L;n,\alpha)$ depends on L and p only via $L+p^2$ and those components of p on which $(kn+pn)^{-1}$ depends, namely, pn for $n^2\neq 0$ and pn and pn^* for $n^2=0$.

Let us first examine the axial gauge with $n^2\neq 0$. Then $I(p,L;n,\alpha)$ will depend on p^2+L,pn,n^2,α. Because of (c) it has to be of the form

$$I(p,L;n,\alpha)=(pn/n^2)G\left((pn)^2/n^2,L+p^2;\alpha\right), \tag{11}$$

with a function G that we want to determine. Equation (9) yields

$$G+2\frac{(pn)^2}{n^2}\left(\frac{\partial G}{\partial((pn)^2/n^2)}+\frac{\partial G}{\partial(L+p^2)}\right)=2\alpha\,\bar{I}\,(L+p^2,\alpha+1), \tag{12}$$

with

 0022-2488/89/061270-03$02.50

$$\bar{I} \equiv \int d^{2\omega}k \frac{1}{(k^2 - 2pk - L)^\alpha}$$
$$= i\pi^\omega (-1)^\alpha \frac{\Gamma(\alpha - \omega)}{\Gamma(\alpha)} (L + p^2)^{\omega - \alpha}. \quad (13)$$

This is an inhomogeneous first-order partial differential equation. Since we know that the PV of $I(p,L;n,\alpha)$ fulfills conditions (a)–(e),

$$G_{\text{part}} = (n^2/pn) I_{\text{PV}}(p,L;n,\alpha) \quad (14)$$

is a particular solution to Eq. (12). The remaining homogeneous equation is easily solved:

$$G_{\text{hom}} = \left(\frac{(pn)^2}{n^2}\right)^{-1/2} f\left(L + p^2 - \frac{(pn)^2}{n^2}\right), \quad (15)$$

where f is *a priori* an arbitrary function. Condition (a) implies

$$G_{\text{hom}} = \left(\frac{(pn)^2}{n^2}\right)^{-1/2} \lambda(\alpha) \left(L + p^2 - \frac{(pn)^2}{n^2}\right)^{\omega - \alpha - 1/2} \quad (16)$$

and therefore

$$I = I_{\text{PV}} + \frac{1}{\sqrt{n^2}} \lambda(\alpha) \left(L + p^2 - \frac{(pn)^2}{n^2}\right)^{\omega - \alpha - 1/2}. \quad (17)$$

To allow compatibility with differentiation with respect to L [cf. Eq. (5)], $\lambda(\alpha)$ must fulfill

$$(\omega - \alpha - \tfrac{1}{2})\lambda(\alpha) = \alpha\lambda(\alpha + 1). \quad (18)$$

In Ref. 6 the integral

$$I_{i\varepsilon}(p,L;n,\alpha) \equiv \int d^{2\omega}k \frac{1}{(k^2 - p^2 - L)^\alpha} \frac{1}{kn + i\varepsilon} \quad (19)$$

was calculated. It allows a calculation of PV integrals:

$$I_{\text{PV}}(p,L;n,\alpha)$$
$$= \tfrac{1}{2}(I_{i\varepsilon}(p,L;n,\alpha) - I_{i\varepsilon}(p,L;-n,\alpha)). \quad (20)$$

Then $I(p,L;n,\alpha)$ as in (17) with any λ that fulfills (18) is equal to

$$\mu I_{i\varepsilon}(p,L;n,\alpha) - (1 - \mu) I_{i\varepsilon}(p,L;-n,\alpha) \quad (21)$$

with some μ. Condition (c), however, implies that $I(p,L;n,\alpha)$ must be odd in n_μ. This is only satisfied for $\lambda = 0$ (corresponding to $\mu = \frac{1}{2}$), showing that the only permissible choice for I is I_{PV}.

Let us now turn our attention to the light-cone case. We have $n^2 = 0$ and allow our integrals to depend on another vector n^*_μ with $(n^*)^2 = 0$. Now $I(p,L;n,n^*,\alpha)$ will depend on $L + p^2$, pn, pn^*, nn^* and α. Homogeneity in n^* implies that I depends on n^* only via pn^*/nn^*. Because of (c) I must be of the form

$$(1/pn) G(pn\, pn^*/nn^*, L + p^2; \alpha). \quad (22)$$

Analogously to Eq. (12) we derive

$$\frac{\partial G}{\partial(pn\, pn^*/nn^*)} + 2\frac{\partial G}{\partial(L + p^2)} = 2\alpha \bar{I}(L + p^2; \alpha + 1). \quad (23)$$

With

$$\frac{\partial \bar{I}(L + p^2;\alpha)}{\partial(L + p^2)} = \alpha \bar{I}(L + p^2; \alpha + 1), \quad (24)$$

we see that

$$G_{\text{part}} = \bar{I}(L + p^2; \alpha) \quad (25)$$

is a particular solution. It corresponds to the PV prescription on the light cone.[7] The solution of the homogeneous equation is

$$G_{\text{hom}} = f(L + p^2 - 2pn\, pn^*/nn^*). \quad (26)$$

Thus (a) implies

$$G_{\text{hom}} = \lambda(\alpha)(L + p^2 - 2pn\, pn^*/nn^*)^{\omega - \alpha} \quad (27)$$

and compatibility with differentiation with respect to L yields

$$(\omega - \alpha)\lambda(\alpha) = \alpha\lambda(\alpha + 1) \quad (28)$$

and therefore

$$\lambda(\alpha) = \lambda'[\Gamma(\alpha - \omega)/\Gamma(\alpha)](-1)^\alpha. \quad (29)$$

We have

$$I(p,L;n,n^*,\alpha)$$
$$= (-1)^\alpha \pi^\omega \frac{\Gamma(\alpha - \omega)}{\Gamma(\alpha)} \frac{1}{pn}$$
$$\times \left[(L + p^2)^{\omega - \alpha} - \bar{\lambda}\left(L + p^2 - 2\frac{pn\, pn^*}{nn^*}\right)^{\omega - \alpha}\right], \quad (30)$$

with $\bar{\lambda} = -\pi^{-\omega}\lambda'$. For $\bar{\lambda} = 0$ we get the PV result and for $\bar{\lambda} = 1$ the LM result (see, e.g., Ref. 8). Thus

$$I(\bar{\lambda}) = \bar{\lambda} I_{\text{LM}} + (1 - \bar{\lambda}) I_{\text{PV}}. \quad (31)$$

All these values are internally consistent for noninteger ω. It is well known, however, that on the light cone the PV (but not the LM) results develop double poles in $(2 - \omega)$ and also poles proportional to $\ln p^2$ in the context of dimensional regularization, which cannot be treated within the usual renormalization schemes. Therefore in physical applications we must not allow our integrals to have PV-like parts and set $\bar{\lambda} = 1$. In this manner we have arrived again at a unique result for $I(p,L;n,n^*,\alpha)$.

Let us now examine integrals with one multiple spurious pole both on and off the light cone. It seems natural to define

$$I(p,L;n,\alpha,\beta) \equiv \int d^{2\omega}k \frac{1}{(k^2 - 2pk - L)^\alpha} \frac{1}{(kn)^\beta}$$
$$= \int d^{2\omega}k \frac{1}{(k^2 - L - p^2)^\alpha} \frac{1}{(kn + pn)^\beta}$$
$$\equiv \frac{(-1)^{\beta-1}}{\Gamma(\beta)} \frac{\partial^{\beta-1}}{\partial(pn)^{\beta-1}} \tilde{I}(L + p^2, pn, \cdots), \quad (32)$$

where $\tilde{I}(L + p^2, pn, \ldots) = I(L,p;n,\alpha)$ or $I(L,p;n,n^*,\alpha)$ as defined previously. Only in this definition multiple spurious poles can be regarded as limits of products of different single spurious poles (for a more detailed presentation of this argument see Ref. 9). We have full compatibility with a generalization of (d) to $\beta > 1$ and, again, we get the PV or LM results.

The simplest integral with different spurious poles is

$$J(p,q,L;n,\alpha)$$
$$= \int d^{2\omega}k \frac{l}{(k_2 - 2pk - L)^{\alpha}} \frac{1}{kn} \frac{1}{kn+qn}. \tag{33}$$

There are two consistency relations analogous to (d):

$$n^{\mu} \frac{\partial J(p,q,L;n,\alpha)}{\partial p^{\mu}}$$
$$= 2\alpha \int d^{2\omega}k \frac{1}{(k^2 - 2pk - L)^{\alpha+1}} \frac{1}{kn+qn},$$
$$2qn \frac{\partial J(p,q,L;n,\alpha)}{\partial L} + n^{\mu} \frac{\partial J(p,q,L;n,\alpha)}{\partial p^{\mu}} \tag{34}$$
$$= 2\alpha \int d^{2\omega}k \frac{1}{(k^2 - 2pk - L)^{\alpha+1}} \frac{1}{kn}.$$

Subtracting these two equations and keeping in mind that

$$\frac{\partial J(p,q,L;n,\alpha)}{\partial L} = \alpha J(p,q,L;n,\alpha+1), \tag{35}$$

we get

$$qnJ(p,q,L;n,\alpha+1)$$
$$= \int d^{2\omega}k \frac{1}{(k^2 - 2pk - L)^{\alpha+1}} \frac{1}{kn+qn}$$
$$- \int d^{2\omega}k \frac{1}{(k^2 - 2pk - L)^{\alpha+1}} \frac{1}{kn}. \tag{36}$$

This implies

$$J(p,q,L;n,\alpha)$$
$$= \frac{1}{qn}\left(\int d^{2\omega}k \frac{1}{(k^2 - 2pk - L)^{\alpha}} \frac{1}{kn+qn} - \int d^{2\omega}k \frac{1}{(k^2 - 2pk - L)^{\alpha}} \frac{1}{kn}\right)$$
$$+ \delta(qn) f(p,q,L;n,\alpha). \tag{37}$$

The $\delta(qn)$ term arises because division of Eq. (36) by qn is an allowed operation only for $qn \neq 0$. Thus, in the context of Feynman integrals,

$$\frac{1}{kn} \frac{1}{kn+qn} - \frac{1}{qn} \frac{1}{kn} + \frac{1}{qn} \frac{1}{kn+qn} = \delta(qn)\delta(kn)g, \tag{38}$$

where g has to be dimensionless and homogeneous in n and, if necessary, in n^*. Here $\delta(kn)$ arises because of the symmetry of the lhs of Eq. (38) between k and q. The δ-function terms in (37) and (38) only become meaningful in the context of integrals over more than one loop. Since the value of such an integral should not depend on the order of integrations and integrals with a single spurious pole are interpreted as PV or LM integrals, g has to take the same value as in the context of these prescriptions, namely, $g = \pi^2$ for $n^2 \neq 0$ and $g = 0$ for $n^2 = 0$.

III. CONCLUSION

Although we have used a higher loop argument here, we cannot give a complete analysis to all orders. The PV prescription is well known to exhibit difficulties in higher loop integrals because $[\mathrm{PV}((kn)^{-1})]^2 \neq \mathrm{PV}((kn)^{-2})$. It seems that our approach, namely, not giving any prescription at the level of propagators and interpreting integrals only in the context of consistency requirements, might be safer. One should recall here that we always interpret $(kn)^{-\beta}$ as in Eq. (32), no matter which propagators contribute to an expression.

It has been shown that mere consistency properties of one loop Feynman integrals with spurious poles determine the values of these integrals completely. If we allow our integrals to depend on only one nonlightlike vector n_{μ}, they turn out to be equal to the PV results, whereas those depending on two lightlike vectors have to be equal to the LM integrals only if we also demand the absence of nonsimple poles in dimensionally regularized integrals. Thus both prescriptions are unique in the contexts of their applications. If we allow our integrals, however, to depend on two unrestricted (not necessarily lightlike) vectors, this uniqueness is lost. Generalizations of the LM prescription[10] fulfill the requirements (a)–(e), but, using the methods of this paper, one finds that adding any function to these results that depends on p and L only via $nn^*pn - n^2pn^*$ and $n^2(L + p^2) - (np)^2$ and has the right dimensionality and homogeneity properties yields values that are also internally consistent.

ACKNOWLEDGMENTS

I would like to thank Professor W. Kummer and G. Pollak for careful readings of the manuscript and valuable suggestions.

[1] W. Kummer, Acta Phys. Austr. **14**, 149 (1961); R. L. Arnowitt and S. J. Fickler, Phys. Rev. **127**, 1821 (1962); J. Schwinger, *ibid.* **130**, 402 (1963); Y. P. Yao, J. Math. Phys. **5**, 1319 (1964).

[2] S. Mandelstam, Nucl. Phys. B **213**, 149 (1983).

[3] P. Goddard, J. Goldstone, C. Rebbi, and C. B. Thorn, Nucl. Phys. B **56**, 109 (1973).

[4] W. Kummer, Acta Phys. Austr. **41**, 315 (1975); W. Konetschny and W. Kummer, Nucl. Phys. B **100**, 106 (1975); **108**, 397 (1976); J. Frenkel, Phys. Rev. D **13**, 2325 (1976); W. Konetschny, *ibid.* **28**, 354 (1983).

[5] G. Leibbrandt, Phys. Rev. D **29**, 1699 (1984); **30**, 2167 (1984).

[6] W. Konetschny, Phys. Rev. D **28**, 354 (1983).

[7] H. C. Lee, lecture notes, AECL-8630, 1984.

[8] P. Gaigg, M. Kreuzer, M. Schweda, and O. Piguet, J. Math. Phys. **28**, 2781 (1987).

[9] H. Skarke and P. Gaigg, Phys. Rev. D **38**, 3205 (1988).

[10] P. Gaigg, M. Kreuzer, M. Schweda, and O. Piguet, J. Math. Phys. **28**, 2781 (1987); M. Kreuzer, G. Pollak, H. Skarke, and M. Schweda, Phys. Lett. B **196**, 557 (1987); P. Gaigg, M. Kreuzer, and G. Pollak, Phys. Rev. D **38**, 2559 (1988); G. Leibbrandt, Nucl. Phys. B **310**, 405 (1988).

A new unitary operator for a pair of coupled oscillators

Hong-yi Fan
Department of Physics, University of New Brunswick, P.O. Box 4400, Fredericton, New Brunswick E3B 5A3, Canada[a] *and Department of Modern Physics, China University of Science and Technology, Hefei, Anhui, People's Republic of China*

(Received 30 August 1988; accepted for publication 18 January 1989)

Using the IWOP (integration within ordered product) technique, a new unitary operator is found that is useful in diagonizing the Hamiltonian of a pair of coupled harmonic oscillators. The coordinate representation of the unitary operator is presented, and is well applied to obtain the wave function of energy eigenstate of the coupled oscillator.

I. INTRODUCTION

Of a system of two interacting spinless particles with coordinates Q_1 and Q_2, the typical one is the coupled harmonic oscillator whose Hamiltonian is given by

$$H=\frac{1}{2m}(P_1^2+P_2^2)+\frac{m}{2}\omega^2(Q_1^2+Q_2^2)+\frac{\kappa}{2}(Q_1-Q_2)^2$$

$$=\hbar\omega(a^\dagger a+b^\dagger b+1)+\frac{\kappa}{2}(Q_1-Q_2)^2, \tag{1}$$

where $a[a^\dagger]$, $b[b^\dagger]$ are two-mode annihilation [creation] operators of the uncoupled harmonic oscillator, and they are related to $Q_1[Q_2]$ and $P_1[P_2]$ by

$$Q_1=\sqrt{\frac{\hbar}{2m\omega}}(a+a^\dagger),\quad P_1=\sqrt{\frac{m\omega\hbar}{2}}\,\frac{a-a^\dagger}{i}, \tag{2}$$

$$Q_2=\sqrt{\frac{\hbar}{2m\omega}}(b+b^\dagger),\quad P_2=\sqrt{\frac{m\omega\hbar}{2}}\,\frac{b-b^\dagger}{i}. \tag{3}$$

The coupled oscillator problem has some intrinsic interest, because in the strong interaction limit, it is a crude model of the states of a tightly bound diatomic molecule in a crystal.[1] It is well known that by introducing the center-of-mass and relative coordinates respectively, as

$$X=\tfrac{1}{2}(Q_1+Q_2),\quad x=Q_1-Q_2, \tag{4}$$

and their conjugate momentum,

$$P=P_1+P_2,\quad [X,P]=i\hbar, \tag{5}$$

$$p=\tfrac{1}{2}(P_1-P_2),\quad [x,p]=i\hbar, \tag{6}$$

(1) can be put into

$$H=H_{\rm cm}+H_{\rm rel}, \tag{7}$$

$$H_{\rm cm}=\frac{P^2}{2M}+\frac{M}{2}\omega^2X^2,\quad M=2m;$$

$$H_{\rm rel}=\frac{p^2}{2\mu}+\frac{\mu}{2}\bar{\omega}^2x^2,\quad \mu=\frac{m}{2}, \tag{8}$$

where $\bar{\omega}$ is the new frequency, defined by

$$\bar{\omega}=(\omega^2+\kappa/\mu)^{1/2}. \tag{9}$$

Let the respective harmonic oscillator eigenstates of $H_{\rm rel}$ and $H_{\rm cm}$ be $|n_1\rangle'$ and $|n_2\rangle'$,

$$|n_1\rangle'|n_2\rangle'\equiv|n_1n_2\rangle'=(A^{\dagger n_1}B^{\dagger n_2}/\sqrt{n_1!n_2!})|00\rangle', \tag{10}$$

where $A^\dagger$, $B^\dagger$ are defined by

$$B^\dagger=\frac{1}{\sqrt{2}}\left(\sqrt{\frac{M\omega}{\hbar}}X-i\frac{P}{\sqrt{M\omega\hbar}}\right),\quad [B,B^\dagger]=1, \tag{11}$$

$$A^\dagger=\frac{1}{\sqrt{2}}\left(\sqrt{\frac{\mu\bar{\omega}}{\hbar}}x-i\frac{p}{\sqrt{\mu\bar{\omega}\hbar}}\right),\quad [A,A^\dagger]=1, \tag{12}$$

and $|00\rangle'$ is the ground state in the new Fock space spanned by $|n_1n_2\rangle'$. By virtue of Eqs. (2)–(6), A, B become

$$B=(1/\sqrt{2})(a+b),$$

$$A=(1/2\sqrt{2\omega\bar{\omega}})[(\bar{\omega}+\omega)(a-b)+(\bar{\omega}-\omega)(a^\dagger-b^\dagger)]. \tag{13}$$

Obviously, the Hamiltonian of the coupled oscillator is diagonized and its spectrum is given by

$$H|n_1n_2\rangle'=[\hbar\bar{\omega}(n_1+\tfrac{1}{2})+\hbar\omega(n_2+\tfrac{1}{2})]|n_1n_2\rangle'. \tag{14}$$

The wave function of $|n_1n_2\rangle'$, though a very simple function of the eigenvalues of X and x, is a rather complicated function of the eigenvalues q_1, q_2 of the particles coordinates Q_1 and Q_2.[1] In this paper, we find that the operator $U\exp[i(\pi/2)J_y]$, where $J_y=1/2i\,(a^\dagger b-b^\dagger a)$ and U is a new unitary operator whose coordinate representation is given in Sec. II, can play the role of transforming two isotropic dimensional harmonic oscillator's Fock basis,

$$|n_1n_2\rangle=(a^{\dagger n_1}b^{\dagger n_2}/\sqrt{n_1!n_2!})|00\rangle, \tag{15}$$

into the Fock basis $|n_1n_2\rangle'$. In terms of the transformation, the wave function $\langle q_1q_2|n_1n_2\rangle'$ can be easily obtained. The paper is arranged as follows.

In Sec. II, we begin by identifying the coordinate representation of U. Then the calculation is carried out, employing the newly developed IWOP technique[2–6] to obtain a normally ordered form of U. This form is further simplified by using an operator identity that is derived in Sec. III, following the way given by Ref. 7. The transforming properties of a, b under the U transformation are thus obtained, which manifestly show that $U\exp[i(\pi/2)J_y]$ is just the generating operator changing $|n_1n_2\rangle$ into $|n_1n_2\rangle'$. In Sec. IV, with the help of the coordinate representation of U, we give the explicit form of the wave function $\langle q_1q_2|n_1n_2\rangle'$.

II. COORDINATE REPRESENTATION OF U

The new unitary operator, which is useful in transforming (15) into (10), is introduced here by identifying the

[a] Present address.

 0022-2488/89/061273-04$02.50

following coordinate representation:

$$U=\left(\frac{\omega}{\overline{\omega}}\right)^{1/4}\int_{-\infty}^{\infty}dq_1\int_{-\infty}^{\infty}dq_2 \times\left|\frac{q_1+q_2}{2}+\frac{q_1-q_2}{2}\sqrt{\frac{\omega}{\overline{\omega}}},\ \frac{q_2+q_1}{2}+\frac{q_2-q_1}{2}\sqrt{\frac{\omega}{\overline{\omega}}}\right\rangle\left\langle q_1q_2\right|, \tag{16}$$

where $|q_1,q_2\rangle\equiv|q_1\rangle|q_2\rangle$ is the two-mode coordinate eigenstate. In the Fock space of two-dimensional harmonic oscillator, they are given by

$$|q_1\rangle=\left(\frac{m\omega}{\pi\hbar}\right)^{1/4}\exp\left\{-\frac{m\omega}{2\hbar}q_1^2+\sqrt{\frac{2m\omega}{\hbar}}q_1a^\dagger-\frac{a^{\dagger2}}{2}\right\}|0\rangle_1, \tag{17}$$

$$|q_2\rangle=\left(\frac{m\omega}{\pi\hbar}\right)^{1/4}\exp\left\{-\frac{m\omega}{2\hbar}q_2^2+\sqrt{\frac{2m\omega}{\hbar}}q_2b^\dagger-\frac{b^{\dagger2}}{2}\right\}|0\rangle_2, \tag{18}$$

where $|00\rangle\equiv|0\rangle_1|0\rangle_2$ is the ground state, satisfying

$$a|00\rangle=b|00\rangle=0, \tag{19}$$

$$|00\rangle\langle00|=:e^{-a^\dagger a-b^\dagger b}: \tag{20}$$

here : : stands for normal product.

The completeness relation of $|q_1\rangle$ can be reformed as an integration within normal product,

$$\int_{-\infty}^{\infty}dq_1|q_1\rangle\langle q_1| =\int_{-\infty}^{\infty}dq_1\left(\frac{m\omega}{\pi\hbar}\right)^{1/2}:\exp\left\{-\frac{m\omega}{\hbar}q_1^2+\sqrt{\frac{2m\omega}{\hbar}}q_1(a^\dagger+a)-\frac{1}{2}(a^\dagger+a)^2\right\}: =\int_{-\infty}^{\infty}\frac{dq}{\sqrt{\pi}}:\exp\left\{-\left[q-\frac{a^\dagger+a}{\sqrt{2}}\right]^2\right\}:=1. \tag{21}$$

By virtue of (17), (18), (20), and the IWOP technique, we can perform the integration in (16) similar in spirit to (21), and obtain

$$U=\pi^{-1}\left(\frac{\omega}{\overline{\omega}}\right)^{1/4}\int_{-\infty}^{\infty}dq_1\int_{-\infty}^{\infty}dq_2:\exp\left\{-\frac{1}{4}\left[\left(3+\frac{\omega}{\overline{\omega}}\right)(q_1^2+q_2^2)+2q_1q_2\left(1-\frac{\omega}{\overline{\omega}}\right)\right]+\frac{q_1}{\sqrt{2}}\left[\left(1+\sqrt{\frac{\omega}{\overline{\omega}}}\right)a^\dagger+\left(1-\sqrt{\frac{\omega}{\overline{\omega}}}\right)b^\dagger\right]+\frac{q_2}{\sqrt{2}}\left[\left(1-\sqrt{\frac{\omega}{\overline{\omega}}}\right)a^\dagger+\left(1+\sqrt{\frac{\omega}{\overline{\omega}}}\right)b^\dagger\right]+\sqrt{2}(q_1a+q_2b)-\frac{1}{2}(a^{\dagger2}+b^{\dagger2}+a^2+b^2)-a^\dagger a-b^\dagger b\right\}:$$
$$=\left(\frac{2\sqrt{\omega\overline{\omega}}}{\omega+\overline{\omega}}\right)^{1/2}\exp\left\{\frac{\omega-\overline{\omega}}{4(\omega+\overline{\omega})}(a^\dagger-b^\dagger)^2\right\}:\exp\left\{\frac{(\sqrt{\overline{\omega}}-\sqrt{\omega})^2}{2(\omega+\overline{\omega})}(a^\dagger b+b^\dagger a-a^\dagger a-b^\dagger b)\right\}:\exp\left\{\frac{\overline{\omega}-\omega}{4(\overline{\omega}+\omega)}(a-b)^2\right\}. \tag{22}$$

Here q_1, q_2 are dimensionless integration variables. Equation (22) is a normally ordered form and is not seen in the earlier literature. Obviously, when $\kappa=0, \overline{\omega}=\omega$, (22) reduces to 1, as expected. In order to prove that U is unitary, we rewrite (16) as

$$U=\left(\frac{\omega}{\overline{\omega}}\right)^{1/4}\int_{-\infty}^{\infty}dq_1\int_{-\infty}^{\infty}dq_2\left|g\begin{pmatrix}q_1\\q_2\end{pmatrix}\right\rangle\left\langle\begin{pmatrix}q_1\\q_2\end{pmatrix}\right|, \tag{23}$$

where

$$\left|\begin{pmatrix}q_1\\q_2\end{pmatrix}\right\rangle\equiv|q_1q_2\rangle,\quad g\equiv\frac{1}{2\sqrt{\overline{\omega}}}\begin{pmatrix}\sqrt{\overline{\omega}}+\sqrt{\omega} & \sqrt{\overline{\omega}}-\sqrt{\omega}\\ \sqrt{\overline{\omega}}-\sqrt{\omega} & \sqrt{\overline{\omega}}+\sqrt{\omega}\end{pmatrix}, \tag{24}$$

then we have

$$UU^\dagger=\sqrt{\frac{\omega}{\overline{\omega}}}\int_{-\infty}^{\infty}dq_1\int_{-\infty}^{\infty}dq_2\left|g\begin{pmatrix}q_1\\q_2\end{pmatrix}\right\rangle\left\langle g\begin{pmatrix}q_1\\q_2\end{pmatrix}\right| =1=U^\dagger U, \tag{25}$$

since the Jacobian for the integration variables' transformation is $\det g=\sqrt{\omega}/\sqrt{\overline{\omega}}$. To see how $a[b]$ changes under the transformation U, we need to further simplify the right-hand side of (22).

III. TRANSFORMATION PROPERTY OF a, b UNDER THE U TRANSFORMATION

In order to remove the : : which appears on the right-hand side of (22), we need a new operator identity, e.g.,

$$:\exp[\sigma(a^\dagger b+b^\dagger a-a^\dagger a-b^\dagger b)]: =\exp\left\{\left[-\frac{1}{2}\ln(1-2\sigma)\right](a^\dagger b+b^\dagger a-a^\dagger a-b^\dagger b)\right\}, \tag{26}$$

where σ is a parameter. To prove (26), we follow Ref. 7 by introducing

$$\Lambda=(a-b)/\sqrt{2},\quad \Lambda^\dagger=(a^\dagger-b^\dagger)/\sqrt{2}, \tag{27}$$

with $[\Lambda,\Lambda^\dagger]=1$. Consider canonical coherent state generated by Λ, i.e.,

$$\Lambda|f\rangle=f|f\rangle,\quad \langle f|f\rangle=1. \tag{28}$$

Then, obviously,

$$\langle f|e^{\lambda\Lambda^\dagger\Lambda}|f'\rangle=e^{-(1/2)(|f|^2+|f'|^2)e^{f^*f'e^\lambda}} =\langle f|:e^{(e^\lambda-1)\Lambda^\dagger\Lambda}:|f'\rangle. \tag{29}$$

By the overcompleteness relation of $|f\rangle$ we obtain

$$\exp[\lambda\Lambda^\dagger\Lambda] = :\exp[(e^\lambda - 1)\Lambda^\dagger\Lambda]: . \tag{30}$$

Setting $\lambda = \ln(1-2\sigma)$ in (30), we get (26). Note that (26) can also be proved by using the IWOP technique. In terms of (26), U becomes

$$U = \left(\frac{2\sqrt{\omega\overline{\omega}}}{\omega+\overline{\omega}}\right)^{1/2} \exp\left\{\frac{\omega-\overline{\omega}}{4(\omega+\overline{\omega})}(a^\dagger - b^\dagger)^2\right\} \times\exp\left(\frac{1}{2}(a^\dagger - b^\dagger)(a-b)\ln\frac{2\sqrt{\overline{\omega}\omega}}{\omega+\overline{\omega}}\right) \times\exp\left\{\frac{\overline{\omega}-\omega}{4(\omega+\overline{\omega})}(a-b)^2\right\} . \tag{31}$$

Using

$$e^{-iJy\theta}a^\dagger e^{iJy\theta} = a^\dagger\cos(\theta/2) + b^\dagger\sin(\theta/2) , \tag{32}$$

$$e^{-iJy\theta}b^\dagger e^{iJy\theta} = b^\dagger\cos(\theta/2) - a^\dagger\sin(\theta/2) , \tag{33}$$

we have

$$S \equiv \exp[\lambda(a^\dagger b + b^\dagger a - a^\dagger a - b^\dagger b)] = e^{i(\pi/2)Jy}e^{-2\lambda a^\dagger a}e^{-i(\pi/2)Jy} . \tag{34}$$

It then follows from (32), (33), and (34) that

$$\begin{aligned} SaS^{-1} &= \tfrac{1}{2}[a^\dagger(e^{-2\lambda}+1) + b^\dagger(1-e^{-2\lambda})] , \\ SaS^{-1} &= \tfrac{1}{2}[a(1+e^{2\lambda}) + b(1-e^{2\lambda})] , \end{aligned} \tag{35}$$

and

$$\begin{aligned} Sb^\dagger S^{-1} &= \tfrac{1}{2}[b^\dagger(e^{-2\lambda}+1) + a^\dagger(1-e^{-2\lambda})] , \\ SbS^{-1} &= \tfrac{1}{2}[b(1+e^{2\lambda}) + a(1-e^{2\lambda})] . \end{aligned} \tag{36}$$

By virtue of (35) and (36), as well as the formulas

$$e^{\lambda(a^\dagger - b^\dagger)^2}ae^{-\lambda(a^\dagger - b^\dagger)^2} = a + 2\lambda(b^\dagger - a^\dagger) , \tag{37}$$

$$e^{\lambda(a^\dagger - b^\dagger)^2}be^{-\lambda(a^\dagger - b^\dagger)^2} = b + 2\lambda(a^\dagger - b^\dagger) , \tag{38}$$

we know that under the U transformation a, b turns to

$$UaU^{-1} = (1/4\sqrt{\overline{\omega}\omega})[(a^\dagger - b^\dagger)(\overline{\omega}-\omega) + (\sqrt{\overline{\omega}}+\sqrt{\omega})^2 a - (\sqrt{\overline{\omega}}-\sqrt{\omega})^2 b] , \tag{39}$$

$$UbU^{-1} = (1/4\sqrt{\overline{\omega}\omega})[(b^\dagger - a^\dagger)(\overline{\omega}-\omega) + (\sqrt{\overline{\omega}}+\sqrt{\omega})^2 b - (\sqrt{\overline{\omega}}-\sqrt{\omega})^2 a] . \tag{40}$$

It then follows from (39), (40), and (13) that

$$\begin{aligned} U(1/\sqrt{2})(a+b)U^{-1} &= Ue^{i(\pi/2)Jy}be^{-i(\pi/2)Jy}U^{-1} \\ &= (1/\sqrt{2})(a+b) = B , \end{aligned} \tag{41}$$

$$\begin{aligned} U(1/\sqrt{2})(a-b)U^{-1} &= Ue^{i(\pi/2)Jy}ae^{-i(\pi/2)Jy}U^{-1} \\ &= (1/2\sqrt{2\omega\overline{\omega}})[(\overline{\omega}+\omega)(a-b) \\ &\quad + (\overline{\omega}-\omega)(a^\dagger - b^\dagger)] = A. \end{aligned} \tag{42}$$

Further, by virtue of (31), we have

$$Ue^{i(\pi/2)Jy}|00\rangle = \left(\frac{2\sqrt{\omega\overline{\omega}}}{\omega+\overline{\omega}}\right)^{1/2} \times\exp\left[\frac{\omega-\overline{\omega}}{4(\omega+\overline{\omega})}(a^\dagger - b^\dagger)^2\right]|00\rangle . \tag{43}$$

One can easily obtain

$$aUe^{i(\pi/2)Jy}|00\rangle = \frac{\omega-\overline{\omega}}{2(\omega+\overline{\omega})}(a^\dagger - b^\dagger)Ue^{i(\pi/2)Jy}|00\rangle , \tag{44}$$

$$bUe^{i(\pi/2)Jy}|00\rangle = \frac{\omega-\overline{\omega}}{2(\omega+\overline{\omega})}(b^\dagger - a^\dagger)Ue^{i(\pi/2)Jy}|00\rangle . \tag{45}$$

It then follows that

$$(1/\sqrt{2})(a+b)Ue^{i(\pi/2)Jy}|00\rangle = 0 , \tag{46}$$

$$(1/2\sqrt{2\omega\overline{\omega}})[(\overline{\omega}+\omega(a-b) + (\overline{\omega}-\omega)(a^\dagger - b^\dagger)]Ue^{i(\pi/2)Jy}|00\rangle = 0 . \tag{47}$$

By comparing (46) and (47) with (41) and (42), we can identify $|00\rangle'$ as

$$|00\rangle' = Ue^{i(\pi/2)Jy}|00\rangle . \tag{48}$$

Thus we obtain the result

$$Ue^{i(\pi/2)Jy}|n_1n_2\rangle = |n_1n_2\rangle' . \tag{49}$$

Note that (36) has squeeze operatorlike form.[4,8]

IV. CALCULATING WAVE FUNCTION $\langle q_1q_2|n_1n_2\rangle'$

In this section we show that the coordinate representation of U provides us with a convenient way to derive the wave function $\langle q_1q_2|n_1n_2\rangle'$. Using (49) and (16), we obtain

$$\begin{aligned} \langle q_1q_2|n_1n_2\rangle' &= \langle q_1q_2|Ue^{i(\pi/2)Jy}|n_1n_2\rangle \\ &= \left(\frac{\omega}{\overline{\omega}}\right)^{1/4}\left\langle q_1q_2\left|\int_{-\infty}^{\infty}dq_1'\int_{-\infty}^{\infty}dq_2'\right|\frac{q_1'+q_2'}{2}+\frac{q_1'-q_2'}{2}\sqrt{\frac{\omega}{\overline{\omega}}},\frac{q_2'+q_1'}{2} + \frac{q_2'-q_1'}{2}\sqrt{\frac{\omega}{\overline{\omega}}}\right\rangle\langle q_1'q_2'|e^{i(\pi/2)Jy}|n_1n_2\rangle \\ &= \left(\frac{\omega}{\overline{\omega}}\right)^{1/4}\int_{-\infty}^{\infty}dq_1'\int_{-\infty}^{\infty}dq_2'\,\delta\left(q_1-\frac{q_1'+q_2'}{2}-\frac{q_1'-q_2'}{2}\sqrt{\frac{\omega}{\overline{\omega}}}\right)\delta\left(q_2-\frac{q_2'+q_1'}{2}-\frac{q_2'-q_1'}{2}\sqrt{\frac{\omega}{\overline{\omega}}}\right) \times\langle q_1'q_2'|e^{i(\pi/2)Jy}|n_1n_2\rangle \\ &= \left(\frac{\overline{\omega}}{\omega}\right)^{1/4}\left\langle\left(\sqrt{\frac{\overline{\omega}}{\omega}}+1\right)\frac{q_1}{2}-\left(\sqrt{\frac{\overline{\omega}}{\omega}}-1\right)\frac{q_2}{2},\left(\sqrt{\frac{\overline{\omega}}{\omega}}+1\right)\frac{q_2}{2}-\left(\sqrt{\frac{\overline{\omega}}{\omega}}-1\right)\frac{q_1}{2}\right|e^{i(\pi/2)Jy}\left|n_1n_2\right\rangle , \end{aligned} \tag{50}$$

in which

$$e^{i(\pi/2)Jy}|n_1n_2\rangle = \left(\frac{1}{\sqrt{2}}\right)^{n_1+n_2}\frac{(a^\dagger - b^\dagger)^{n_1}(a^\dagger + b^\dagger)^{n_2}}{\sqrt{n_1!n_2!}}\Big|00\Big\rangle$$

$$= \left(\frac{1}{\sqrt{2}}\right)^{n_1+n_2}\sum_{l=0}^{n_1}\sum_{k=0}^{n_2}\binom{n_1}{l}\binom{n_2}{k}(-)^{n_1-l}\frac{[(l+k)!(n_1+n_2-l-k)!]^{1/2}}{\sqrt{n_1!n_2!}}\Big|l+k, n_1+n_2-l-k\Big\rangle. \quad (51)$$

Substituting (51) into (50), and using

$$\left\langle\left(\sqrt{\frac{\overline{\omega}}{\omega}}+1\right)\frac{q_1}{2}-\left(\sqrt{\frac{\overline{\omega}}{\omega}}-1\right)\frac{q_2}{2}, \left(\sqrt{\frac{\overline{\omega}}{\omega}}+1\right)\frac{q_2}{2}-\left(\sqrt{\frac{\overline{\omega}}{\omega}}-1\right)\frac{q_1}{2}\Big| l+k, n_1+n_2-l-k\right\rangle$$

$$= \left(\frac{m\omega}{\pi\hbar}\right)^{1/2}[(l+k)!(n_1+n_2-l-k)!2^{n_1+n_2}]^{-1/2}\exp\left[-\frac{m}{4\hbar}[(q_1^2+q_2^2)(\omega+\overline{\omega})-2q_1q_2(\overline{\omega}-\omega)]\right]$$

$$\times H_{l+k}\left\{\sqrt{\frac{m}{\hbar}}\left[(\sqrt{\overline{\omega}}+\sqrt{\omega})\frac{q_1}{2}-(\sqrt{\overline{\omega}}-\sqrt{\omega})\frac{q_2}{2}\right]\right\}H_{n_1+n_2-l-k}\left\{\sqrt{\frac{m}{\hbar}}\left[(\sqrt{\overline{\omega}}+\sqrt{\omega})\frac{q_2}{2}-(\sqrt{\overline{\omega}}-\sqrt{\omega})\frac{q_1}{2}\right]\right\}. \quad (52)$$

Where H_{l+k} is the Hermite polynomial, we get

$$\langle q_1q_2|n_1n_2\rangle' = \left(\frac{1}{2}\right)^{n_1+n_2}\left(\frac{m^2\omega\overline{\omega}}{\pi^2\hbar^2}\right)^{1/4}\sum_{l=0}^{n_1}\sum_{k=0}^{n_2}\binom{n_1}{l}\binom{n_2}{k}(-)^{n_1-l}(n_1!n_2!)^{-1/2}$$

$$\times\exp\left\{-\frac{m}{4\hbar}[(q_1^2+q_2^2)(\omega+\overline{\omega})-2q_1q_2(\overline{\omega}-\omega)]\right\}H_{l+k}\left\{\sqrt{\frac{m}{\hbar}}\left[(\sqrt{\overline{\omega}}+\sqrt{\omega})\frac{q_1}{2}-(\sqrt{\overline{\omega}}-\sqrt{\omega})\frac{q_2}{2}\right]\right\}$$

$$\times H_{n_1+n_2-l-k}\left\{\sqrt{\frac{m}{\hbar}}\left[(\sqrt{\overline{\omega}}+\sqrt{\omega})\frac{q_2}{2}-(\sqrt{\overline{\omega}}-\sqrt{\omega})\frac{q_1}{2}\right]\right\}. \quad (53)$$

Finally, it is worthwhile to point out that although the separate form (7) of the coupled oscillator has been known for a long time, the problem of what is the unitary transformation changing $|n_1n_2\rangle$ into $|n_1n_2\rangle'$ had not been paid enough attention before this work solved the problem.

ACKNOWLEDGMENT

The author wishes to thank the referee for bringing Ref. 7 to his attention.

[1]D. S. Saxon, *Elementary Quantum Mechanics* (Holden–Day, San Francisco, 1968).

[2]Hong-yi Fan and Tu-nan Ruan, Sci. Sin. A **27**, 392 (1984); **28**, 1067 (1985).

[3]Hong-yi Fan and Tu-nan Ruan, Commun. Theor. Phys. (Beijing) **4**, 181 (1985); **3**, 345 (1984); **2**, 1563 (1983).

[4]Hong-yi Fan, H. R. Zaidi, and J. R. Klauder, Phys. Rev. D **35**, 1831 (1987).

[5]Hong-yi Fan and H. R. Zaidi, Phys. Lett. A **124**, 303 (1987).

[6]Hong-yi Fan, Phys. Lett. A **126**, 145, 150 (1987); **127**, 403 (1988); **129**, 4 (1988).

[7]J. R. Klauder and B.-S. Skagerstam, *Coherent States* (World Scientific, Singapore, 1985), p. 678; see also B.-S. Skagerstam, Z. Phys. C **24**, 97 (1984).

[8]Hong-yi Fan and J. R. Klauder, J. Phys. A **21**, L725 (1988).

Long-time behavior of two-point functions of a quantum harmonic oscillator interacting with bosons

Asao Arai
Department of Mathematics, Hokkaido University, Sapporo 060, Japan

(Received 18 August 1988; accepted for publication 8 February 1989)

A class of exactly soluble models of a one-dimensional quantum harmonic oscillator interacting with bosons moving in the d-dimensional space $\mathbf{R}^d$ is considered and the long-time behavior of the two-point function of the oscillator at zero temperature and at finite temperatures is analyzed. It is shown that the two-point functions decay with a power-law respectively as the time tends to infinity and that, in the case where the boson is massless, the two-point function at zero temperature decays faster than those at finite temperatures, while, in the case where the boson is massive, they decay with the same order. Further, the dependence of the decay order on d as well as on the infrared behavior of the one-boson energy and the momentum cutoff function in the interaction Hamiltonian is clarified in each case.

I. INTRODUCTION AND THE MAIN RESULTS

In a previous paper,[1] we considered a general class of models of a one-dimensional quantum harmonic oscillator coupled to infinitely many bosons in the Hilbert space

$$\mathscr{F} = L^2(\mathbf{R}) \otimes \mathscr{F}_s(\mathscr{H}), \tag{1.1}$$

where $\mathscr{F}_s(\mathscr{H})$ is the symmetric (Boson) Fock space over a complex Hilbert space $\mathscr{H}$ [e.g., Ref. 2 (§II.4)]. The unperturbed (free) Hamiltonian of each model in the class has the common form

$$H_0 = \omega_0 a^* a \otimes I + I \otimes d\Gamma(h). \tag{1.2}$$

Here a is the annihilation operator of the oscillator acting in $L^2(\mathbf{R})$, $d\Gamma(h)$ is the second quantization of a non-negative self-adjoint operator h (the one-particle free Hamiltonian of the boson) [e.g., Ref. 3 (§X.7)], I denotes identity, and $\omega_0 > 0$ is a constant parameter denoting the circular frequency of the oscillator. The interaction part (perturbation) H_I is given by a general form quadratic in the annihilation and creation operators for the bosons and the oscillator. (For the detailed form of H_I, see Ref. 1. But, in the present paper, we do not need it.) It was proved in Ref. 1 that, under some conditions, the total Hamiltonian

$$H = H_0 + H_I \tag{1.3}$$

is unitarily equivalent to $d\Gamma(h) + E_0$ [acting in $\mathscr{F}_s(\mathscr{H})$] with a real constant E_0 and hence, in particular, all the possible embedded eigenvalues of H_0, except for the zero eigenvalue, disappear under the perturbation. Further, in the case $\mathscr{H} = L^2(\mathbf{R}^d)$, taking H_0 as

$$H_0 = \omega_0 a^* a \otimes I + I \otimes d\Gamma(\omega^r) \tag{1.4}$$

with a rotation invariant function ω on $\mathbf{R}^d$ and a parameter $r > 0$ and changing the parameters in H_I, we showed that the class gives a unified description of standard models of a one-dimensional quantum harmonic oscillator coupled to bosons, containing the following ones:

(M1) (The RWA oscillator; $r = 1$) (e.g., Refs. 4–7)

$$H = \omega_0 a^* a \otimes I + I \otimes d\Gamma(\omega) + a \otimes b(\rho)^* + a^* \otimes b(\rho).$$

(M2) (The Schwabl–Thirring model; $r = 1/2$) (e.g., Refs. 6, 8–13)

$$H = \omega_0 a^* a \otimes I + I \otimes d\Gamma(\omega^{1/2}) + q \otimes \phi(\rho).$$

(M3) ($r = 1/2$)

$$H = \tfrac{1}{2} p^2 \otimes I + I \otimes d\Gamma(\omega^{1/2}) + (\omega_0^2/2)(q \otimes I - I \otimes \phi(\rho))^2 - \tfrac{1}{2}\omega_0.$$

Here $b(f), f \in L^2(\mathbf{R}^d)$, is the (smeared) boson annihilation operator acting in $\mathscr{F}_s(L^2(\mathbf{R}^d))$, $q \in \mathbf{R}$ denotes the position of the oscillator, $p = -id/dq$,

$$\phi(f) = (1/\sqrt{2})\{b(\omega^{-1/4} f)^* + b(\omega^{-1/4} f)\},$$

and ρ is a cutoff function.

The model (M3) may be regarded as a simplified version of a three-dimensional quantum harmonic oscillator minimally coupled to a quantized radiation field with an ultraviolet cutoff (e.g., Ref. 14 and references therein).

In the present paper, taking $\mathscr{H} = L^2(\mathbf{R}^d)$ with H_0 given by (1.4), we investigate the long-time behavior of the two-point function at the zero temperature defined by

$$W(t_1, t_2) = (\Omega, q(t_1) q(t_2) \Omega), \quad t_1, t_2 \in \mathbf{R}, \tag{1.5}$$

and that at a finite temperature $\beta^{-1} > 0$ given symbolically by

$$W_\beta(t_1, t_2) = \frac{\mathrm{Tr}(e^{-\beta H} q(t_1) q(t_2))}{\mathrm{Tr}\, e^{-\beta H}}, \tag{1.6}$$

where $q(t)$ is the time evolution of q by the total Hamiltonian H given by (1.3)

$$q(t) = e^{itH} q e^{-itH}, \quad q \in \mathbf{R}, \tag{1.7}$$

Ω is the ground state of H, and Tr denotes the trace [strictly speaking, the rhs of (1.6) is defined as an infinite volume limit of a finite volume approximation of the quantum system under consideration]. As we shall see below, $W(t_1, t_2)$ [resp. $W_\beta(t_1, t_2)$] has a characteristic form common to all the models in the class. We are mainly interested in the following aspects:

(I) The dependence of the order of the time decay on the dimension d, r, and on the infrared property of the functions of ω and ρ [i.e., the behavior of $\omega(\mathbf{k})$ and $\rho(\mathbf{k})$ as $\mathbf{k} \to 0$].

(II) The difference of the time decay between the case of finite temperatures ($0 < \beta < \infty$) and that of the zero temperature ($\beta = \infty$).

(III) The difference of the time decay between the

 0022-2488/89/061277-12$02.50

massless case [$\inf_{\mathbf{k}} \omega(\mathbf{k}) = 0$] and the massive case [$\inf_{\mathbf{k}} \omega(\mathbf{k}) > 0$].

We remark that we can not expect an exponential decay for the two-point functions, because we assume that H is bounded from below to define them; this follows from a general theorem (Ref. 15, §7.3, Theorem 3.3). See also Appendix C in the present paper.

The problem of the long-time behavior of two-point functions (or correlation functions) of a quantum harmonic oscillator coupled to a heat bath (an infinite system of bosons) has recently been discussed in the statistical physics literature (e.g., Refs. 6, 12, 16, and 17; cf. also Ref. 10 for a field theoretical discussion) and some partial results have been obtained with some concrete models [mainly those related to the models (M1) and (M2)]. In these works, however, the authors make an *ad hoc* assumption on the spectral density of the correlation function or on the memory kernel in the Langevin equation. This procedure formally corresponds to taking a special form for the cutoff function ρ in our Hamiltonian formalism. In the present paper, we do not make such an *ad hoc* assumption and consider the models as generally as possible. Accordingly, our results include as special cases the partial results on the long-time behavior mentioned above and generalize on them. At the end of this section, after stating our results, we shall give a more detailed comparison of our results with those obtained in the other works.

We now proceed to describe the two-point functions in our models.

Let ω_1 be a non-negative, continuously differentiable, and monotone increasing function on $(0,\infty)$ such that $\omega_1(x) \to \infty$ as $x \to \infty$ with the derivative $\omega_1'(x) > 0$ for all $x > 0$. We set

$$m = \inf_{x>0} \omega_1(x) \geqslant 0. \tag{1.8}$$

The rotation invariant function ω on $\mathbf{R}^d$ is defined by

$$\omega(\mathbf{k}) = \omega_1(|\mathbf{k}|), \quad \mathbf{k} \in \mathbf{R}^d. \tag{1.9}$$

The cutoff function ρ is a real-valued continuous function in $L^2(\mathbf{R}^d)$ such that

$$\int_{\mathbf{R}^d} \frac{\rho(\mathbf{k})^2}{\omega(\mathbf{k})} d\mathbf{k} < \infty, \tag{1.10}$$

and

$$\int_{\mathbf{R}^d} \frac{\rho(\mathbf{k})^2}{|m - \omega(\mathbf{k})|} d\mathbf{k} < \infty. \tag{1.11}$$

Remark: If $m > 0$ (the massive case), then (1.10) is automatically satisfied because of $\rho \in L^2(\mathbf{R}^d)$ [more generally, we have $\omega^{-\lambda}\rho \in L^2(\mathbf{R}^d)$ for all $\lambda > 0$], but, in the case $m = 0$ (the massless case), this is not true.

For $j = 0,1$, we introduce the function $\Phi_\rho^{(j)}(z)$ by

$$\Phi_\rho^{(j)}(z) = \int_{\mathbf{R}^d} \frac{\omega(\mathbf{k})^j \rho(\mathbf{k})^2}{z - \omega(\mathbf{k})} d\mathbf{k}, \tag{1.12}$$

which is analytic in the cut plane

$$\mathbf{C}_m = \mathbf{C} \setminus [m, \infty). \tag{1.13}$$

Let

$$D^{(j)}(z) = a_0^{(j)} - a_1^{(j)} z + \delta_{1,j} \int \rho(\mathbf{k})^2 d\mathbf{k} + \Phi_\rho^{(j)}(z), \tag{1.14}$$

where $a_i^{(j)}$, $i,j = 0,1$, are real constants with $a_1^{(j)} > 0$, $j = 0,1$, and $\delta_{1,j}$ is the Kronecker delta function. We assume that

$$d_m^{(j)} \equiv a_0^{(j)} - a_1^{(j)} m + \delta_{1,j} \int \rho(\mathbf{k})^2 d\mathbf{k} + \Phi_\rho^{(j)}(m) > 0. \tag{1.15}$$

Then it is easy to see that

$$D^{(j)}(z) \neq 0, \quad z \in \mathbf{C}_m. \tag{1.16}$$

Remark: By renormalizing the parameter $a_0^{(j)}$ as

$$\tilde{a}_0^{(j)} = a_0^{(j)} + a_1^{(j)} m - \delta_{1,j} \int \rho(\mathbf{k})^2 d\mathbf{k} - \Phi_\rho^{(j)}(m), \quad a_0^{(j)} > 0,$$

$d_m^{(j)}$ can be made positive.

For technical reasons, we assume the following:

(AI) (a) $\sup_{\substack{\epsilon>0 \\ x\in[m,\infty)}} |\Phi_\rho^{(j)}(x - i\epsilon)| < \infty$,

(b) $\inf_{\substack{\epsilon>0 \\ x\in[m,\infty)}} |D^{(j)}(x - i\epsilon)| > 0$.

For sufficient conditions for (AI) to hold, see Appendices A and B.

It follows from (AI)(a) that, for all sufficiently large $x > 0$,

$$|D^{(j)}(x - i\epsilon)| \geqslant cx \tag{1.17}$$

with a constant $c > 0$ independent of ϵ.

One can show using the theory of the Hilbert transform (e.g., Ref. 18) that the limits

$$D_\pm^{(j)}(x) \equiv \lim_{\epsilon \downarrow 0} D^{(j)}(x \pm i\epsilon), \quad j = 0, 1, \tag{1.18}$$

exist for a.e. $x \in (m,\infty)$, which, by assumption (AI)(b), cannot be zero.

We are now ready to give an explicit form of $W(t_1,t_2)$ [resp. $W_\beta(t_1,t_2)$] (up to a constant multiple) which follows from Ref. 1:

$$W(t_1,t_2) = W(t_1 - t_2), \tag{1.19}$$

$$W_\beta(t_1,t_2) = W_\beta(t_1 - t_2), \tag{1.20}$$

with

$$W(t) = \int_{\mathbf{R}^d} \frac{\rho(\mathbf{k})^2 e^{-it\omega(\mathbf{k})^r}}{\omega(\mathbf{k})^{2r\alpha} |D_+^{(j)}(\omega(\mathbf{k}))|^2} d\mathbf{k}, \quad j = 0, 1, \tag{1.21}$$

and

$$W_\beta(t) = \int_{\mathbf{R}^d} \frac{\rho(\mathbf{k})^2 (e^{(\beta - it)\omega(\mathbf{k})^r} + e^{it\omega(\mathbf{k})^r})}{\omega(\mathbf{k})^{2r\alpha} |D_+^{(j)}(\omega(\mathbf{k}))|^2 (e^{\beta\omega(\mathbf{k})^r} - 1)} d\mathbf{k}, \quad j = 0, 1. \tag{1.22}$$

Here $r > 0$ is the parameter in H_0 given by (1.4) and $\alpha \in \mathbf{R}$ is a parameter appearing in $b(\omega^{-r\alpha}\rho)$ and $b(\omega^{-r\alpha}\rho)^*$ in the interaction H_I, and j depends on the form of H_I; for example, in the model (M1) [resp. (M2), (M3)], we have $(r, \alpha, j) = (1, 0, 0)$ [resp. $(\frac{1}{2},\frac{1}{2},0)$, $(\frac{1}{2},\frac{1}{2},1)$]. Of course, the con-

stants $a_0^{(j)}$ and $a_1^{(j)}$ also change according to models : In the model (M1) [resp. (M2), (M3)], we have $a_0^{(0)} = \omega_0$ and $a_1^{(0)} = 1$ [resp. $a_0^{(0)} = \omega_0^2 a_1^{(0)}$, $a_0^{(1)} = \omega_0^2 a_1^{(1)}$]. In deriving (1.21) [resp. (1.22)] with the case $m = 0$, an additional assumption that $\rho^2/(\omega^{2r\alpha}|D_+^{(j)}(\omega)|^2)$ [resp. $\rho^2/\omega^{2r\alpha+1}|D_+^{(j)}(\omega)|^2$] is integrable near the origin is made. Note that, for all $t\in\mathbf{R}$, $W_\beta(t) \to W(t)$ as $\beta \to \infty$ (the zero-temperature limit).

To make the present paper self-contained, we take formulas (1.19)–(1.22) as the starting point. Therefore the reader is not required to have any detailed knowledge in Ref. 1.

We shall denote by $\eta(x)$ the inverse function of $\omega_1(x)$,

$$\omega_1(\eta(x)) = x \tag{1.23}$$

for $x > m$. By the inverse function theorem, $\eta(x)$ is differentiable and monotone increasing in (m,∞) with

$$\lim_{x \downarrow m} \eta(x) = 0, \quad \eta'(x) = [\omega_1'(\eta(x))]^{-1}, \quad x\in(m,\infty). \tag{1.24}$$

For a measurable function f on $\mathbf{R}^d$, we define the function $[f]$ on $[0,\infty)$ by

$$[f](x) = \int_{S^{d-1}} dS(\theta) f(x\theta), \quad x\in[0,\infty), \tag{1.25}$$

where S^{d-1} is the $d-1$ sphere and dS is the surface integral measure on S^{d-1}. Let

$$I^{(j)}(x) = \eta'(x)\eta(x)^{d-1}x^j[\rho^2](\eta(x)), \quad x\in(m,\infty). \tag{1.26}$$

Since ρ is continuous by assumption, it follows that $I^{(j)}(x)$ is continuous in $x\in(m,\infty)$.

As another technical assumption, we take the following:

(AII) There exists a constant $\theta\in(0,\pi/2)$ such that the function $I^{(j)}(x)$ has an analytic continuation $I^{(j)}(z)$ onto the domain

$$\mathbf{D}_{m,\theta} = \{z\in\mathbf{C} \mid \mathrm{Re}\, z > m, \ -\theta < \arg z < 0\} \tag{1.27}$$

with the following properties:

(a) $\lim_{\epsilon \downarrow 0} I^{(j)}(x - i\epsilon) = I(x)$, $x\in(m,\infty)$.

(b) $|I^{(j)}(z)| \leqslant \text{const}\ |z|^{-q_j}$

for all sufficiently large $|z|$ ($z\in\mathbf{D}_{m,\theta}$) with a constant $q_j \geqslant 0$.

(c) $$\lim_{\substack{z\to 0 \\ z\in\mathbf{D}_{0,\theta}}} \frac{I^{(j)}(m+z)}{z^{p_j(m)}} = A_m^{(j)}$$

with constants $A_m^{(j)} \neq 0$ and $p_j(m) \geqslant 0$.

(d) For all sufficiently small $\epsilon_0 > 0$,

$$\inf_{\substack{0<\epsilon<\epsilon_0 \\ x\in[m,\infty)}} |D^{(j)}(x - i\epsilon) - 2i\pi I^{(j)}(x - i\epsilon)| > 0.$$

Remark: As is seen from the definition of $I^{(j)}(x)$, $p_j(m)$ is determined by the dimension d and the infrared property of ω_1, ω_1', and ρ. See Sec. IV for an example.

In order to investigate the decay property of $W(t)$ and $W_\beta(t)$ in a unified way, we introduce the following more general function:

$$W(t,f) = \int_{\mathbf{R}^d} \frac{f(\omega(\mathbf{k}))\rho(\mathbf{k})^2 e^{-it\omega(\mathbf{k})^r}}{\omega(\mathbf{k})^{2r\alpha}|D_+^{(j)}(\omega(\mathbf{k}))|^2}\, d\mathbf{k}. \tag{1.28}$$

Here f is any meromorphic function in $\mathbf{D}_{m,\theta}$ with the following properties:

(f I) The number of poles of $f(z)$ is finite.

(f II) The limit

$$\lim_{\epsilon \downarrow 0} f(x - i\epsilon) \equiv f(x)$$

exists for a.e. $x\in(m,\infty)$.

(f III) $\sup\{|f(z)| \mid z\in\mathbf{D}_{m,\theta}\cup(m,\infty), |z|\geqslant r_0\} < \infty$ with a constant $r_0 > 0$.

(f IV) $$\lim_{\substack{z\to 0 \\ z\in\mathbf{D}_{0,\theta}}} z^{\alpha_m(f)} f(m+z) = f_m$$

with constants $\alpha_m(f)\in\mathbf{R}$ and $f_m \neq 0$.

We now state the main results. We mean by

$$g(t) \underset{t\to\infty}{\sim} h(t)$$

that

$$\lim_{t\to\infty} \frac{g(t)}{h(t)} = 1.$$

Let

$$B_m^{(j)} = (D^{(j)}(m) - 2i\pi\delta_{0,p_j(m)}A_m^{(j)})D^{(j)}(m). \tag{1.29}$$

Theorem 1.1: Suppose that $2r\alpha + j + q_j + 1 > 0$.

(a) Let $m = 0$ and suppose that

$$\mu_j \equiv \frac{p_j(0) - \alpha_0(f) - 2r\alpha - j + 1}{r} > 0. \tag{1.30}$$

Then

$$W(t,f) \underset{t\to\infty}{\sim} \frac{A_0^{(j)} f_0 e^{-i\pi\mu_j/2}\Gamma(\mu_j)}{rB_0^{(j)}}\, t^{-\mu_j}, \tag{1.31}$$

where $\Gamma(z)$ is the gamma function.

(b) Let $m > 0$ and

$$\nu_j = p_j(m) - \alpha_m(f) + 1 > 0. \tag{1.32}$$

Then

$$W(t,f) \underset{t\to\infty}{\sim} \frac{A_m^{(j)} f_m e^{-i\pi\nu_j/2}\Gamma(\nu_j)e^{-itm^r}}{r^{\nu_j} m^{2r\alpha+j+(r-1)\nu_j} B_m^{(j)}}\, t^{-\nu_j}. \tag{1.33}$$

Remark: Under the assumption of Theorem 1.1 in each case $m = 0$ or $m > 0$, the integral of the rhs of (1.28) is absolutely convergent.

As corollaries of Theorem 1.1, we can derive the asymptotic behavior of $W(t)$ and $W_\beta(t)$ as $t \to \infty$.

Theorem 1.2 (the massless case): Let $m = 0$ and $2r\alpha + j + q_j + 1 > 0$.

(a) Suppose that

$$\lambda_j \equiv \frac{p_j(0) - 2r\alpha - j + 1}{r} > 0. \tag{1.34}$$

Then

$$W(t) \underset{t\to\infty}{\sim} \frac{A_0^{(j)} e^{-i\pi\lambda_j/2}\Gamma(\lambda_j)}{rB_0^{(j)}}\, t^{-\lambda_j}. \tag{1.35}$$

(b) Suppose that $\lambda_j > 1$. Then

$$W_\beta(t) \underset{t\to\infty}{\sim} 2\,\mathrm{Re}\left\{\frac{A_0^{(j)} e^{-i\pi(\lambda_j-1)/2}\Gamma(\lambda_j-1)}{\beta r B_0^{(j)}}\right\} t^{-\lambda_j+1}. \tag{1.36}$$

Proof: (a) We have

$$W(t) = W(t, f\equiv 1).$$

On the other hand, the function $f(z)\equiv 1$ obviously possesses properties (f I) and (f II) with $\alpha_0(f)=0$ and $f_0=1$. Therefore, in this case, we have $\mu_j=\lambda_j$. Thus (1.31) yields (1.35).

(b) Let

$$f(z) = e^{\beta z^r}/(e^{\beta z^r}-1)$$

and

$$g(z) = 1/(e^{\beta \bar{z}^r}-1).$$

Then we have

$$W_\beta(t) = W(t,f) + \overline{W(t,\bar{g})}.$$

It is easy to see that each of f and $\bar{g}$ satisfies (f I) and (f II) with $\alpha_0(f)=\alpha_0(\bar{g})=r$ and $f_0=\bar{g}_0=1/\beta$. Hence we have $\mu_j=\lambda_j-1$. Thus we get (1.36) from (1.31). □

Theorem 1.3 (the massive case): Let $m>0$ and $2r\alpha+j+q_j+1>0$. Then

$$\text{(a)}\quad W(t) \underset{t\to\infty}{\sim} \frac{A_m^{(j)} e^{-i\pi(p_j(m)+1)/2}\Gamma(p_j(m)+1)}{r^{p_j(m)+1} m^{2r\alpha+j+(r-1)(p_j(m)+1)} B_m^{(j)}} \times e^{-itm^r} t^{-p_j(m)-1}. \tag{1.37}$$

$$\text{(b)}\quad W_\beta(t) \underset{t\to\infty}{\sim} \frac{A_m^{(j)}\Gamma(p_j(m)+1)}{(e^{\beta m^r}-1) r^{p_j(m)+1} m^{2r\alpha+j+(r-1)(p_j(m)+1)}} \times\left\{\frac{e^{(\beta-it)m^r - i\pi(p_j(m)+1)/2}}{B_m^{(j)}} + \frac{e^{itm^r + i\pi(p_j(m)+1)/2}}{\overline{B_m^{(j)}}}\right\} t^{-p_j(m)-1}. \tag{1.38}$$

Proof: Similar to the proof of theorem 1.2. □

Theorems 1.2 and 1.3 show that the two-point functions decay with a power-law respectively as $t\to\infty$. In both cases $m=0$ and $m>0$, the order of the decay increases as $p_j(m)$ does. On the other hand, the dependence of the decay order on the dimension d and the infrared behavior of ω and ρ comes only from $p_j(m)$ [see (1.26) and (AII)(c)]. In particular, it follows from (1.26) that $p_j(m)$ is monotone increasing in d. Therefore the higher d becomes, the faster the two-point functions decay.

In the massless case, Theorem 1.2 shows that $W(t)$ decays faster than $W_\beta(t)$. On the other hand, in the massive case, the order of the decay of $W(t)$ coincides with that of $W_\beta(t)$. We note also that, in the massive case, the oscillating factors $\exp(\pm itm^r)$ appear in the asymptotic behavior of the two-point functions.

Remark: We have assumed $d_m^{(j)}>0$ [(1.15)] to obtain Theorems 1.1–1.3. In fact, in the case $d_m^{(j)}<0, D^{(j)}(z)$ has a simple unique zero $\nu_0\in(-\infty, m)$ and the oscillating terms proportional to $\exp(\pm i\nu_0 t)$ appear in the two-point functions. This corresponds to the nondisappearence of embedded eigenvalues of H_0 under the perturbation (in other words, the oscillator mode persists under the perturbation) (cf. Ref. 7).

We now try to compare in some detail our results with those on the long-time behavior obtained in the other works.[6,12,16,17]

In Ref. 16, a phenomenological model based on an averaged Langevin equation rather than on a Hamiltonian was presented; the correlation function at the inverse temperature β is given as

$$\langle q(t)q(0)\rangle = \int_{-\infty}^{\infty} \frac{d\omega}{\pi}\chi''(\omega)\frac{\hbar e^{-i\omega t}}{1-e^{-\beta\hbar\omega}}, \tag{1.39}$$

where $\chi''(\omega)$ is the imaginary part of the Fourier transform of the response function,

$$\chi''(\omega) = \frac{1}{M}\cdot\frac{\gamma\omega}{(\omega^2-\omega_0^2)^2+\gamma^2\omega^2} \tag{1.40}$$

with positive constants γ (the damping constant) and M (the mass of the oscillator) and $\hbar$ is the Planck constant divided by 2π. They showed that, at the zero temperature $\beta\to\infty$, the real part $S(t)$ (the symmetrized correlation function) decays as

$$S(t) \underset{t\to\infty}{\sim} -\frac{\hbar\gamma}{\pi M\omega_0^4}\frac{1}{t^2}. \tag{1.41}$$

We can show that this result is a special case of our results: in fact, at the zero temperature ($\beta\to\infty$), one has from (1.39)

$$\langle q(t)q(0)\rangle = \int_0^{\infty}\frac{d\omega}{\pi}\chi''(\omega)\hbar e^{-i\omega t}. \tag{1.42}$$

On the other hand, the function $W(t)$ given by (1.21) is written as

$$W(t) = \int_m^{\infty} w(x)e^{-itx^r}dx \tag{1.43}$$

with

$$w(x) = \frac{I^{(j)}(x)}{x^{2rd\alpha+j}|D_x^{(j)}(x)|^2}. \tag{1.44}$$

Therefore, by considering the case with $m=0$ and $r=1$ and by setting formally

$$w(x) = (\hbar/\pi)\chi''(x),$$

we have at the zero temperature

$$W(t) = \langle q(t)q(0)\rangle.$$

Since $\chi''(x)\sim\gamma x/(M\omega_0^4)$ as $x\to 0$, we have

$$A_0^{(j)}/B_0^{(j)} = \hbar\gamma/\pi M\omega_0^4, \quad \lambda_j = 2.$$

Then, (1.41) follows from (1.35). On the other hand, at finite temperatures $\beta^{-1}>0$, $\langle q(t)q(0)\rangle$ is not of the form $W_\beta(t)$ given by (1.22), because the frequency spectrum of $\langle q(t)q(0)\rangle$ is equal to the whole real line $\mathbf{R}$ as is seen from (1.39). [The frequency (energy) spectrum of our models is equal to $[m,\infty)$.] This is a big difference between the model under consideration and ours. They also showed that, at finite temperatures, $\langle q(t)q(0)\rangle$ decays exponentially. We remark that, if one modifies the correlation function as

$$\langle q(t)q(0)\rangle_+ = \int_0^{\infty}\frac{d\omega}{\pi}\chi''(\omega)\frac{\hbar e^{-i\omega t}}{1-e^{-\beta\hbar\omega}}$$

or

$$\langle q(t)q(0)\rangle_- = \int_{-\infty}^{0} \frac{d\omega}{\pi}\chi''(\omega)\frac{\hbar e^{-i\omega t}}{1-e^{-\beta\hbar\omega}},$$

then one gets

$$\langle q(t)q(0)\rangle_+ \underset{t\to\infty}{\sim} \frac{(-i)\gamma}{\pi M\omega_0^4}\cdot\frac{1}{t},$$

$$\langle q(t)q(0)\rangle_- \underset{t\to\infty}{\sim} \frac{i\gamma}{\pi M\omega_0^4\beta}\cdot\frac{1}{t}.$$

Namely, $\langle q(t)q(0)\rangle_\pm$ have a power-law decay. This follows from an application of Theorem 2.1(a) in Sec. II.

In Ref. 17, a generalized version of the model in Ref. 16 was considered and a more detailed analysis on the long-time behavior of the correlation functions was done. As in the model in Ref. 16, the correlation function at zero temperature is a special case of our model and the result on the long-time behavior coincides with ours, at least, up to t^{-2} order. As for the case at finite temperatures, the same remark as that concerning the model in Ref. 16 applies.

In Ref. 12, they start from a discrete version of the model (M2) with respect to the boson degrees of freedom and derive the dynamical equation for the position operator of the oscillator. Then, they make an *ad hoc* assumption for the memory kernel so that the symmetrized autocorrelation function of the fluctuating force per unit mass $\Phi_T(t)$ and the symmetrized autocorrelation function of the particle velocity $C_{vv}(t)$ take the form

$$\Phi_T(t) = \frac{\hbar\gamma^2}{M\tau_R}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\cdot\frac{\omega}{\omega^2+\gamma^2}\coth\frac{\beta\hbar\omega}{2}e^{i\omega t} \tag{1.45}$$

and

$$C_{vv}(t) = \frac{\hbar\gamma^2}{M\tau_R}\int_{-\infty}^{\infty}\frac{d\omega}{2\pi}\frac{\omega^3}{\omega^2+\gamma^2} \times\frac{1}{|\omega^2+(\gamma/\tau_R)[i\omega/(\gamma-i\omega)]|^2} \times\coth\frac{\beta\hbar\omega}{2}e^{i\omega t}, \tag{1.46}$$

respectively, where $\tau_R>0$ is a constant. They showed that, at finite temperatures, $\Phi_T(t)$ and $C_{vv}(t)$ decay exponentially as $t\to\infty$ and, only in an intermediate time region, display a long time tail as const t^{-2}. These results do not coincide with our results. This is due to the fact that, in the present case, the frequency spectrum runs from $-\infty$ to $+\infty$ as is seen from (1.45) and (1.46), which is an effect of the symmetrization. Let $\Phi_T^{(+)}(t)$ and $C_{vv}^{(+)}(t)$ be the functions defined by the rhs of (1.45) and (1.46) with the integral interval replaced by $[0,\infty)$, respectively. Then, we see that $\Phi_T^{(+)}(t)$ and $C_{vv}^{(+)}(t)$ display a power-law decay as

$$\Phi_T^{(+)}(t) \underset{t\to\infty}{\sim} \frac{i}{\pi M\tau_R\beta}\cdot\frac{1}{t}$$

and

$$C_{vv}^{(+)}(t) \underset{t\to\infty}{\sim} \frac{i\tau_R}{\pi M\beta}\cdot\frac{1}{t}.$$

At zero temperature, we have

$$\Phi_T(t) = \frac{\hbar\gamma^2}{M\tau_R\pi}\mathrm{Re}\left\{\int_0^\infty d\omega\frac{\omega}{\omega^2+\gamma^2}e^{i\omega t}\right\},$$

$$C_{vv}(t) = \frac{\hbar\gamma^2}{M\tau_R\pi}\mathrm{Re}\left\{\int_0^\infty d\omega\frac{\omega^3}{\omega^2+\gamma^2} \times\frac{e^{i\omega t}}{|\omega^2+(\gamma/\tau_R)[i\omega/(\gamma-i\omega)]|^2}\right\}.$$

By integration by parts and applying Theorem 2.1(a), we get

$$\Phi_T(t) \underset{t\to\infty}{\sim} -\frac{\hbar}{M\tau_R\pi}\cdot\frac{1}{t^2}, \tag{1.47}$$

$$C_{VV}(t) \underset{t\to\infty}{\sim} -\frac{\hbar\tau_R}{M\pi\gamma^2}\cdot\frac{1}{t^2}. \tag{1.48}$$

These results are not given in Ref. 12.

In Ref. 6, the authors start from the models (M1) and (M2) with a discrete boson degrees of freedom and calculate the momentum autocorrelation function $C(t)$ of the oscillator

$$C(t) = C_0(t) + C_1(t) \tag{1.49}$$

with

$$C_1(t) = \int_0^\infty \frac{1}{e^{\beta\hbar\lambda}-1}G(\lambda)\cos(\lambda t)d\lambda, \tag{1.50}$$

$$C_0(t) = \frac{1}{2}\int_0^\infty G(\lambda)e^{-i\lambda t}d\lambda, \tag{1.51}$$

where $G(\lambda)$ is a function depending on the choice of the models (M1) and (M2). The exact form of $G(\lambda)$ was not given and, by assuming that $G(\lambda)\sim\lambda^m$, they showed that, at low temperatures, $C_1(t)$ decays with a power law as $t\to\infty$. However, analysis on the long-time behavior of $C_1(t)$ at nonlow temperatures and of $C_0(t)$ (the correlation function at zero temperature) was not given. We note that (1.49) with (1.50) and (1.51) is exactly of the same form as $W_\beta(t)$ given by (1.22) with $m=0$. This is easily seen by change of variable. Thus we can apply our results and solve the problem left in Ref. 6. One can easily check that our general results yield as a special case the partial result on the long-time behavior of $C_1(t)$ obtained in Ref. 6.

The rest of the present paper is organized as follows: In Sec. II, we establish a limit theorem on an integral of Fourier's type. In Sec. III, applying the limit theorem, we prove Theorem 1.1. In Sec. IV, we discuss an example of ω. In Appendices A and B, sufficient conditions for (AI) to hold are given. In Appendix C, a necessary condition for exponential decay of Fourier transforms is given.

II. A LIMIT THEOREM

In this section, we prove a general limit theorem concerning an integral of Fourier's type.

Let g be a measurable function in $L^1((m,\infty),dx)$. We are concerned with the asymptotic behavior of the function

$$G(t) = \int_m^\infty g(x)e^{-itx^r}dx \quad (t>0) \tag{2.1}$$

as $t\to\infty$, where $m\geqslant 0$ and $r>0$ are constants. We assume the following:

$(A\text{-}g)_m$ There exists a meromorphic function $g(z)$ in $\mathbf{D}_{m,\theta}$ with the following properties:

(a) $\lim_{\epsilon \downarrow 0} g(x - i\epsilon) = g(x)$, a.e. $x \in (m, \infty)$,

(b) For all sufficiently small $\epsilon > 0$ and every $a,b \in (m, \infty)$ $(a < b)$,

$$\sup_{a < x < b} |g(x - i\epsilon)| \leqslant C_{a,b}$$

with a constant $C_{a,b} > 0$ independent of ϵ.

(c) For all sufficiently large $|z|$ $(z \in \mathbf{D}_{m,\theta})$

$$|g(z)| \leqslant C/|z|^q$$

with constants $C > 0$ and $q > 1$.

(d) $g(z)$ has no poles in $\{z \in \mathbf{D}_{m,\theta} \mid |z - m| < \epsilon_0\}$ with a constant $\epsilon_0 > 0$ and the limit

$$\lim_{z \to 0} \frac{g(m+z)}{z^{\mu_m}} \equiv g_m \neq 0$$

exists with constants g_m and $\mu_m > -1$.

Under the assumption $(A\text{-}g)_m$, we have the following theorem.

Theorem 2.1: (a) Let $m = 0$. Then

$$G(t) \underset{t \to \infty}{\sim} \frac{g_0 e^{-i\pi(\mu_0 + 1)/2r} \Gamma((\mu_0 + 1)/r)}{r} t^{-(\mu_0 + 1)/r}. \tag{2.2}$$

(b) Let $m > 0$. Then

$$G(t) \underset{t \to \infty}{\sim} \frac{g_m m^{(1-r)(\mu_m + 1)} e^{-it[m^r + (\mu_m + 1)/2]} \Gamma(\mu_m + 1)}{r^{\mu_m + 1}} \times t^{-(\mu_m + 1)}. \tag{2.3}$$

Proof: We first consider the case $m = 0$. Let

$$h_t(z) = e^{-itz^r} g(z), \quad t > 0.$$

Then $h_t(z)$ is meromorphic in $\mathbf{D}_{0,\theta}$. Let $\epsilon > 0$, $\delta > 0$ be sufficiently small, $L > 0$ be sufficiently large, and $0 < \theta_0 < \theta$. Let $\Gamma(\epsilon, \delta, L, \theta_0)$ be the curve in $\mathbf{D}_{0,\theta}$ given by

$$\Gamma(\epsilon, \delta, L, \theta_0) = \{x - i\epsilon \mid \delta \leqslant x \leqslant L\} \cup \{\delta - i\epsilon + se^{-i\theta_0} \mid 0 \leqslant s \leqslant L\} \cup \{\delta - i\epsilon + Le^{iu} \mid -\theta_0 \leqslant u \leqslant 0\}$$

with the anticlockwise orientation and $\mathbf{D}(\epsilon, \delta, L, \theta_0)$ be the interior domain of $\Gamma(\epsilon,\delta,L,\theta_0)$. Let $\{a_k\}_{k=1}^N$ be poles in $\mathbf{D}_{0,\theta_0}$ [N is finite by $(\mathrm{Ag})_0$ (b)–(d)]. Then, by applying the Cauchy integral theorem to the integral of $h_t(z)$ along $\Gamma(\epsilon,\delta,L,\theta_0)$, we get

$$\int_\delta^L h_t(x - i\epsilon)dx = \int_0^L h_t(\delta - i\epsilon + se^{-i\theta_0})e^{-i\theta_0} ds + \int_{-\theta_0}^0 h_t(\delta - i\epsilon + Le^{iu})iLe^{iu} du - 2\pi i \sum_{a_k \in \mathbf{D}(\epsilon,\delta,L,\theta_0)} \mathrm{Res}(h_t(z), a_k), \tag{2.4}$$

where $\mathrm{Res}(h_t(z), a_k)$ denotes the residue of $h_t(z)$ at $z = a_k$ [we take $\Gamma(\epsilon,\delta,L,\theta_0)$ so that $a_k \notin \Gamma(\epsilon,\delta,L,\theta)$ for all k]. We take θ_0 to satisfy $0 < r\theta_0 < \pi/2$. Then we have by $(A\text{-}g)_0$–(c)

$$|h_t(\delta - i\epsilon + Le^{iu})| \leqslant C/L^q$$

for all sufficiently large L with a constant $C > 0$ independent of L. Hence we get

$$\lim_{L \to \infty} \int_{-\theta_0}^0 h_t(\delta - i\epsilon + Le^{iu})iLe^{iu} du = 0. \tag{2.5}$$

For $s \geqslant s_0 > 0$, we have by $(\mathrm{Ag})_0$–(c)

$$|h_t(\delta - i\epsilon + se^{-i\theta_0})| \leqslant ae^{-bs^r}$$

with constants $a > 0$ and $b > 0$ independent of δ and ϵ. For $0 < s \leqslant s_0$, we have by $(A\text{-}g)_0$–(d)

$$|h_t(\delta - i\epsilon + se^{-i\theta_0})| \leqslant C$$

if $\mu_0 \geqslant 0$ and

$$|h_t(\delta - i\epsilon + se^{-i\theta_0})| \leqslant Cs^{\mu_0}$$

if $-1 < \mu_0 < 0$, where $C > 0$ is a constant independent of δ and ϵ. Therefore, $|h_t(\delta - i\epsilon + se^{-i\theta_0})|$ is dominated by an integrable function on $(0, \infty)$ independent of δ and ϵ. Thus by the Lebesgue dominated convergence theorem we get

$$\lim_{\epsilon,\delta \downarrow 0} \int_0^\infty h_t(\delta - i\epsilon + se^{-i\theta_0})e^{-i\theta_0} ds = \int_0^\infty h_t(se^{-i\theta_0})e^{-i\theta_0} ds. \tag{2.6}$$

Similarly, using $(A\text{-}g)_0$–(b)–(d), we can show that $|h_t(x - i\epsilon)|$ is dominated by an integrable function on $(0, \infty)$ independent of δ and $\epsilon > 0$. Thus by the Lebesgue dominated convergence theorem we get

$$\lim_{\delta \downarrow 0} \lim_{\epsilon \downarrow 0} \int_\delta^\infty h_t(x - i\epsilon)dx = \int_0^\infty h_t(x)dx. \tag{2.7}$$

Taking the limit $L \to \infty$ first, $\epsilon \downarrow 0$ second, and $\delta \downarrow 0$ finally in (2.4) and using (2.5)–(2.7), we obtain

$$G(t) = G_1(t) + G_2(t) \tag{2.8}$$

with

$$G_1(t) = \int_0^\infty h_t(se^{-i\theta_0})e^{-i\theta_0} ds, \tag{2.9}$$

$$G_2(t) = -2\pi i \sum_{a_k \in \mathbf{D}_{0,\theta_0}} \mathrm{Res}(h_t(z), a_k). \tag{2.10}$$

Note that $(A\text{-}g)_0$ implies that the number of poles of $g(z)$ in $\mathbf{D}_{0,\theta_0}$ is finite. Hence the sum with respect to the residue in (2.10) is a finite one and we have

$$|G_2(t)| \leqslant e^{-at} P(t) \tag{2.11}$$

with a constant $a > 0$ and a polynomial $P(t)$ in t. On the other hand, by the change of variable $s \to t^{1/r}s$, we have

$$G_1(t)=\frac{1}{t^{1/r}}\int_0^\infty \exp[-i(se^{-i\theta_0})^r]g\left(\frac{se^{-i\theta_0}}{t^{1/r}}\right)e^{-i\theta_0}\,ds$$

$$=\frac{1}{t^{(\mu_0+1)/r}}\int_0^\infty \exp[-i(se^{-i\theta_0})^r]$$

$$\times(se^{-i\theta_0})^{\mu_0}u\left(\frac{se^{-i\theta_0}}{t^{1/r}}\right)e^{-i\theta_0}\,ds, \tag{2.12}$$

where

$$u(z)=g(z)/z^{\mu_0}.$$

It follows from $(A\text{–}g)_0$–(c) and (d) that, for all $s>0$,

$$|u(se^{-i\theta_0})|\leqslant C/(1+s)^{\mu_0+q}<C$$

with a constant $C>0$ (note that $\mu_0+q>\mu_0+1>0$). Therefore, by the dominated convergence theorem, we get from (2.12)

$$\lim_{t\to\infty} t^{(\mu_0+1)/r}G_1(t)=g_0I(\theta_0) \tag{2.13}$$

with

$$I(\tau)=\int_0^\infty \exp[-i(se^{-i\tau})^r]$$

$$\times(se^{-i\tau})^{\mu_0}e^{-i\tau}\,ds. \tag{2.14}$$

The function $z^{\mu_0}\exp(-iz^r)$ is analytic in $\{z\in\mathbf{C}|\ -\pi<\arg z<\pi,\ z\neq 0\}$ and, if $-\pi<r(\arg z)<0$, then it decays exponentially as $|z|\to\infty$. Therefore, by the Cauchy theorem, we have

$$I(\theta_0)=I(\pi/2r)$$

$$=e^{-i(\mu_0+1)\pi/2r}\int_0^\infty e^{-s^r}s^{\mu_0}\,ds$$

$$=\frac{e^{-i(\mu_0+1)\pi/2r}\Gamma((\mu_0+1)/r)}{r}. \tag{2.15}$$

Combining (2.11), (2.13)–(2.15) with (2.8), we obtain (2.2).

We next consider the case $m>0$. By the change of variable $x^r\to x^r-m^r$, we have

$$G(t)=e^{-itm^r}\int_0^\infty \tilde{g}(x)e^{-itx^r}\,dx$$

with

$$\tilde{g}(x)=\frac{x^{r-1}g((x^r+m^r)^{1/r})}{(x^r+m^r)^{(r-1)/r}}.$$

Therefore, the problem is reduced to the case $m=0$; we need only to check that the function $\tilde{g}(x)$ satisfies $(A\text{–}\tilde{g})_0$.

Let $\tilde{g}(z)$ be $\tilde{g}(x)$ with z in place of x. Then $\tilde{g}(z)$ is meromorphic in $\mathbf{D}_{0,\theta}$ and

$$\lim_{\epsilon\downarrow 0}\tilde{g}(x-i\epsilon)=\tilde{g}(x),\quad \text{a.e. } x\in(0,\infty).$$

Hence $\tilde{g}$ satisfies $(A\text{–}\tilde{g})_0$ – (a).

To prove $(A-\tilde{g})$–(b), we note that, for $0<a<x<b<\infty$,

$$[(x-i\epsilon)^r+m^r]^{1/r}=(x^r+m^r)^{1/r}$$

$$-\frac{i\epsilon}{(x^r+m^r)^{(r-1)/r}x^{1-r}}+O(\epsilon^2)$$

as $\epsilon\downarrow 0$. Therefore, by the assumption $(A\text{–}g)_m$–(b), we get

$$\sup_{a<x<b}|g([(x-i\epsilon)^r+m^r]^{1/r})|\leqslant C_1$$

with a constant $C_1>0$ independent of ϵ. Further, it is easy to see that

$$\sup_{a<x<b}\left|\frac{(x-i\epsilon)^{r-1}}{[(x-i\epsilon)^r+m^r]^{(r-1)/r}}\right|\leqslant C_2$$

with a constant $C_2>0$ independent of ϵ. Therefore we get

$$\sup_{a<x<b}|\tilde{g}(x-i\epsilon)|\leqslant C_3$$

with a constant $C_3>0$ independent of ϵ. Thus $\tilde{g}$ satisfies $(A\text{–}\tilde{g})_0$–(b).

For sufficiently large $|z|$, we have from $(A\text{–}g)_m$–(c)

$$|\tilde{g}(z)|\leqslant\frac{\text{const}|z|^{r-1}}{|z^r+m^r|^{(r-1)/r}(z^r+m^r)^{q/r}}$$

$$\leqslant\frac{\text{const}}{|z|^q}.$$

Therefore $(A\text{–}\tilde{g})_0$–(c) follows.

We have

$$(z^r+m^r)^{1/r}\sim m+m^{1-r}z^r/r$$

as $z\to 0$. Hence $(A\text{–}g)_m$–(d) implies that $\tilde{g}(z)$ has no poles in $\{z\in\mathbf{D}_{0,\theta}|\ |z|<\epsilon_0\}$ with a sufficiently small ϵ_0 and

$$\lim_{z\to 0}\frac{\tilde{g}(z)}{z^{r(\mu_m+1)-1}}=\frac{g^m m^{(1-r)(\mu_m+1)}}{r^{\mu_m}}.$$

Therefore $(A\text{–}\tilde{g})_0$–(d) holds with $\tilde{g}_0=g_m m^{(1-r)(\mu_m+1)}/r^{\mu_m}$ and $\mu_0=r(\mu_m+1)-1$.

Thus we can apply the result in the case $m=0$ to $\tilde{g}(x)$ and get (2.3).

III. PROOF OF THEOREM 1.1

We can write

$$W(t,f)=\int_m^\infty g(x)e^{-itx^r}\,dx$$

with

$$g(x)=\frac{f(x)I^{(j)}(x)}{x^{2r\alpha+j}|D_+^{(j)}(x)|^2}.$$

We shall show that g satisfies $(A\text{–}g)_m$ in Sec. II. Then Theorem 1.1 follows from an application of Theorem 2.1.

Let

$$D_+^{(j)}(z)=D^{(j)}(z)-2\pi iI^{(j)}(z),\quad z\in\mathbf{C}_m,$$

and define

$$g(z)=\frac{f(z)I^{(j)}(z)}{z^{2r\alpha+j}D_+^{(j)}(z)D^{(j)}(z)}.$$

Then, $g(z)$ is meromorphic in $\mathbf{D}_{m,\theta}$. Note that $D_\pm^{(j)}(x)$ is given as

$$D_\pm^{(j)}(x)=a_0^{(j)}-a_1^{(j)}x+\delta_{1,j}\int\rho(\mathbf{k})^2\,d\mathbf{k}$$

$$+\mathrm{P}\int\frac{\omega(\mathbf{k})^j\rho(\mathbf{k})^2}{x-\omega(\mathbf{k})}\,d\mathbf{k}$$

$$+2\pi iI^{(j)}(x),$$

where P denotes the principal value. Hence we have

$$\lim_{\epsilon \downarrow 0} D_+^{(j)}(x - i\epsilon) = D_+^{(j)}(x), \quad \text{a.e. } x \in (m, \infty).$$

Therefore, together with the fact $D_+(x) = \overline{D_-(x)}$, we get

$$\lim_{\epsilon \downarrow 0} g(x - i\epsilon) = g(x).$$

Thus g satisfies $(A\text{–}g)_m$–(a).

Property $(A\text{–}g)_m$–(b) follows from (AI)(b), (AII)(d) and properties of f.

By (1.17) and (AII–b), we have for all sufficiently large $|x|$

$$\inf_{\epsilon > 0} |D_+^{(j)}(x - i\epsilon)| \geqslant Cx$$

with a constant $C > 0$. Further, it is easy to see that, for sufficiently large $|z|$ with $\operatorname{Im} z < -\epsilon$, $z \in \mathbf{D}_{m,\theta}$,

$$|D_+^{(j)}(z)| \geqslant C_\epsilon |z|, \quad |D^{(j)}(z)| \geqslant C_\epsilon |z|$$

with a constant C_ϵ dependent on $\epsilon > 0$. Combining these properties with (AI)(b), (AII)(b), and (f I), we have for sufficiently large $|z|$ $(z \in \mathbf{D}_{m,\theta})$

$$|g(z)| \leqslant \text{const}/|z|^{2r\alpha + j + q_j + 2}.$$

Since we assume that $2r\alpha + j + q_j + 1 > 0$, we have $2r\alpha + j + q_j + 2 > 1$. Therefore, $(A\text{–}g)_m$–(c) follows.

The poles of $g(z)$ come only for the zeros of $D_+^{(j)}(z)$ and the poles of $f(z)$. The zeros of $D_+^{(j)}(z)$ are discrete and do not accumulate in $\mathbf{D}_{m,\theta}$. It is obvious that $D_+^{(j)}(z)$ [resp. $f(z)$] has no zeros (resp. poles) in a small neighborhood of $z = m$ in $\mathbf{D}_{m,\theta}$ [recall that we assume $d_m^{(j)} > 0$, see (1.15)]. As for the asymptotic behavior of $g(z)$ as $z \to m$, we have

$$g(z) \underset{z \to 0}{\sim} \frac{A_0^{(j)} f_0}{D_+^{(j)}(0) D^{(j)}(0)} z^{-2r\alpha - j - \alpha_0(f) + p_j(0)}$$

in the case $m = 0$ and

$$g(m + z) \underset{z \to 0}{\sim} \frac{A_m^{(j)} f_m}{m^{2r\alpha + j} D_+^{(j)}(m) D^{(j)}(m)} z^{-\alpha_m(f) + p_j(m)}$$

in the case $m > 0$. Therefore, g satisfies $(A\text{–}g)_m$–(d).

Thus we can apply Theorem 2.1 to obtain Theorem 1.1.

IV. AN EXAMPLE

In this section we consider an example of ω and see how $p_j(m)$ is determined by the dimension d and the infrared property of ω and ρ, where $p_j(m)$ is defined by (AII)–(c).

Let

$$\omega_1(x) = (x^2 + m^{1/\lambda})^\lambda, \quad x \geqslant 0, \tag{4.1}$$

with a parameter $\lambda > 0$ (see also Appendix A). Then the inverse function $\eta(x)$ of $\omega_1(x)$ is given by

$$\eta(x) = (x^{1/\lambda} - m^{1/\lambda})^{1/2}, \quad x \geqslant m. \tag{4.2}$$

Let ρ_1 be a continuous function on $[0, \infty)$ satisfying

$$\rho_1(x) > 0, \quad x \in (0, \infty),$$

and ρ be given by

$$\rho(\mathbf{k}) = \rho_1(|\mathbf{k}|), \quad k \in \mathbf{R}^d. \tag{4.3}$$

We assume that ρ_1 has an analytic continuation $\rho_1(z)$ onto the domain $\mathbf{D}_{0,\theta_\rho}$ with a constant $\theta_\rho \in (0, \pi/2)$ such that

$$\lim_{\epsilon \downarrow 0} \rho_1(x - i\epsilon) = \rho_1(x), \quad x \in (0, \infty).$$

With ω_1 and ρ_1 given above, the function $I^{(j)}(x)$ defined by (1.26) takes the form

$$I^{(j)}(x) = \frac{V_d \eta(x)^{d-2} x^{j + (1-\lambda)/\lambda} \rho_1(\eta(x))^2}{2\lambda}, \quad x \in (m, \infty), \tag{4.4}$$

where

$$V_d = \int_{S^{d-1}} dS. \tag{4.5}$$

Lemma 4.1: Assume that, for sufficiently large $|z|$ $(z \in \mathbf{D}_{0,\theta_\rho})$,

$$|\rho_1(z)| \leqslant C/|z|^{q(\rho)} \tag{4.6}$$

with constants $C > 0$ and $q(\rho) > 0$ and that

$$\lim_{z \to 0} \frac{\rho_1(z)}{z^{\gamma(\rho)}} = \rho_0 \tag{4.7}$$

with constants $\gamma(\rho) \geqslant 0$ and $\rho_0 \neq 0$. Suppose that

$$q(\rho)/\lambda - j + 1 - d/2\lambda \geqslant 0 \tag{4.8}$$

and, in the case $m = 0$,

$$(d-2)/2\lambda + j + (1-\lambda)/\lambda + \gamma(\rho)/\lambda \geqslant 0. \tag{4.9}$$

Then $I^{(j)}(x)$ satisfies (AII)(a)–(c) with q_j given by

$$q_j = q(\rho)/\lambda - j + 1 - d/2\lambda \tag{4.10}$$

and

$$p_j(m) = \begin{cases} (d-2)/2\lambda + j + (1-\lambda)/\lambda + \gamma(\rho)/\lambda, \\ \quad m = 0, \\ (d-2)/2 + \gamma(\rho), \\ \quad m > 0 \end{cases} \tag{4.11}$$

Proof: For a sufficiently small $\theta \in (0, \pi/2)$, $\eta(z)$ is defined and analytic in $\mathbf{D}_{m,\theta}$ and the image of η is included in $\mathbf{D}_{0,\theta_\rho}$. Therefore the function

$$I^{(j)}(z) = \frac{V_d \eta(z)^{d-2} z^{j + (1-\lambda)/\lambda} \rho_1(\eta(z))^2}{2\lambda} \tag{4.12}$$

is analytic in $\mathbf{D}_{m,\theta}$. It is obvious that (AII)(a) holds.

For sufficiently large $|z|$ $(z \in \mathbf{D}_{m,\theta})$, we have

$$C_1 |z|^{1/2\lambda} \leqslant |\eta(z)| \leqslant C_2 |z|^{1/2\lambda}$$

with constants $C_k > 0$, $k = 1, 2$, and hence

$$|I^{(j)}(z)| \leqslant \text{const} \quad |z|^{-q_j},$$

with q_j given by (4.10). Therefore, (AII)–(b) follows.

Let $m > 0$. Then we have

$$\eta(m + z) \underset{z \to 0}{\sim} (m^{(1-\lambda)/\lambda}/\lambda)^{1/2} z^{1/2}$$

and hence

$$\rho_1(\eta(m + z)) \underset{z \to 0}{\sim} \rho_0 (m^{(1-\lambda)/\lambda}/\lambda)^{\gamma(\rho)/2} z^{\gamma(\rho)/2}.$$

Therefore we get for $m > 0$

$$I^{(j)}(m + z) \underset{z \to 0}{\sim} \frac{\rho_0^2 V_d}{2\lambda} (m^{(1-\lambda)/\lambda}/\lambda)^{\gamma(\rho) + (d-2)/2} \times m^{j + (1-\lambda)/\lambda} z^{\gamma(\rho) + (d-2)/2}.$$

In the case $m=0$, we have

$$I^{(j)}(z) \underset{z\to 0}{\sim} \frac{\rho_0^2 V_d}{2\lambda} z^{(d-2)/2\lambda + j + (1-\lambda)/\lambda + \gamma(\rho)/\lambda}.$$

Thus (AII)–(c) is satisfied with $p_j(m)$ given by Eq. (4.11). □

Remark: Condition (AII)(d) follows from the same assumption as in Lemma B1 (see Appendix B).

It is interesting to compare the decay order of $W(t)$ in the case $m=0$ [i.e., λ_j given by (1.34)] with that in the case $m>0$ [i.e., $p_j(m)+1$]. By (4.11), we have

$$p_j(0) = p_j(m)/\lambda + j + (1-\lambda)/\lambda,$$

where we take $m>0$.

Hence we get

$$\Delta_j \equiv \lambda_j - (p_j(m)+1)$$
$$= ((1-\lambda r)/\lambda r)(p_j(m)+1) - 2\alpha.$$

If $\Delta_j > 0$ (resp. $\Delta_j < 0$), then $W(t)$ with $m=0$ decays faster (resp. slower) than $W(t)$ with $m>0$.

In the models (M1)–(M3), Δ_j is given as follows:

(M1): In this case, we have $r=1$, $\alpha = j = 0$. Hence we get

$$\Delta_0 = ((1-\lambda)/\lambda)(p_0(m)+1).$$

(M2): In this case, we have $r=\alpha=\frac{1}{2}$, $j=0$. Hence we get

$$\Delta_0 = ((2-\lambda)/\lambda)(p_0(m)+1) - 1.$$

(M3): In this case, we have $r=\alpha=\frac{1}{2}$, $j=1$. Hence we get

$$\Delta_1 = ((2-\lambda)/\lambda)(p_1(m)+1) - 1.$$

In the standard case, λ is taken as $\lambda = \frac{1}{2}$ (resp. 1, 1) in the model (M1) [resp. (M2), (M3)]. Therefore, in this case, $W(t)$ with $m=0$ decays faster than $W(t)$ with $m>0$ in each model.

ACKNOWLEDGMENT

This work was supported in part by the Grant-In-Aid 63740063 and 62460001 for science research from the Ministry of Education, Japan.

APPENDIX A: UNIFORM BOUNDEDNESS OF $\Phi_\rho^{(j)}(x-i\epsilon)$

In this appendix, we consider a sufficient condition for $\Phi_\rho^{(j)}(z)$ to satisfy (AI)(a). We define the function $J^{(j)}(x)$ on $\mathbf{R}$ by

$$J^{(j)}(x) = \begin{cases} I^{(j)}(x), & x>m, \\ 0, & x\leqslant m, \end{cases} \tag{A1}$$

where $I^{(j)}(x)$ is given by (1.26).

Lemma A1: Suppose that $J^{(j)}$ is in $L^p(\mathbf{R})$ with some $p\in(1,\infty)$ and satisfies the Lipschitz condition

$$|J^{(j)}(x+h) - J^{(j)}(x)| < K|h|^a \tag{A2}$$

uniformly in x, as $h\to 0$, with constants $a\in(0,1)$ and $K>0$. Then there exists a constant $C>0$ such that, for all $x\in\mathbf{R}$ and $\epsilon>0$,

$$|\Phi_\rho^{(j)}(x-i\epsilon)| \leqslant C. \tag{A3}$$

Proof: We can write

$$\Phi_\rho^{(j)}(x-i\epsilon) = (Q_\epsilon * J^{(j)})(x) + i(P_\epsilon * J^{(j)}), \tag{A4}$$

where P_ϵ and Q_ϵ are the Poisson and the conjugate Poisson kernel, respectively,

$$P_\epsilon(x) = \epsilon/(x^2+\epsilon^2), \quad Q_\epsilon(x) = x/(x^2+\epsilon^2),$$

and $*$ denotes the convolution. Let $\Phi_\rho^{(j)}(x)$ be the Hilbert transform of $J^{(j)}$,

$$\Phi_\rho^{(j)}(x) = P\int \frac{J^{(j)}(s)}{x-s}\,ds.$$

Then, by the assumption $J^{(j)}\in L^p(\mathbf{R})$ and a general theorem [e.g., Ref. 19 (Chap. VI, Lemma 1.5)], we have

$$Q_\epsilon * J^{(j)} = P_\epsilon * \Phi_\rho^{(j)}.$$

Further, condition (A2) implies that $\Phi_\rho^{(j)}(x)$ is continuous in $x\in\mathbf{R}$ with $\Phi_\rho^{(j)}(x)\to 0$ ($|x|\to\infty$) and hence, in particular, $\Phi_\rho^{(j)}$ is bounded [e.g., Ref. 18 (§5.15)]. Similarly, it follows that $J^{(j)}(x)$ is bounded. It is well-known (or easy to prove) that

$$|(P_\epsilon * f)(x)| \leqslant \pi \|f\|_\infty$$

for all $f\in L^\infty(\mathbf{R})$. Therefore, we get

$$|(Q_\epsilon * J^{(j)})(x)| \leqslant \pi \|\Phi_\rho^{(j)}\|_\infty,$$
$$|(P_\epsilon * J^{(j)})(x)| \leqslant \pi \|J^{(j)}\|_\infty.$$

From these estimates and (A4), we obtain (A3). □

Remark: Condition (A2) implies the continuity of $I^{(j)}(x)$ and hence

$$\lim_{x\downarrow m} I^{(j)}(x) = 0.$$

A sufficient condition for (A2) to hold is the following lemma.

Lemma A2: Let

$$u(x) = \eta(x)^{d-1}\eta'(x), \quad x\in(m,\infty), \tag{A5}$$

and ρ be given by

$$\rho(\mathbf{k}) = \rho_1(|\mathbf{k}|), \quad k\in\mathbf{R}^d, \tag{A6}$$

with a continuous function ρ_1 on $[0,\infty)$. Assume that

(a) $\lim_{x\downarrow m} u(x) = 0$,

(b) $\eta(x) \sim C_\eta x^{\delta(\eta)}$, (A7)

$u(x) \sim C_u x^{\delta(u)}$, (A8)

$\rho_1(x) \sim C_\rho x^{-\delta(\rho)}$, (A9)

as $x\to\infty$, with strictly positive constants $C_\eta, C_u, C_\rho, \delta(\eta)$, $\delta(u)$, and $\delta(\rho)$.

(c) For all sufficiently small $|h|<1$ ($h\in\mathbf{R}$)

$$|\eta(x+h)-\eta(x)| \leqslant P_\eta(x)|h|^{\epsilon(\eta)}, \quad x\in(m,\infty), \tag{A10}$$

$$|u(x+h)-u(x)| \leqslant P_u(x)|h|^{\epsilon(u)}, \quad x\in(m,\infty), \tag{A11}$$

$$|\rho_1(x+h)-\rho_1(x)| \leqslant P_\rho(x)|h|^{\epsilon(\rho)}, \quad x\in[0,\infty), \tag{A12}$$

with strictly positive constants $\epsilon(\eta)$, $\epsilon(u)$, $\epsilon(\rho)$, and non-negative polynomially bounded continuous functions P_η, P_u, and P_ρ on $\mathbf{R}$. Suppose that

$$2\delta(\eta)\delta(\rho) - \delta(u) \geqslant 0, \tag{A13}$$

$$\sup_{x>m} P_u(x) x^{j-2\delta(\eta)\delta(\rho)} < \infty, \tag{A14}$$

$$\operatorname*{sum}_{x>m} P_\rho(\eta(x))P_\eta(x)^{\epsilon(\rho)}x^{j+\delta(u)-\delta(\eta)\delta(\rho)}<\infty. \quad \text{(A15)}$$

Then, (A2) holds.

Proof: Let $x\geqslant m$ and $|h|<1$ be sufficiently small. We set $I^{(j)}(m)\equiv 0$. Then we have

$$\begin{aligned}J^{(j)}(x+h)-J^{(j)}(x)&=I^{(j)}(x+h)-I^{(j)}(x)\\&=V_d(F_1(x,h)+F_2(x,h)+F_3(x,h)),\end{aligned}$$

where V_d is given by (4.5) and

$$\begin{aligned}F_1(x,h)&=(u(x+h)-u(x))(x+h)^j\rho_1(\eta(x))^2,\\F_2(x,h)&=\delta_{1,j}u(x)\rho_1(\eta(x))^2h,\\F_3(x,h)&=u(x+h)(x+h)^j[\rho_1(\eta(x+h))+\rho_1(\eta(x))]\\&\quad\times[\rho_1(\eta(x+h))-\rho_1(\eta(x))].\end{aligned}$$

(In the case $x=m$, we take $h>0$.) By the assumption, we can estimate as

$$|F_1(x,h)|\leqslant\text{const}\frac{(x^j+1)P_u(x)|h|^{\epsilon(u)}}{1+x^{2\delta(\eta)\delta(\rho)}},$$

$$|F_2(x,h)|\leqslant\text{const}\frac{x^{\delta(u)}|h|}{1+x^{2\delta(\eta)\delta(\rho)}},$$

$$|F_3(x,h)|\leqslant\text{const}\frac{(1+x)^{\delta(u)+j}P_\rho(\eta(x))P_\eta(x)^{\epsilon(\rho)}|h|^{\epsilon(\eta)\epsilon(\rho)}}{(1+x^{\delta(\eta)\delta(\rho)})}.$$

These estimates together with (A13)–(A15) imply (A2). In the case $x<m$, (A2) trivially holds. □

We conclude this section with an example of ω_1 and ρ_1.

Let ω_1 be given by (4.1). We suppose that $\lambda\in(0,1]$. Then, $u(x)$ defined by (A5) takes the form

$$\begin{aligned}u(x)&=\frac{x^{(1-\lambda)/\lambda}(x^{1/\lambda}-m^{1/\lambda})^{(d-2)/2}}{2\lambda}\\&=\frac{x^{(1-\lambda)/\lambda}\eta(x)^{d-2}}{2\lambda},\end{aligned} \quad \text{(A16)}$$

where $\eta(x)$ is the inverse function of ω_1 [see (4.2)]. If $d>2\lambda$, then we have

$$\eta(x)\sim x^{1/2\lambda},\quad u(x)\sim\frac{x^{(d-2\lambda)/2\lambda}}{2\lambda} \quad \text{(A17)}$$

as $x\to\infty$ and hence (A7) and (A8) hold.

To prove (A10) and (A11), we first note the following elementary estimate.

Lemma A3: Let $\alpha>0$. Then, for all sufficiently small $\epsilon>0$ and $|h|>0$,

$$|(x+h)^\alpha-x^\alpha|\leqslant C(1+x^{\alpha-\epsilon})|h|^\epsilon,\quad x\geqslant 0, \quad \text{(A18)}$$

with a constant $C>0$, where, in the case $x=0$, we take $h>0$.

Proof: It is sufficient to prove (A18) for $h>0$ and $x>0$. Let $\delta\in(0,1)$ be fixed and $x\geqslant h/\delta$. Then we have $h/x\leqslant\delta<1$ and hence

$$(x+h)^\alpha=x^\alpha(1+h/x)^\alpha\leqslant x^\alpha(1+C(h/x))$$

with a constant $C>0$. Therefore we get

$$\begin{aligned}0<(x+h)^\alpha-x^\alpha&\leqslant Cx^\alpha(h/x)\\&\leqslant C\delta^{1-\epsilon}x^{\alpha-\epsilon}h^\epsilon.\end{aligned}$$

In the case $x\leqslant h/\delta$, we have

$$0<(x+h)^\alpha-x^\alpha\leqslant\{(1+1/\delta)^\alpha+1/\delta^\alpha\}h^\alpha.$$

Thus (A18) follows. □

Lemma A4: Let $\eta(x)$ be the inverse function of ω_1 [see (4.2)]. Then, for all sufficiently small $\epsilon>0$ and $|h|$, we have

$$|\eta(x+h)-\eta(x)|\leqslant\text{const}(1+x^{(1/2\lambda)-\epsilon})|h|^\epsilon \quad \text{(A19)}$$

for the case $m=0$, and

$$|\eta(x+h)-\eta(x)|\leqslant\text{const}[x^{(1-\lambda)/2\lambda}+(x^{1/\lambda}-m^{1/\lambda})^{(1/2)-\epsilon}x^{(1-\lambda)\epsilon/\lambda}]|h|^\epsilon \quad \text{(A20)}$$

for the case $m>0$.

Proof: We first consider the case $m=0$. Then we have

$$\eta(x+h)-\eta(x)=(x+h)^{1/2\lambda}-x^{1/2\lambda},\quad x>0.$$

Therefore, by Lemma A3, we get (A19).

Let $m>0$ and $m>h>0$. Then we have

$$\eta(x+h)\leqslant\{x^{1/\lambda}(1+C(h/x))-m^{1/\lambda}\}^{1/2},\quad x\geqslant m,$$

with a constant $C>0$. Therefore, we get

$$0<\eta(x+h)-\eta(x)\leqslant x^{(1-\lambda)/2\lambda}\{(X+Ch)^{1/2}-X^{1/2}\}$$

with

$$X=(x^{1/\lambda}-m^{1/\lambda})x^{(\lambda-1)/\lambda}.$$

Therefore, using Lemma A3, we get (A20). □

We introduce a function $\theta_+(x)$ on $[0,\infty)$ by

$$\theta_+(x)=\begin{cases}1, & x>0,\\0, & x=0.\end{cases} \quad \text{(A21)}$$

Lemma A5: Let u be given by (A16).

(a) Let $m=0$ and $d\geqslant 2\lambda$. Then, for all sufficiently small $\epsilon>0$ and $|h|>0$,

$$|u(x+h)-u(x)|\leqslant\text{const}(1+\theta_+(d-2\lambda)x^{(d/2\lambda)-1-\epsilon})|h|^\epsilon,\quad x\geqslant 0, \quad \text{(A22)}$$

(b) Let $m>0$ and $d\geqslant 2$. Then, for all sufficiently small $\epsilon>0$ and $|h|>0$,

$$|u(x+h)-u(x)|\leqslant\text{const}\{1+\theta_+(d-2)(x+1)^{[d-2-2\lambda\epsilon+2(1-\lambda)\epsilon]/2\lambda}+\theta_+(1-\lambda)x^{(d-2\lambda-2\epsilon)/2\lambda}\}|h|^\epsilon,\quad x\geqslant m. \quad \text{(A23)}$$

Proof: (a) The case $d=2\lambda$ is trivial, since we have $u(x)=1/2\lambda$. Let $d>2\lambda$. Then, we have

$$u(x+h)-u(x)=(1/2\lambda)\{(x+h)^{(d-2\lambda)/2\lambda}-x^{(d-2\lambda)/2\lambda}\}.$$

Therefore, applying Lemma A3, we get (A.22).

(b) We prove (A.23) only in the case $d>2$ and $0<\lambda<1$. The other cases are easier. We write

$$\begin{aligned}u(x+h)-u(x)&=(1/2\lambda)[(x+h)^{(1-\lambda)/\lambda}-x^{(1-\lambda)/\lambda}]\eta(x+h)^{d-2}\\&\quad+(1/2\lambda)x^{(1-\lambda)/\lambda}(\eta(x+h)-\eta(x))\\&\quad\times\sum_{i=0}^{d-3}\eta(x+h)^{d-3-i}\eta(x)^i.\end{aligned}$$

Then, Lemmas A3, A4, (A.19), and (A.20) give (A.23). □

Lemmas A4 and A5 show that (A.10) and (A.11) hold.

As for ρ_1, one can easily find a number of examples which satisfy the conditions required. For example, one can take functions of the form

$$\rho_1(x) = Q(x)/[1 + P(x)], \quad x \geqslant 0$$

where $P(x)$ and $Q(x)$ are polynomials with positive coefficients and $\deg P - \deg Q > 0$ is sufficiently large.

APPENDIX B: ESTIMATE FOR $|D^{(j)}(x - i\epsilon)|$

In this appendix, we consider a sufficient condition for (AI)(b) to hold.

Lemma B1: Suppose that $\rho(\mathbf{k}) > 0$ for all $\mathbf{k} \neq 0$. Then, under the assumption in Lemma A1, (AI)(b) holds.

Proof: We have

$$D_{-}^{(j)}(x) = a_0^{(j)} - a_1^{(j)} x + \delta_{1j} \int \rho(\mathbf{k})^2 \, d\mathbf{k} + \Phi_\rho^{(j)}(x) + i\pi I^{(j)}(x). \tag{B1}$$

As stated in the proof of Lemma A1, $\Phi_\rho^{(j)}(x)$ is continuous in $x \in \mathbf{R}$ and

$$\Phi_\rho^{(j)}(x) \to 0 \tag{B2}$$

as $|x| \to \infty$. Therefore $D_{-}^{(j)}(x)$ is continuous on $\mathbf{R}$. [For $x < m$, $D_{-}^{(j)}(x) = D^{(j)}(x)$.] By the Remark after the proof of Lemma A1 and condition $d_m^{(j)} > 0$ [see (1.15)], we have

$$\inf_{x \leqslant m + \delta} |D_{-}^{(j)}(x)| > 0$$

with some constant $\delta > 0$ and hence, taking (1.16) into account, we get

$$\inf_{\substack{\epsilon > 0 \\ x \in [m, m+\delta)}} |D^{(j)}(x - i\epsilon)| > 0. \tag{B3}$$

It follows from (B1) and (B2) that

$$|D_{-}^{(j)}(x)| \geqslant \text{const}\, x$$

for $x \geqslant R$ with a sufficiently large $R > 0$. Combining this estimate with (1.16), we get

$$\inf_{\substack{\epsilon > 0 \\ x > R}} |D^{(j)}(x - i\epsilon)| > 0. \tag{B4}$$

The positivity of ρ together with that of η and η' implies that

$$\inf_{\delta < x < R} I^{(j)}(x) > 0.$$

Hence we have

$$\inf_{\delta < x < R} |D_{-}^{(j)}(x)| > 0,$$

which, together with (1.16), gives

$$\inf_{\substack{\epsilon > 0 \\ \delta < x < R}} |D^{(j)}(x - i\epsilon)| > 0. \tag{B5}$$

Estimates (B3)–(B5) yields (AI)(b). □

APPENDIX C: A NECESSARY CONDITION FOR EXPONENTIAL DECAY OF FOURIER TRANSFORMS

In this section, we give a necessary condition for the Fourier transform $\hat{f}(t)$ of an $L^2(\mathbf{R})$ function $f(x)$ to decay exponentially as $t \to \infty$ and hence a sufficient condition for f under which $\hat{f}$ does not decay exponentially.

Proposition C1: Let $f \neq 0$ be in $L^2(\mathbf{R})$ and $\hat{f}(t)$ be the Fourier transform of f,

$$\hat{f}(t) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-itx} f(x) dx, \quad t \in \mathbf{R}.$$

Suppose that there exist positive constants $C > 0$ and $a > 0$ such that, for all $t \in \mathbf{R}$,

$$|\hat{f}(t)| \leqslant Ce^{-a|t|} \tag{C1}$$

Then, $\operatorname{supp} f$, the support of f, equals the whole real line $\mathbf{R}$.

Proof: [cf. Ref. 15 (§7.3, Theorem 3.3)] It follows from (C.1) and a general theorem [e.g., Ref. 3 (§9.3, Theorem 9.13)] that $f(x)$ has an analytic continuation $\tilde{f}(z)$ to the set $S_a \equiv \{z \mid |\operatorname{Im} z| < a\}$. Suppose that $\operatorname{supp} f \neq \mathbf{R}$. Then, the set $A = \mathbf{R}/\operatorname{supp} f$ is an open set in $\mathbf{R}$ and $f(x) = 0$ for all $x \in A$. Hence $\tilde{f}(z) = 0$ for all $z \in A$. Therefore, it follows from an elementary property of holomorphic function that $f(z) = 0$ for all $z \in S_a$. But this contradicts the assumption that $f(x)$ does not vanish identically. Thus $\operatorname{supp} f$ must be equal to $\mathbf{R}$. □

From Proposition C.1, we have

Corollary C2: Let $g \neq 0$ be in $L^2(\mathbf{R})$ and put

$$G_{a,b}(t) = \int_a^b e^{-itx} g(x) dx.$$

Suppose that $-\infty \leqslant a < b < \infty$ or $-\infty < a < b \leqslant \infty$. Then, $G_{a,b}(t)$ does not decay exponentially as $|t| \to \infty$.

Proof: We need only to apply Proposition C1 with $f = \sqrt{2\pi}\chi_{[a,b]} g$, where $\chi_{[a,b]}$ is the characteristic function on $[a,b]$: $\chi_{[a,b]}(x) = 1$, $x \in [a,b]$; $\chi_{[a,b]}(x) = 0$, $x \notin [a,b]$. □

[1] A. Arai, "Perturbation of embedded eigenvalues in Fock spaces: A general class of exactly soluble models." Hokkaido Univ. Preprint Series in Math. No. 21, February, 1988.

[2] M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. I: Functional Analysis (Academic, New York, 1972).

[3] M. Reed and B. Simon, *Methods of Modern Mathematical Physics*, Vol. II; Fourier Analysis, Self-Adjointness (Academic, New York, 1975).

[4] K. Lindenberg and B. J. West, "Statistical properties of quantum systems: The linear oscillator," Phys. Rev. A **30**, 568 (1984).

[5] E. Braun, "Irreversible behavior of a quantum harmonic oscillator coupled to a heat bath." Physica A **129**, 262 (1985).

[6] E. Braun and P. A. Mello, "The correlation function for a quantum oscillator in a low-temperature heat bath," Physica A **143**, 547 (1987)..

[7] A. Arai, "Spectral analysis of a quantum harmonic oscillator coupled to infinitely many scalar bosons," J. Math. Anal. Appl. **139** (1989).

[8] F. Schwabl and W. Thirring, "Quantum theory of laser radiation," Ergeb. Exact. Naturwiss. **36**, 219 (1964).

[9] A. Arai, "On a model of a harmonic oscillator coupled to a quantized, massless, scalar field. I," J. Math. Phys. **22**, 2539 (1981).

[10] A. Arai, "On a model of a harmonic oscillator coupled to a quantized, massless, scalar field. II," J. Math. Phys. **22**, 2549 (1981).

[11] H. Maassen, "A quantum field acting as a heat bath," Phys. Lett. A **91**, 107 (1982).

[12] C. Aslangul, N. Pittier, and D. Saint-James, "Time behavior of the correlation functions in a simple dissipative model," J. Stat. Phys. **40**, 167 (1985).

[13] P. Schramm and H. Grabert, "Low-temperature and long-time anomalies of a damped quantum particle," J. Stat. Phys. **49**, 767 (1987).

[14]A. Arai, "Rigorous theory of spectra and radiation for a model in quantum electrodynamics," J. Math. Phys. **24**, 1896 (1983).
[15]E. B. Davies, *Quantum Theory of Open Systems* (Academic, London, 1976).
[16]H. Grabert, U. Weiss, and P. Talkner, "Quantum theory of the damped harmonic oscillator," Z. Phys. B **55**, 87 (1984).
[17]R. Jung, G. L. Ingold, and H. Grabert, "Long-time tails in quantum Brownian motion," Phys. Rev. A **32**, 2510 (1985).
[18]E. C. Titchmarsh, *Theory of Fourier Integrals*, 2nd ed. (Oxford, U. P., Oxford, 1948).
[19]E. M. Stein and G. Weiss, *Fourier Analysis on Euclidean Spaces* (Princeton U.P., Princeton, NJ, 1971).

Classical Yang–Baxter equations and quantum integrable systems

Branislav Jurčo
Joint Laboratory of Optics of Czechoslovak Academy of Sciences and Palacký University, Gottwaldova 15, CS-77146 Olomouc, Czechoslovakia

(Received 23 August 1988; accepted for publication 4 January 1989)

Quantum integrable models associated with nondegenerate solutions of classical Yang–Baxter equations related to the simple Lie algebras are investigated. These models are diagonalized for rational and trigonometric solutions in the cases of $sl(N)/gl(N)/$, $o(N)$ and $sp(N)$ algebras. The analogy with the quantum inverse scattering method is demonstrated.

I. INTRODUCTION

As is well known, the classical and quantum Yang–Baxter equations (CYBE and QYBE, respectively) play a central role in the theory of classical and quantum integrable systems.[1,2] In the case of CYBE a classification of their nondegenerate solutions related to the simple Lie algebras was given.[3] In this paper we investigate quantum integrable systems associated with these solutions. They are generalizations of Gaudin's models.[4,5] Given a finite-dimensional Lie algebra g and its representation $T = \otimes\, T_j$, Gaudin looked for sets of mutually commuting operators H_j—special elements of corresponding representation of enveloping algebra (Ref. 5, Sec. 13.2.2),

$$H_j = \sum_{\substack{l=1 \\ l\neq j}}^{M} \sum_{a} w_a^{jl} X_a^j X_a^l, \tag{1.1}$$

where X_a^j represents generator X_a in T_j and then simultaneously diagonalized these operators in the case of su (2) when w_a^{jl} were trigonometric and rational functions of some coupling constants (Sec. IV) and for $u(n)$ in the rational case. We show that for such systems the classical r matrix plays an analogous role as the quantum R matrix in quantum integrable systems associated with QYBE (see Ref. 2). We also diagonalize some models of interest [e.g., for trigonometric solutions of CYBE for generalized Toda systems of types $A_n^{(1)}$, $B_n^{(1)}$, $C_n^{(1)}$, and $D_n^{(1)}$ (Ref. 6)].

The crucial inspirations of this paper were Faddeev's footnote,[7] that Hamiltonians of type (1.1) solve CYBE as well as Gaudin's work on the su (2) case[4,5] and Refs. 8 and 9 dealing with group-invariant generalizations of Heisenberg spin chains.

II. CLASSICAL YANG–BAXTER EQUATIONS

CYBE is a functional equation[1]

$$[r_{01}(\lambda), r_{0'1}(\mu)] + [r_{00'}(\lambda-\mu), r_{01}(\lambda) + r_{0'1}(\mu)] = 0, \tag{2.1}$$

where $r(\lambda)$ is a $g\otimes g$-valued function of complex parameter λ, g being a finite-dimensional Lie algebra and $r_{00'}(\lambda) = r(\lambda)\otimes I$, $r_{0'1}(\lambda) = I\otimes r(\lambda)$, etc. If $r(\lambda)$ is a solution of (2.1) then $(\rho_i\otimes\rho_j)\, r(\lambda)$, where (ρ_i, V_i), (ρ_j, V_j) are irreducible representations of g that also solves (2.1). Classification of nondegenerate solutions for simple Lie algebras was given in Ref. 3. Such a solution is a meromophic function that has a pole of first order at $\lambda = 0$ with residue

$$\operatorname*{res}_{\lambda=0} r(\lambda) = K^{ab} X_a \otimes X_b, \tag{2.2}$$

where K^{ab} is the matrix inverse to the Killing matrix K_{ab}. In the following we shall use two properties of nondegenerate solutions $r(\lambda)$[3]:

(1) $r(\lambda)$ fulfills the so-called unitarity condition

$$r_{01}(-\lambda) = -r_{10}(\lambda); \tag{2.3}$$

(2) if $\gamma\in\Gamma$ [Γ is the set of discrete poles of $r(\lambda)$], then

$$\begin{aligned} &r(\lambda+\gamma) = (A_\gamma\otimes I) r(\lambda) = (I\otimes A^{-1}) r(\lambda), \\ &(A_\gamma\otimes A_\gamma) r(\lambda) = r(\lambda), \end{aligned} \tag{2.4}$$

where A_γ is an automorphism of g (for its form see Ref. 3).

Well-known nondegenerate solutions are, for instance, the rational solution

$$r(\lambda) = (1/\lambda) K^{ab} X_a \otimes X_b \tag{2.5}$$

and the trigonometric solution

$$\begin{aligned} r(\lambda) = \operatorname{cotan}\lambda \Sigma H_a \times H_a + \frac{e^{i\lambda}}{\sin\lambda} \sum_{\alpha>0} E_\alpha \otimes E_{-\alpha} \\ + \frac{e^{-i\lambda}}{\sin\lambda} \sum_{\alpha>0} E_{-\alpha}\otimes E_\alpha, \end{aligned} \tag{2.6}$$

where H_a, E_α is the Cartan–Weyl basis for g.[2,6] Solution (2.6) is equivalent (in sense of Ref. 3) to

$$\begin{aligned} r(\lambda) = \operatorname{cotan}\lambda \sum H_a\otimes H_a \\ + \sum_\alpha \frac{\exp[i\lambda(1-(2/n)p(\alpha))]}{\sin\lambda} E_\alpha\otimes E_{-\alpha}, \end{aligned} \tag{2.7}$$

where $p(\alpha)$ is the height of the root α (mod n) with n being the Coxeter number.

III. QUANTUM INTEGRABLE SYSTEMS

Now consider a finite chain of M sites. The space of quantum states at each site $i = 1,\ldots, M$ will be the representation space V_i for representation (ρ_i, V_i) of g. Then the space $\mathscr{H} = V_1\otimes\cdots V_M$ is the quantum state space for the system. With the notation

$$\begin{aligned} &r_{ij}(\lambda) = (\rho_i\otimes\rho_j) r(\lambda), \quad i,j = 0, 0', 1, \ldots, M, \\ &L(\lambda) = \sum_{i=1}^{M} r_{0i}(\lambda-\epsilon_i), \quad r(\lambda) = r_{00'}(\lambda), \end{aligned} \tag{3.1}$$

where ϵ_i are arbitrary constants and the indices 0 and 0′ refer to the fundamental vector representations, we can—as a consequence of (2.1)—write

 0022-2488/89/061289-05$02.50

$$[L(\lambda)\otimes I, I\otimes L(\mu)]$$
$$+[r(\lambda-\mu), L(\lambda)\otimes I + I\otimes L(\mu)] = 0. \tag{3.2}$$

In more compact obvious notation,

$$[L_a(\lambda), L_b(\mu)] + [r_{ab}(\lambda-\mu), L_a(\lambda) + L_b(\mu)] = 0. \tag{3.3}$$

Denoting $T(\lambda) = \frac{1}{2}\,\mathrm{Tr}_0\, L^2(\lambda)$, where Tr_0 is the trace taken only over the first factor in $g\otimes g$, we have the following lemma:

Lemma:

$$[T(\lambda), T(\mu)] = 0. \tag{3.4}$$

Proof: The Laurent expansion of $T(\lambda)$ in $\lambda = \epsilon_i + \gamma$, $\gamma\in\Gamma$ is using (2.2), (2.4) and the invariance of Killing form

$$T(\lambda) = \frac{1}{2}\,\frac{C_2^i}{(\lambda-\epsilon_i-\gamma)^2} + \frac{H_i}{\lambda-\epsilon_i-\gamma} + \cdots, \tag{3.5}$$

where $H_i = \Sigma_{j\neq i}\, r_{ij}(\epsilon_i - \epsilon_j)$ and C_2^i is the second-order Casimir operator $(K^{ab} X^i_a X^i_b)$ in representation (ρ_i, V_i). Further, for $i\neq j$,

$$[H_i, H_j] = \left[\sum_{k\neq j} r_{jk}(\epsilon_j-\epsilon_k), \sum_{l\neq i} r_{il}(\epsilon_i-\epsilon_l)\right]$$
$$= \sum_k \{[r_{jk}(\epsilon_j-\epsilon_k), r_{ik}(\epsilon_i-\epsilon_k)]$$
$$+ [r_{jk}(\epsilon_j-\epsilon_k), r_{ij}(\epsilon_i-\epsilon_j)]$$
$$+ [r_{ji}(\epsilon_j-\epsilon_i), r_{ik}(\epsilon_i-\epsilon_k)] = 0, \tag{3.6}$$

where we used CYBE and unitarity of $r(\lambda)$, (2.3). So the function $[T(\lambda), T(\mu)]$ has no poles and according to the Liouville lemma it is identically zero.

The following relation can be proved using only (3.2) and the trace properties[10]

$$[T(\lambda), L(\mu)] + [\mathrm{Tr}_0(r(\lambda-\mu)(L(\lambda)\otimes I))$$
$$+ \tfrac{1}{2}\,\mathrm{Tr}_0\, r^2(\lambda-\mu), L(\mu)] = 0. \tag{3.7}$$

Note that computing the residue of (3.7) at $\lambda = \epsilon_i$ yields with (3.5) quantum equations of motion $i(dL(\mu)/dt) = [H_i, L(\mu)]$ in the Lax form

$$i\frac{dL(\mu)}{dt} = \left[\operatorname*{res}_{\lambda=\epsilon_i}(\mathrm{Tr}_0(r(\lambda-\mu)(L(\lambda)\otimes I)) + \frac{1}{2}\,\mathrm{Tr}_0\, r^2(\lambda-\mu)), L(\mu)\right] = 0. \tag{3.8}$$

IV. PREPARATION OF DIAGONALIZATION OF H_i

Let $e_{\alpha\beta}$ be the basis of $\mathrm{gl}(N)$, $(e_{\alpha\beta})_{ij} = \delta_{\alpha i}\,\delta_{\beta j}$, $\alpha_l, \beta_l, i, j = 1, \ldots, N$. We embed $g = A_n, B_n, C_n, D_n$ into $\mathrm{gl}(N)$ with $N = n+1, 2n+1, 2n, 2n$, respectively. Here $L(\lambda)$ determined by (3.1), (2.5), and (2.6) (with an appropriate normalization) can be written in one formula,

$$L(\lambda) = \sum_i \Big\{ f(\lambda-\epsilon_i)\sum_\alpha e_{\alpha\alpha}\otimes X^i_{\alpha\alpha} + g(\lambda-\epsilon_i) \times \sum_{\alpha<\beta} e_{\alpha\beta}\otimes X^i_{\beta\alpha} + \bar g(\lambda-\epsilon_i)\sum_{\alpha>\beta} e_{\alpha\beta}\otimes X^i_{\beta\alpha}\Big\}, \tag{4.1}$$

where $f(\lambda) = g(\lambda) = 1/\lambda$ for the rational case, and $f(\lambda) = \mathrm{cotan}\,\lambda$, $g(\lambda) = e^{i\lambda}/\sin\lambda$ for the trigonometric case. The elements $X_{\alpha\beta}$ of g satisfy the commutation relations:

$$(1)\,[X_{\alpha\beta}, X_{\gamma\delta}] = \delta_{\beta\gamma}X_{\alpha\delta} - \delta_{\alpha\delta}X_{\gamma\beta}; \tag{4.2}$$

for $g = A_n$;

(2) for $g = B_n$, C_n, D_n,

$$[X_{\alpha\beta}, X_{\gamma\delta}] = \delta_{\beta\gamma}X_{\alpha\delta} - \delta_{\alpha\delta}X_{\gamma\beta} + \epsilon_\alpha\epsilon_\beta\delta_{\delta'\beta}X_{\gamma\alpha'} - \epsilon_\alpha\epsilon_\beta\delta_{\alpha'\gamma}X_{\beta'\delta},$$
$$X_{\alpha\beta} = -\epsilon_\alpha\epsilon_\beta X_{\beta'\alpha'}, \quad \alpha' = N+1-\alpha, \tag{4.3}$$

where $\epsilon_\alpha = 1$ for B_n and D_n; $\epsilon_\alpha = 1 (1\leqslant\alpha\leqslant n)$, $\epsilon_\alpha = -1 (n+1\leqslant\alpha\leqslant 2n)$ for C_n. In the case of A_n, we have to assume the representations (ρ_i, V_i) of $\mathrm{sl}(N)$ appearing in (4.1) and (4.2) are obtained as reductions of representations $(\tilde\rho_i, \tilde V_i)$ of $\mathrm{gl}(N)$ with generators $X^i_{\alpha\beta}$; explicitly,

$$X_{\alpha\alpha} = \tilde X_{\alpha\alpha} - (1/N)\Sigma\tilde X_{\beta\beta},$$
$$X_{\alpha\beta} = \tilde X_{\alpha\beta}, \quad \alpha\neq\beta. \tag{4.4}$$

Since the commutation relations for these two sets of generators are identical, we can always take $\mathrm{gl}(N)$ instead of $\mathrm{sl}(N)$; if it were of interest to pass to the $\mathrm{sl}(N)$; one would only pass from the weights m^i to $m^i - (1/N)\Sigma m^j$ [in formulas (5.11) and (5.12) below].

The corresponding r matrices can be written similarly as (4.1)[6]: for $\mathrm{gl}(N)$,

$$r(\lambda) = f(\lambda)\Sigma e_{\alpha\alpha}\otimes e_{\alpha\alpha} + g(\lambda)\sum_{\alpha<\beta} e_{\alpha\beta} \otimes e_{\beta\alpha} + \bar g(\lambda)\sum_{\alpha>\beta} e_{\alpha\beta}\otimes e_{\beta\alpha} \tag{4.5}$$

and for $\mathrm{o}(N)$ and $\mathrm{sp}(N)$

$$r(\lambda) = f(\lambda)\sum e_{\alpha\alpha}\otimes(e_{\alpha\alpha} - e_{\alpha'\alpha'}) + g(\lambda)\sum_{\alpha<\beta} e_{\alpha\beta} \otimes (e_{\beta\alpha} - \epsilon_\alpha\epsilon_\beta e_{\alpha'\beta'}) + \bar g(\lambda)\sum_{\alpha>\beta} e_{\alpha\beta} \otimes (e_{\beta\alpha} - \epsilon_\alpha\epsilon_\beta e_{\alpha'\beta'}). \tag{4.6}$$

Note that it is not hard to see, using explicit forms (4.1), (4.5), and (4.6) and the Liouville lemma, that

$$[\mathrm{Tr}_0\, r^2(\lambda-\mu), L(\mu)] = 0, \tag{4.7}$$

which simplifies (3.7) in the considered cases. For all cases, the Hamiltonians have the form

$$H_i = \sum_{j\neq i}\Big\{ f(\epsilon_i-\epsilon_j)\sum X^i_{\alpha\alpha}X^j_{\alpha\alpha} + g(\epsilon_i-\epsilon_j)\sum_{\alpha<\beta} X^i_{\alpha\beta}X^j_{\beta\alpha} + \bar g(\epsilon_i-\epsilon_j) \times \sum_{\alpha>\beta} X^i_{\alpha\beta}X^i_{\beta\alpha}\Big\}. \tag{4.8}$$

It is now clear from the above formulas, that it does not matter that $\mathrm{o}(4)$ is not simple.

V. DIAGONALIZATION OF H_i IN THE CASE $g=\mathrm{gl}(N)(\mathrm{sl}(N))$

In analogy with Ref. 8 we write the $N\times N$ matrix L in block form as

$$L(\lambda)=\begin{pmatrix}A(\lambda) & B(\lambda)\\ C(\lambda) & D(\lambda)\end{pmatrix}, \tag{5.1}$$

where $B(\lambda)$ is a row with $N-1$ components $(B^1,\ldots,B^{N-1})$, $C(\lambda)$ is a column, and $D(\lambda)$ is an $(N-1)\times(N-1)$ matrix. The $N^2\times N^2$ r matrix (4.5) takes in this basis the form

$$r(\lambda)=\begin{pmatrix}f(\lambda) & 0 & 0 & 0\\ 0 & 0 & g(\lambda)I_{N-1} & 0\\ 0 & \bar{g}(\lambda)I_{N-1} & 0 & 0\\ 0 & 0 & 0 & s(\lambda)\end{pmatrix}, \tag{5.2}$$

where $s(\lambda)$ is an $(N-1)\times(N-1)$ matrix and has the same structure as $r(\lambda)$. The following commutation relations [derived from (3.2) in the block form] will be useful:

$$[A(\lambda),B(\mu)]=-f(\lambda-\mu)B(\mu)+\bar{g}(\lambda-\mu)B(\lambda), \tag{5.3}$$

$$[D_a(\lambda),B_b(\mu)]=-\bar{g}(\lambda-\mu)B_a(\lambda)+B_b(\mu)s_{ab}(\lambda-\mu). \tag{5.4}$$

From (3.7)

$$\begin{aligned}[T(\lambda),B(\mu)]=&\bar{g}(\lambda-\mu)A(\mu)B(\lambda)\\ &+B(\mu)\mathrm{Tr}_0[s(\lambda-\mu)D(\lambda)\otimes I)\\ &-f(\lambda-\mu)A(\lambda)B(\mu)\\ &-\bar{g}(\lambda-\mu)B(\lambda)D(\mu).\end{aligned} \tag{5.5}$$

Now let the spaces $V_1,\ldots,V_M$ be the representation spaces of $\mathrm{gl}(N)$, corresponding to the highest weights $(m_i^1,\ldots,m_i^N)$. Let $\mathscr{H}^0\subset\mathscr{H}$ denote the subspace of vectors $v\in\mathscr{H}$ such that

$$X_{11}^i v=m_i^1 v,\quad X_{\alpha 1}^i v=0,\quad \alpha=2,\ldots,N, \tag{5.6}$$

i.e., $\mathscr{H}^0=V_1{}^0\otimes\cdots V_M^0$, where $V_i^0\subset V_i$ corresponds to the embedding of $\mathrm{gl}(N-1)\subset\mathrm{gl}(N)$ and is an irreducible $\mathrm{gl}(N-1)$ representation space of the highest weight $(m_i^2,\ldots,m_i^N)$. We construct (in analogy with Ref. 8) vectors

$$F=B^{i_1}(\lambda_1^{(1)})\cdots B^{i_{n_1}}(\lambda_{n_1}^{(1)})F_{i_1\cdots i_{n_1}}^{(1)}. \tag{5.7}$$

We use the summation convention for pairs of upper and lower indices and introduce the space $\tilde{\mathscr{H}}=\mathscr{H}^0\otimes V_{M+1}^0\otimes\cdots V_{M+n_1}^0$, where V_{M+i}^0 are the representation spaces C^{N-1} for the fundamental representation of gl $(N-1)$. Here, $F^{(1)}\in\tilde{\mathscr{H}}$ will be specified later. Further we denote

$$\tilde{H}_{\tilde{\epsilon}_i}=\tfrac{1}{2}\operatorname*{res}_{\lambda=\tilde{\epsilon}_i}\mathrm{Tr}_0\tilde{L}^2(\lambda)=\operatorname*{res}_{\lambda=\tilde{\epsilon}_i}\tilde{T}(\lambda), \tag{5.8}$$

where $\{\tilde{\epsilon}_j\}_{j=1}^{M+n_1}=\{\epsilon_1,\ldots,\epsilon_M,\lambda_1^{(1)},\ldots,\lambda_{n_1}^{(1)}\}$ and $\tilde{L}(\lambda)$ is defined as in (3.1) for an extended chain of $(M+n_1)$ sites with $\mathrm{gl}(N-1)$ spins with representation spaces $V_i=V_i^0$, $i=1,\ldots,M+n_1$. The aim is now to commute the $T(\lambda)$ over all B^i's in F using the commutation relation (5.5) and then using (5.3) and (5.4) to commute the occurring A's and D's over the remaining B^i's to the right. Using (5.6), sampling together the "wanted" terms [containing the B^i's with the same arguments as in (5.7)] and the "unwanted" terms (there occur for every k such terms in which the argument $\lambda_k^{(1)}$ in B_{i_k} is changed by λ), demanding that the "unwanted" terms cancel and assuming that according to (3.5) $H_i=\operatorname*{res}_{\lambda=\epsilon_i}T(\lambda)$, we can prove the following theorem [for analogy with GL(N)-invariant generalization of Heisenberg magnetic chain see Ref. 8].

Theorem 1: Vector (5.7) is an eigenvector of H_i with eigenvalue

$$h_i=-\sum_k f(\epsilon_i-\lambda_k^{(1)})m_i^1+\sum_{j\neq i}f(\epsilon_i-\epsilon_j)m_i^1m_j^1+\tilde{h}_{\epsilon_i}, \tag{5.9}$$

if $F^{(1)}\in\tilde{\mathscr{H}}$ is an eigenvector of $\tilde{\mathscr{H}}_{\epsilon_i}$ with eigenvalue $\tilde{h}_{\epsilon_i}$ and the set $\{\lambda_k^{(1)}\}_{k=1}^{n_1}$ satisfies the equations

$$\sum_{l\neq k}f(\lambda_k^{(1)}-\lambda_l^{(1)})-\sum_j f(\lambda_k^{(1)}-\epsilon_j)m_j^1+\tilde{h}_{\lambda_k^{(1)}}=0. \tag{5.10}$$

Then, using embeddings $\mathrm{gl}(N)\supset\mathrm{gl}(N-1)\cdots\supset\mathrm{gl}(2)$ and repeating the recursive procedure following from Theorem 1, we have for the eigenvalues of H_i,

$$\begin{aligned}h_i=&-\sum_k f(\epsilon_i-\lambda_k^{(1)})m_i^1+\sum_{j\neq i}f(\epsilon_i-\epsilon_j)m_i^1m_j^1\\ &+\sum_{S=2}^{N-1}\Big(\sum_{j\neq i}f(\epsilon_i-\epsilon_j)m_i^s m_j^s+\sum_k f(\epsilon_i-\lambda_k^{(s-1)})m_i^s-\sum_k f(\epsilon_i-\lambda_k^{(s)})m_i^s\Big)\\ &+\sum_{j\neq i}f(\epsilon_i-\epsilon_j)m_i^Nm_j^N+\sum_k f(\epsilon_i-\lambda_k^{(N-1)})m_i^N,\end{aligned} \tag{5.11}$$

and for sets $\{\lambda_k^{(1)}\}_{i=1\ k=1}^{N-1\ n_i}$ conditions that guarantee the canceling of "unwanted" terms in each step of this recursive procedure:

$$2\sum_{l\neq k}f(\lambda_k^{(1)}-\lambda_l^{(1)})-\sum_j f(\lambda_k^{(1)}-\epsilon_j)m_j^1+\sum_j f(\lambda_k^{(1)}-\epsilon_j)m_j^2-\sum_l f(\lambda_k^{(1)}-\lambda_l^{(2)})=0,$$

$$2\sum_{l\neq k}f(\lambda_k^{(s)}-\lambda_l^{(s)})-\sum_j(\lambda_k^{(s)}-\epsilon_j)m_j^s-\sum_l f(\lambda_k^{(s)}-\lambda_l^{(s-1)})+\sum_j f(\lambda_k^{(s)}-\epsilon_j)\times m_j^{s+1}-\sum_l f(\lambda_k^{(s)}-\lambda_l^{(s+1)})=0,$$

$$2\sum_{l\neq k} f(\lambda_k^{(N-1)}-\lambda_l^{(N-1)})-\sum_j f(\lambda_k^{(N-1)}-\epsilon_j)m_j^{N-1}$$
$$-\sum_l f(\lambda_k^{(N-1)}-\lambda_l^{(N-2)})$$
$$+\sum_j f(\lambda_k^{(N-1)}-\epsilon_j)m_j^N=0. \tag{5.12}$$

These equations are also conditions that the poles of T at $\lambda=\lambda_k^{(i)}$ disappear.

VI. DIAGONALIZATION OF H_i IN THE CASES $g=o(N)$, $sp(N)$

When $g=o(N)$, we write $L(\lambda)$ in the block form

$$L(\lambda)=\begin{pmatrix} A(\lambda) & B(\lambda) & 0 \\ C(\lambda) & D(\lambda) & B^*(\lambda) \\ 0 & C^*(\lambda) & A^*(\lambda)\end{pmatrix}, \tag{6.1}$$

where $B(\lambda)$ is a row with $(N-2)$ components $(B^1,\ldots,B^{N-2})$, $C(\lambda)$ is a column, and $D(\lambda)$ is an $(N-2)\times(N-2)$ matrix, and $A^*(\lambda)=-A(\lambda)$, $B^i(\lambda)=-B^{*N+1-i}(\lambda)$, $C^i(\lambda)=-C^{*N+1-i}(\lambda)$. The useful commutation relations in block form follow from (3.2) and (3.7):

$$[A(\lambda),B(\mu)]=-f(\lambda-\mu)B(\mu)+\bar g(\lambda-\mu)B(\lambda), \tag{6.2}$$

$$[D_a(\lambda),B_b(\mu)]=-\bar g(\lambda-\mu)B_a(\lambda)+B_a^*(\lambda)\tilde g(\lambda-\mu)+B_b(\mu)s_{ab}(\lambda-\mu), \tag{6.3}$$

$$[T(\lambda),B(\mu)]=2\bar g(\lambda-\mu)A(\mu)B(\lambda)+B(\mu)\mathrm{Tr}_0(s(\lambda-\mu)D(\lambda)\otimes I)-2f(\lambda-\mu)A(\lambda)B(\mu)-2\bar g(\lambda-\mu)B(\lambda)D(\mu). \tag{6.4}$$

Here again $s(\lambda)$ is an $(N-2)^2\times(N-2)^2$ matrix with the same structure as $r(\lambda)$ in (4.6), and $\tilde g(\lambda)$ is an $N\times(N-2)^2$ matrix with elements $\tilde g(\lambda)_{i,jk}=\bar g(\lambda)\,\delta_{i1}\,\delta_{jk'}$.

Theorem 1 holds also in this case but with the following obvious modifications: m_i^k; $i=1,\ldots,M$, $k=1,\ldots,[N/2]$ stands now for the highest weights of irreducibile representations of $o(N)$, $\tilde H_{\epsilon_i}$ is now the Hamiltonian associated with chain of $o(N-2)$ spins, and we must multiply the first two terms in rhs of (5.9) with a factor 2 [see (6.4)]. Also the equations for the eigenvalues (5.11) and (5.12) hold with these changes.

In the case $g=sp(N)$ we write (in analogy with Ref. 9) the $L(\lambda)$ matrix in form

$$L(\lambda)=\begin{pmatrix} A(\lambda) & B(\lambda) \\ C(\lambda) & D(\lambda)\end{pmatrix}, \tag{6.5}$$

where A, B, C, D are $n\times n$ $(n=N/2)$ matrices and

$$A_{ij}=D_{j'i'},\quad B_{ij}=-B_{j'i'},\quad C_{ij}=-C_{j'i'},$$
$$i,j=1,\ldots,n,\quad i'=n+1-i.$$

We shall use the commutation relations in block form

$$[A_a(\lambda),B_b(\mu)]=-s_{ab}(\lambda-\mu)B_b(\mu)+B_a(\lambda)\bar u_{ab}(\lambda-\mu)+B_b(\mu)q_{ab}(\lambda-u), \tag{6.6}$$

$$[D_a(\lambda),B_b(\mu)]=-q_{ab}(\lambda-\mu)B_b(\mu)-\bar u_{ab}(\lambda-\mu)B_a(\lambda)+B_b(\mu)s_{ab}(\lambda-\mu), \tag{6.7}$$

$$[T(\lambda),B(\mu)]=B(\mu)\mathrm{Tr}_0(q(\lambda-\mu)(A(\lambda)\otimes I)+s(\lambda-\mu)(D(\lambda)\otimes I))+A(\mu)\mathrm{Tr}_0(\bar u(\lambda-\mu)(B(\lambda)\otimes I))-\mathrm{Tr}_0(s(\lambda-\mu)(A(\lambda)\otimes I)+q(\lambda-\mu)(D(\lambda)\otimes I))B(\mu)-\mathrm{Tr}_0(\bar u(\lambda-\mu)(B(\lambda)\otimes I))D(\mu). \tag{6.8}$$

Here $s(\lambda)$ is an $n^2\times n^2$ matrix with the same structure as the r matrix (4.5) for the gl (n) case, $q(\lambda)$ is an $n^2\times n^2$ matrix,

$$q(\lambda)=-f(\lambda)\sum_{\alpha<1}^{n} e_{\alpha\alpha}\otimes e_{\alpha'\alpha'}-g(\lambda)\sum_{\alpha<\beta}^{n} e_{\alpha\beta}\otimes e_{\alpha'\beta'}-\bar g(\lambda)\sum_{\alpha>\beta}^{n} e_{\alpha\beta}\otimes e_{\alpha'\beta'}, \tag{6.9}$$

and $u(\lambda)$ is an $n^2\times n^2$ matrix $u(\lambda)_{ijkl}=g(\lambda)(\delta_{i+jn+1}\delta_{k+ln+1}-\delta_{ik}\delta_{jl})$. Now, let the spaces V_i $i=1,\ldots,M$ be the representation spaces for sp $(2n)$ with the highest weights $(m_i^1,\ldots,m_i^n)$. We denote $\mathscr{H}^0\subset\mathscr{H}$ as the subspace of quantum space that is invariant under the action of $A_{ij}(\lambda)$ and $D_{ij}(\lambda)$ such that

$$C_{ij}(\lambda)v=0,\quad \forall v\in\mathscr{H}^0.$$

It is not hard to see that $\mathscr{H}^0=\otimes_{i=1}^{M}\tilde V_i$, where $\tilde V_i$ are the representation spaces for gl(n) with the highest weights $(m_i^1,\ldots,m_i^n)$. We denote $\tilde A_{ij}(\lambda)$ resp. $\tilde D_{ij}(\lambda)$ as the reductions of $A_{ij}(\lambda)$ and $D_{ij}(\lambda)$ to $\mathscr{H}_0$. Further we take

$$F=B^{i_1j_1}(\lambda_1)\cdots B^{i_kj_k}(\lambda_k)F^{(1)}_{i_1\cdots i_kj_1\cdots j_k},$$
$$F^{(1)}\in\mathscr{H}^0\otimes_{i=1}^{k}V_i^0\otimes_{i'=1}^{k}V_{i'}^0, \tag{6.10}$$

$$\tilde L(\lambda)=\sum_{i=1}^{k}q_{0i}(\lambda-\lambda_i)+\tilde A(\lambda)-\sum_{i'=1}^{k}s_{0i'}^{T_{i'}}(\lambda-\lambda_i), \tag{6.11}$$

$$\tilde{\tilde L}(\lambda)=\sum_{i=1}^{k}s_{0i}(\lambda-\lambda_i)+\tilde D(\lambda)-\sum_{i'=1}^{k}q_{0i'}^{T_{i'}}(\lambda-\lambda_i), \tag{6.12}$$

where $T_{i'}$ stands for transposition in $V_{i'}^0$ — i'th representation space for fundamental representation for gl(n) and denote

$$\tilde T(\lambda)=\tfrac12\,\mathrm{Tr}_0\,\tilde L^2(\lambda),\quad \tilde H_i=\operatorname*{res}_{\lambda=\epsilon_i}\tilde T(\lambda),$$
$$\tilde H_{\lambda_k}=\operatorname*{res}_{\lambda=\lambda_k}\tilde T(\lambda),$$
$$\tilde{\tilde T}(\lambda)=\tfrac12\,\mathrm{Tr}_0\,\tilde{\tilde L}^2(\lambda),\quad \tilde{\tilde H}_i=\operatorname*{res}_{\lambda=\epsilon_i}\tilde{\tilde T}(\lambda), \tag{6.13}$$
$$\tilde{\tilde H}_{\lambda_k}=\operatorname*{res}_{\lambda=\lambda_k}\tilde{\tilde T}(\lambda).$$

Now using the commutation relations (6.6)–(6.8) and notations introduced above, we can prove the following theorem [for analogy with Sp($2n$)-invariant "magnets" see the Appendix of Ref. 9].

Theorem 2: The state (6.10) is an eigenstate of H_i, (4.7), with eigenvalue h_i iff $F^{(1)}$ is an eigenstate of $\tilde{H}_i$ and $\tilde{\tilde{H}}_i$ with eigenvalues $\tilde{\tilde{h}}_i$ resp. $\tilde{\tilde{h}}_i$, $h_i = \tilde{h}_i + \tilde{\tilde{h}}_i$ and the numbers λ_k satisfy the following conditions:

$$\tilde{h}_{\lambda_k} + \tilde{\tilde{h}}_{\lambda_k} = 0, \quad \forall k, \tag{6.14}$$

where $\tilde{h}_{\lambda_k}$, $\tilde{\tilde{h}}_{\lambda_k}$ are eigenvalues of $\tilde{H}_{\lambda_k}$ resp. $\tilde{\tilde{H}}_{\lambda_k}$.

Now using automorphisms of gl(n),

$$e_{\alpha\beta} \to -e_{\beta'\alpha'}, \quad e_{\alpha\beta} \to -e_{\beta\alpha}, \quad e_{\alpha\beta} \to e_{\alpha'\beta'},$$

we can see that all Hamiltonians in Theorem 2 are of the gl(n) type and can be diagonalized according to Sec. V. We shall not write out explicit results because they are cumbersome and their structure is obvious.

In this paper we are not dealing with such problems as completeness of obtained eigenstates, their norms, or calculation of some correlation functions [for the su(2) case see Refs. 4 and 5]. Further we note that in cases when the (semiclassical) solutions $R(\lambda,\eta)$ of QYBE, such that $r(\lambda) = \lim_{\eta\to 0}(d/d\eta)\, R(\lambda, \eta)$, is known,[6,9,11] all results of interest can be obtained in this limit from the quantum inverse scattering method.[5,12]

We do not give here any example of a system in the described class, because we shall devote forthcoming papers to such examples of physical interest [four-boson interactions, three-boson interactions (also multiple), second harmonics generation, the Dicke model].

ACKNOWLEDGMENTS

It is a pleasure to acknowledge useful discussions with J. Tolar and L. Hlavatý.

[1] L. A. Takhtajan and L. D. Faddeev, *Hamiltonian Approach in the Soliton Theory* (Nauka, Moscow, 1986), in Russian.
[2] P. P. Kulish and E. K. Sklyanin, Lect. Notes Phys. **151**, 61 (1982).
[3] A. A. Belavin and V. G. Drinfel'd, Funct. Anal. Appl. **16**, 159 (1982).
[4] M. Gaudin, J. Phys. (Paris) **37**, 1087 (1976).
[5] M. Gaudin, *La Fonction d'Onde de Bethe* (Commissariat à l'Energie Atomique, Paris, 1983).
[6] M. Jimbo, Commun. Math. Phys. **102**, 537 (1986).
[7] L. D. Faddeev, footnote p. 296 in Russian edition of Ref. 5 (Mir, Moscow, 1987).
[8] P. P. Kulish and N. Yu. Reshetikhin, J. Phys. A **16**, L591 (1983).
[9] N. Yu. Reshetikhin, Teor. Mat. Fiz. **63**, 347 (1985).
[10] N. Ganoulis, Commun. Math. Phys. **109**, 23 (1987).
[11] V. V. Bazhanov, preprint 86-206, Serpukhov, 1986.
[12] P. P. Kulish, Physica D **3**, 246 (1981).

Lanczos potential

Gonzálo Ares de Parga,[a] Oscar Chavoya A., and José L. López Bonilla
Departamento de Física, Escuela Superior de Física y Matemáticas, Instituto Politécnico Nacional, Edif. #6, U. P. "Zacatenco," Mexico 07738, D. F., Mexico

(Received 8 November 1988; accepted for publication 15 February 1989)

The Lanczos spin tensor [C. Lanczos, Rev. Mod. Phys. **34**, 379 (1962)] is obtained for arbitrary space-times type III, *N*, or *o*.

I. INTRODUCTION

In his paper of 1962, Lanczos[1] shows that for any Riemannian four-geometry, there always exists a third-order tensor K_{abc} which functions as a potential or generator of the Weyl tensor $C_{i,jkr}$ in the following way:

$$C_{pqjb} = K_{pqj;b} - K_{pqb;j} + K_{jbp;q} - K_{jbq;p} + \tfrac{1}{2}[g_{pb}(K_{jq} + K_{qj}) - g_{pj}(K_{qb} + K_{bq}) + g_{qj}(K_{pb} + K_{bq}) - g_{qb}(K_{pj} + K_{jp})] \quad (1a)$$

where

$$K_{pb} \equiv K_p{}^c{}_{b;c}. \quad (1b)$$

K_{abc} is called the Lanczos spin tensor and it possesses the following properties:

$$K_{abc} = -K_{bac}, \quad K_z{}^b{}_b = 0,$$
$$K_{abc} + K_{bca} + K_{cab} = 0. \quad (1c)$$

For a given geometry, the construction of K_{ijr} is equivalent to solving Eqs. (1a) with Eqs. (1c) as constraints. This is not simple in general; nevertheless in this paper K_{abc} is constructed satisfying Eqs. (1a) and (1c) for arbitrary type *o*, *N*, or III space-times.

II. WEYL–LANCZOS EQUATIONS

For simplicity, we will use the notation and conventions attributable to Kramer *et al.*[2] with respect to Newman–Penrose formalism[3] (NP).

The symmetries in (1c) imply that K_{abc} has 16 real independent components; that is, eight complex projections on the NP tetrad,

$$\begin{aligned} \Omega_0 &= K_{(1)(4)(4)}, & \Omega_4 &= K_{(1)(4)(1)}, \\ \Omega_1 &= K_{(1)(4)(2)}, & \Omega_5 &= K_{(1)(4)(3)}, \\ \Omega_2 &= K_{(3)(2)(4)}, & \Omega_6 &= K_{(3)(2)(1)}, \\ \Omega_3 &= K_{(3)(2)(2)}, & \Omega_7 &= K_{(3)(2)(3)}, \end{aligned} \quad (2a)$$

where

$$K_{(a)(b)(c)} \equiv K_{pqr} Z_{(a)}{}^p Z_{(b)}{}^q Z_{(c)}{}^r, \quad (2b)$$

with the null tetrad,

$$(Z_{(a)}{}^r) = (m^r, \bar{m}^r, l^r, n^r). \quad (2c)$$

Hence, once (2a) are known, the Lanczos potential is obtained from

$$K_{abc} = T_{abc} + \bar{T}_{abc}, \quad (3a)$$

with a bar meaning complex conjugation such that

$$\begin{aligned} T_{abc} = {} & \Omega_0 U_{ab} l_c + \Omega_1 (M_{ab} l_c - U_{ab} m_c) \\ & + \Omega_2 (V_{ab} l_c - M_{ab} m_c) - \Omega_3 V_{ab} m_c \\ & - \Omega_4 U_{ab} \bar{m}_c + \Omega_5 (U_{ab} n_c - M_{ab} \bar{m}_c) \\ & - \Omega_6 (M_{ab} n_c - V_{ab} \bar{m}_c) + \Omega_7 V_{ab} n_c, \end{aligned} \quad (3b)$$

where

$$\begin{aligned} V_{ab} &= n_a m_b - n_b m_a, \\ U_{ab} &= -l_a \bar{m}_b + l_b \bar{m}_a, \\ M_{ab} &= m_a \bar{m}_b - m_b \bar{m}_a - n_a l_b + n_b l_a. \end{aligned} \quad (3c)$$

Therefore, projecting (1a) on (2c), we obtain the following relations.

Weyl–Lanczos equations:

$$\begin{aligned} \psi_0 = {} & 2[-\delta\Omega_0 + D\Omega_4 + (\bar{\alpha} + 3\beta - \bar{\pi})\Omega_0 - 3\sigma\Omega_1 \\ & + (-3\epsilon + \bar{\epsilon} - \bar{\rho})\Omega_4 + 3\chi\Omega_5], \\ 2\psi_1 = {} & -\Delta\Omega_0 - 3\delta\Omega_1 + \bar{\delta}\Omega_4 + 3D\Omega_5 \\ & + (3\gamma + \bar{\gamma} + 3\mu - \bar{\mu})\Omega_0 + 3(\bar{\alpha} + \beta - \bar{\pi} - \tau)\Omega_1 \\ & - 6\sigma\Omega_2 + (-3\alpha + \bar{\beta} - 3\pi - \bar{\tau})\Omega_4 \\ & + 3(-\epsilon + \bar{\epsilon} + \rho - \bar{\rho})\Omega_5 + 6\chi\Omega_6, \\ \psi_2 = {} & -\Delta\Omega_1 - \delta\Omega_2 + \bar{\delta}\Omega_5 + D\Omega_6 + \nu\Omega_0 \\ & + (2\mu - \bar{\mu} + \gamma + \bar{\gamma})\Omega_1 + (\bar{\alpha} - \beta - \bar{\pi} - 2\tau)\Omega_2 \\ & - \alpha\Omega_3 - \lambda\Omega_4 + (-\alpha + \bar{\beta} - 2\pi - \bar{\tau})\Omega_5 \\ & + (\epsilon + \bar{\epsilon} - \bar{\rho} + 2\rho)\Omega_6 + \chi\Omega_7, \\ 2\psi_3 = {} & -3\Delta\Omega_2 - \delta\Omega_3 + 3\bar{\delta}\Omega_6 + D\Omega_7 \\ & + 3(-\bar{\mu} + \mu + \bar{\gamma} - \gamma)\Omega_2 + 6\nu\Omega_1 \\ & + (\bar{\alpha} - 3\beta - 3\tau - \bar{\pi})\Omega_3 \\ & - 6\lambda\Omega_5 + 3(\alpha + \bar{\beta} - \bar{\tau} - \pi)\Omega_6 \\ & + (3\epsilon + \bar{\epsilon} - \bar{\rho} + 3\rho)\Omega_7, \\ \psi_4 = {} & 2[-\Delta\Omega_3 + \bar{\delta}\Omega_7 + 3\nu\Omega_2 + (-\bar{\mu} - 3\gamma + \bar{\gamma})\Omega_3 \\ & - 3\lambda\Omega_6 + (3\alpha + \bar{\beta} - \bar{\tau})\Omega_7], \end{aligned} \quad (4)$$

where the Ω_r, $r = 0,...,7$ are related to the NP quantities ψ_a, $a = 0,...,4$ and with the 12 spin coefficients $\chi, \sigma, \rho, \ldots$.

It is difficult to solve Eqs. (4) directly, but comparison of (4) with the NP[2] equations suggests a solution for arbitrary, *N*, III, or *o*, Petrov type.

(a) type o or N: We know that (2c) could be chosen such that $\psi_a = 0$, $a \neq 4$: $\psi_4 \neq 0$ if type *N* and $\psi_4 = 0$ if type *o*. Therefore a solution of (4) is

[a] Area de Física, CBI, Universidad Autónoma Metropolitana-Azcapotzalco, Av. Sn. Pablo 180, Mexico 02200, D. F., Mexico.

 0022-2488/89/061294-02$02.50

$$\Omega_0 = -\chi/2, \quad \Omega_4 = -\sigma/2,$$
$$\Omega_1 = -\rho/6, \quad \Omega_5 = -\tau/6,$$
$$\Omega_2 = \pi/6, \quad \Omega_6 = \mu/6, \tag{5}$$
$$\Omega_3 = \lambda/2, \quad \Omega_7 = \nu/2.$$

(b) type o or III: In this case, there always exists a null tetrad such that $\psi_a = 0$, $a \neq 3$: $\psi_3 \neq 0$ if type III and $\psi_3 = 0$, if type *o*. By comparison of (4) with the eighteen NP[2] equations, it is simple to obtain the solution

$$\Omega_0 = -\chi, \quad \Omega_4 = -\sigma,$$
$$\Omega_1 = -\rho/3, \quad \Omega_5 = -\tau/3,$$
$$\Omega_2 = \pi/3, \quad \Omega_6 = \mu/3, \tag{6}$$
$$\Omega_3 = \lambda, \quad \Omega_7 = \nu.$$

Note that (5) and (6) differ by a factor of 2. To verify that (5) and (6) are solutions of (4), it is helpful to use the 18 well-known NP[2] equations.

In the case of type *o*, we may select $\Omega_r = 0$, $r = 0,\ldots,7$, which corresponds to the trivial solution; (5) and (6) give us a nontrivial solution of (1a) for this Petrov type.

In literature about Lanczos potential, it is common to find the idea that the construction of K_{abc} involves a great deal of work. Nevertheless in this paper we show that NP allows one to express, in a simple way, the Lanczos generator for arbitrary, type *N*, III, or *o* space-times. We are currently seeking the existence of similar solutions to (5) and (6) for arbitrary space-time types I, II, or *D*. However, to the prèsent, we have not found those solutions.

In the general case (arbitrary Petrov type), we conjecture that the Ω_r will be linear combinations of the spin coefficients (with respect to some canonical tetrad).

[1]C. Lanczos, Rev. Mod. Phys. **34**, 379 (1962).
[2]D. Kramer, H. Stephani, M. MacCallum, and E. Herlt, *Exact Solutions of the Einstein Field Equations* (Cambridge U.P., Cambridge, 1980), Eqs. (7.28)–(7.45).
[3]E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

Zeeman's lemma on Robertson–Walker space-times

J. A. Lester
Department of Mathematics and Computer Science, California State University, 5151 State University Drive, Los Angeles, California 90032

(Received 6 November 1987; accepted for publication 21 December 1988)

Zeeman's lemma states that transformations of Minkowski space-time that preserve timelike causality also preserve lightlike causality. The significance of this lemma is that it makes possible an explicit determination of the "causal automorphisms" of Minkowski space-time without recourse to regularity assumptions: they must be orthochronous Lorentz transformations. Here Zeeman's lemma for Robertson–Walker space-times is proved, a result again tantamount to obtaining the causal automorphisms of these space-times explicitly.

I. INTRODUCTION

In general terms, one event in space-time can cause another whenever some influence (i.e., a particle of some sort) can pass between them. Since the world line of this particle must be a nonspacelike curve, one event can cause another whenever some future-directed nonspacelike curve contains both events in the correct order. If the curve is timelike (i.e., if the particle is material) the causality is said to be timelike and if the curve is null, the causality is said to be lightlike. In the following, we also distinguish between *passive* causality (when the nonspacelike curve is a *geodesic*) and ordinary causality as usually defined (when it need not be a geodesic).

Zeeman's lemma[1] states that transformations of Minkowski space-time that preserve passive timelike causality preserve passive lightlike causality. [Note that although Zeeman's proof implicity uses geodesics instead of more general nonspacelike curves, in Minkowski space-time, there is in fact no distinction between passive and ordinary causality. For other space-times, there *is* generally such a distinction: in anti-de Sitter space-time, for example (as illustrated in Fig. 20 of Ref. 2), the events reachable from a typical event p via timelike geodesics clearly form a proper subset of those reachable via more general timelike curves.] The significance of Zeeman's lemma is that it makes it possible to determine the "causal automorphisms" (i.e., the causality-preserving transformations) of Minkowski space-time explicitly, without assuming them to be in any way regular (linear, continuous, etc.). Since transformations preserving passive lightlike causality trivially preserve pairs of events joined by light signals, the problem reduces to showing that transformations that preserve such pairs of points are essentially Lorentz transformations. (This is the substance of Ref. 1; however, the same conclusion follows from earlier, less well-known theorems of Alexandrov.[3] See also the bibliographies of Refs. 4 and 5 for related theorems.)

In the following, we prove Zeeman's lemma for passive causality in Robertson–Walker space-times. We begin in Sec. II by describing these space-times in terms of certain scalar products and continue in Sec. III by finding their nonspacelike geodesics explicitly. In Sec. IV we derive algebraic conditions for passive causality and in Sec. V we use these conditions to prove the lemma. We note that since the transformations of Robertson–Walker space-times preserving light signals were derived explicitly in Ref. 4 (the results are too lengthy to state here), this proof of Zeeman's lemma is tantamount to a derivation of explicit forms for the passive causal automorphisms of Robertson–Walker space-times.

II. ROBERTSON–WALKER SPACE-TIMES

Robertson-Walker space-times are those with exact spherical spatial symmetry about every point (see Ref. 2, Sec. 5.3 for a more detailed description). For a suitable coordinate system, the metric of any such space takes the local form

$$ds^2 = -d\rho^2 + h(\rho)^2\, d\sigma^2,$$

where ρ is a timelike coordinate; $d\sigma^2$ is a three-dimensional spatial metric of constant zero, negative, or positive curvature; and $h = h(\rho)$ is some appropriate "radius" function.

We require a more formal and global description of Robertson–Walker space-times. For some open interval $\mathbf{I}$ of $\mathbb{R}$ ($\mathbf{I}$ may be infinite or semi-infinite), we assume that the function $h{:}\mathbf{I}\to\mathbb{R}$ is everywhere positive and "nice," i.e., such that all limits, derivatives, and integrals required below exist. As described below, the spatial metric $d\sigma^2$ will be expressed in terms of a suitable scalar product.

Consider first the case of constant zero spatial curvature and let $(\ ,\ )$ be the usual dot product on $\mathbb{R}^3$.[4]

Definition 2.1: The Robertson–Walker space-time of constant zero spatial curvature corresponding to the function $h{:}\mathbf{I}\to\mathbb{R}$ is the set

$$\mathbf{R}^0(h) := \{(\rho,\mathbf{r})\,|\,\rho\in\mathbf{I},\ \mathbf{r}\in\mathbb{R}^3\},$$

together with the metric

$$ds^2 := -d\rho^2 + h(\rho)^2(\mathbf{dr},\mathbf{dr}).$$

For constant *positive* spatial curvature, let $(\ ,\ )$ denote the usual dot product on $\mathbb{R}^4$. We allow $\mathbf{r}$ to range over the sphere $(\mathbf{r},\mathbf{r}) = 1$ in $\mathbb{R}^4$.

Definition 2.2: The Robertson–Walker space-time of constant positive curvature corresponding to the function $h{:}\mathbf{I}\to\mathbb{R}$ is the set

$$\mathbf{R}^+(h) = \{(\rho,\mathbf{r})\,|\,\rho\in\mathbf{I},\ \mathbf{r}\in\mathbb{R}^4, (\mathbf{r},\mathbf{r}) = 1\},$$

together with the metric

$$ds^2 := -d\rho^2 + h(\rho)^2(\mathbf{dr},\mathbf{dr}).$$

[By parametrizing the sphere as

 0022-2488/89/061296-05$02.50

$\mathbf{r} = (\cos\chi, \sin\chi\cos\theta, \sin\chi\sin\theta\cos\phi, \sin\chi\sin\theta\sin\phi)$, we obtain the more familiar form $(\mathbf{dr},\mathbf{dr}) = d\chi^2 + \sin^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)$.]

For constant *negative* spatial curvature, define the indefinite scalar product $(\ ,\ )$ on $\mathbb{R}^4$ by

$$(\mathbf{r}_1,\mathbf{r}_2) := -w_1w_2 + x_1x_2 + y_1y_2 + z_1z_2,$$

for all $\mathbf{r}_1 := (w_1,x_1,y_1,z_1)$ and $\mathbf{r}_2 := (w_2,x_2,y_2,z_2) \in \mathbb{R}^4$. The equation $(\mathbf{r},\mathbf{r}) = -1$ represents a hyperboloid of two sheets in $\mathbb{R}^4$. By fixing the sign of the w coordinate, we let $\mathbf{r}$ range over one sheet of this hyperboloid.

Definition 2.3: The Robertson–Walker space-time of constant negative spatial curvature corresponding to the function $h:\mathbf{I}\to\mathbb{R}$ is the set

$$\mathbf{R}^-(h) := \{(\rho,\mathbf{r}) \mid \rho\in\mathbf{I},\ \mathbf{r}\in\mathbb{R}^4, (\mathbf{r},\mathbf{r}) = -1,\ w>0\},$$

together with the metric

$$ds^2 := -d\rho^2 + h(\rho)^2(\mathbf{dr},\mathbf{dr}).$$

[The more conventional form for $d\sigma^2$ can be obtained by parametrizing the sheet of the hyperboloid as

$$\mathbf{r} = (\cosh\chi, \sinh\chi, \cos\theta, \sinh\chi\sin\theta\cos\phi, \sinh\chi\sin\theta\sin\phi);$$

we then calculate that $(\mathbf{dr},\mathbf{dr}) = d\chi^2 + \sinh^2\chi(d\theta^2 + \sin^2\theta\, d\phi^2)$.]

We note some useful properties of this indefinite scalar product.

(i) For distinct $\mathbf{a}$ and $\mathbf{b}$ in $\mathbb{R}^4$ with positive w coordinates satisfying $(\mathbf{a},\mathbf{a}) = (\mathbf{b},\mathbf{b}) = -1$, we have $(\mathbf{a},\mathbf{b}) < -1$.

(ii) For any $\mathbf{a}$, $\mathbf{c}$, and $\mathbf{d}$ in $\mathbb{R}^4$ with at least one of $(\mathbf{a},\mathbf{a})$, $(\mathbf{c},\mathbf{c})$, or $(\mathbf{d},\mathbf{d})$ negative,

$$\begin{vmatrix} (\mathbf{a},\mathbf{a}) & (\mathbf{a},\mathbf{c}) & (\mathbf{a},\mathbf{d}) \\ (\mathbf{c},\mathbf{a}) & (\mathbf{c},\mathbf{c}) & (\mathbf{c},\mathbf{d}) \\ (\mathbf{d},\mathbf{a}) & (\mathbf{d},\mathbf{c}) & (\mathbf{d},\mathbf{d}) \end{vmatrix} \leqslant 0.$$

Before continuing on to our discussion of geodesics and passive causality, we look at a useful modification of the timelike coordinate ρ. Fix an arbitrary number $\nu\in\mathbf{I}$.

Definition 2.4: For any scalar $c\geqslant 0$ and any $\rho\in\mathbf{I}$, we define

$$\rho_c := \int_\nu^\rho \frac{d\lambda}{h(\lambda)\{1 + c[h(\lambda)]^2\}^{1/2}}.$$

We also define $\mathbf{I}_0 := \{\rho_0 \mid \rho\in\mathbf{I}\}$.

The following theorem lists the properties of ρ_c.

Theorem 2.1: For any $\rho, \alpha, \beta\in\mathbf{I}$ with $\beta > \alpha$ and any $c\geqslant 0$, (i) $\beta_c > \alpha_c$; (ii) if $c\neq 0$, then $\rho_0 > \rho_c$; (iii) if $c\neq 0$, then $\beta_0 - \alpha_0 > \beta_c - \alpha_c$; and (iv) if μ is any scalar with $0 < \mu < \beta_0 - \alpha_0$, then there exists a $c>0$ with $\beta_c - \alpha_c = \mu$.

Proof: Statements (i) and (ii) are obvious from the definition of ρ_c and elementary properties of integrals. For (iii) and (iv), define the function $F_{\alpha\beta}:[0,\infty)\to\mathbb{R}$ by

$$F_{\alpha\beta}(x) := \beta_x - \alpha_x = \int_\alpha^\beta \frac{d\lambda}{h(\lambda)\{1 + x[h(\lambda)]^2\}^{1/2}}.$$

Now $F_{\alpha\beta}$ is a strictly decreasing function of x and $F_{\alpha\beta}(0) = \beta_0 - \alpha_0$, from which statement (iii) follows. Furthermore, $F_{\alpha\beta}$ is continuous (since h is "nice") and $\lim_{x\to\infty} F_{\alpha\beta}(x) = 0$, so for any μ with $0 < \mu < \beta_0 - \alpha_0 = F_{\alpha\beta}(0)$, we have statement (iv): $F_{\alpha\beta}(c) = \mu$ for some $c>0$. ■

III. THE NONSPACELIKE GEODESICS OF ROBERTSON–WALKER SPACE-TIMES

Let $\mathbf{R}$ denote whichever of the Robertson–Walker space-times $\mathbf{R}^0(h)$, $\mathbf{R}^-(h)$, or $\mathbf{R}^+(h)$ is under discussion and fix the scalar ϵ to be 0, -1, or $+1$, respectively. Denote derivatives with respect to arc length s by dots and derivatives with respect to ρ by primes. Set $h := h(\rho)$.

Theorem 3.1: Geodesics in $\mathbf{R}$ satisfy the following equations:

$$\ddot{\rho} + hh'(\dot{\mathbf{r}},\dot{\mathbf{r}}) = 0, \quad [h^2\mathbf{r}]^{\cdot} = -\epsilon h^2(\dot{\mathbf{r}},\dot{\mathbf{r}})\mathbf{r}.$$

Proof: Use the standard variational argument:

$$2\,ds\,\delta(ds) = -2\,d\rho\,\delta(d\rho) + 2hh'\,\delta\rho(\mathbf{dr},\mathbf{dr}) + 2h^2(\mathbf{dr},\delta(\mathbf{dr})),$$

so

$$0 = \int \delta(ds) = \int \dot{\rho}(\delta\rho)^{\cdot}\,ds + \int hh'(\dot{\mathbf{r}},\mathbf{r}')\delta\rho\,ds + \int h^2(\dot{\mathbf{r}},(\delta\mathbf{r})^{\cdot})ds.$$

Integrate the first and third integrals by parts and simplify:

$$0 = \int \{[\ddot{\rho} + hh'(\dot{\mathbf{r}},\dot{\mathbf{r}})]\delta\rho + ([h^2\dot{\mathbf{r}}]^{\cdot},\delta\mathbf{r})\}ds,$$

true for all $\delta\rho\in\mathbb{R}$ and all *allowable* $\delta\mathbf{r}$.

If $\mathbf{R} = \mathbf{R}^0(h)$, there is no restriction on $\mathbf{r}$, so all $\delta\mathbf{r}\in\mathbb{R}^3$ are allowable. Then for $\delta\mathbf{r} = \mathbf{0}$, we obtain $\ddot{\rho} + hh'(\dot{\mathbf{r}},\dot{\mathbf{r}}) = 0$ (the first equation) and for $\delta\rho = 0$, we obtain $([h^2\dot{\mathbf{r}}]^{\cdot},\delta\mathbf{r}) = 0$ for all $\delta\mathbf{r}\in\mathbb{R}^3$, which implies that $[h^2\dot{\mathbf{r}}]^{\cdot} = 0 = \epsilon h^2(\dot{\mathbf{r}},\dot{\mathbf{r}})$, the second equation.

If $\mathbf{R} = \mathbf{R}^{\pm}(h)$, then since $(\mathbf{r},\mathbf{r}) = \epsilon$, only those $\delta\mathbf{r}\in\mathbb{R}^4$ with $(\delta\mathbf{r},\mathbf{r}) = 0$ are allowable. For $\delta\mathbf{r} = \mathbf{0}$, we again obtain the first equation. For $\delta\rho = 0$, we have that $[h^2\dot{\mathbf{r}}]^{\cdot}$ is orthogonal to all that are orthogonal to $\mathbf{r}$, so $[h^2\dot{\mathbf{r}}]^{\cdot} = \lambda\mathbf{r}$ for some scalar λ. Then $\lambda\mathbf{r} = h^2\ddot{\mathbf{r}} + 2h\dot{h}\dot{\mathbf{r}}$, from which $\lambda(\mathbf{r},\mathbf{r}) = (h^2\ddot{\mathbf{r}} + 2h\dot{h}\dot{\mathbf{r}},\mathbf{r})$, i.e., $\epsilon\lambda = h^2(\ddot{\mathbf{r}},\mathbf{r}) + 2h\dot{h}(\dot{\mathbf{r}},\mathbf{r})$. However, $(\mathbf{r},\mathbf{r}) = \epsilon$ implies that $(\dot{\mathbf{r}},\mathbf{r}) = 0$ and $(\dot{\mathbf{r}},\dot{\mathbf{r}}) + (\ddot{\mathbf{r}},\mathbf{r}) = 0$, so $\lambda = -\epsilon h^2(\dot{\mathbf{r}},\dot{\mathbf{r}})$ and $[h^2\dot{\mathbf{r}}]^{\cdot} = -\epsilon h^2(\dot{\mathbf{r}},\dot{\mathbf{r}})\mathbf{r}$. ■

We now use these geodesic equations with appropriate initial conditions to find the nonspacelike geodesics of $\mathbf{R}$. Define δ to be $+1$ or 0 whenever the geodesic under discussion is timelike or null, respectively.

Theorem 3.2: Assume that the following initial conditions for the geodesic equations of Theorem 3.1 hold at $s = 0$:

$$\rho = \alpha, \quad \dot{\rho} = \beta, \quad \mathbf{r} = \mathbf{a}, \quad \dot{\mathbf{r}} = \Psi\mathbf{b},$$

where

$$(\mathbf{b},\mathbf{b}) = 1, \quad \Psi\geqslant 0, \quad -\beta^2 + [h(\alpha)]^2\Psi^2 = -\delta.$$

If $\mathbf{R}\neq\mathbf{R}^0(h)$, then

$$(\mathbf{a},\mathbf{a}) = \epsilon, \quad (\mathbf{a},\mathbf{b}) = 0.$$

Then (taking $\dot{\rho}\geqslant 0$) the nonspacelike geodesics in $\mathbf{R}$ are as follows.

If $\Psi = 0$, then $\mathbf{r} = \mathbf{a}$, $\rho = s + \alpha$, and the geodesic is timelike.

If $\Psi > 0$, then for $c := \delta\Psi^{-2}[h(\alpha)]^{-4} \geqslant 0$, the geodesic takes the form

$$\begin{cases} \text{for } \mathbf{R} = \mathbf{R}^0(h) & \mathbf{r}(\rho) = (\rho_c - \alpha_c)\mathbf{b} + \mathbf{a} \\ \text{for } \mathbf{R} = \mathbf{R}^-(h), & \mathbf{r}(\rho) = \cosh(\rho_c - \alpha_c)\mathbf{a} \\ & \quad + \sinh(\rho_c - \alpha_c)\mathbf{b}, \\ \text{for } \mathbf{R} = \mathbf{R}^+(h), & \mathbf{r}(\rho) = \cos(\rho_c - \alpha_c)\mathbf{a} \\ & \quad + \sin(\rho_c - \alpha_c)\mathbf{b} \end{cases}$$

and is timelike if and only if $c \neq 0$.

Proof: Expand the second geodesic equation and multiply by $2h^2$:

$$2h^4\ddot{\mathbf{r}} + 4h^3\dot{h}\dot{\mathbf{r}} = -2\epsilon h^3(\dot{\mathbf{r}},\dot{\mathbf{r}})\mathbf{r}.$$

Take the scalar product with $\dot{\mathbf{r}}$ and rearrange; then

$$[h^4(\dot{\mathbf{r}},\dot{\mathbf{r}})]^{\cdot} = -2\epsilon h^3(\dot{\mathbf{r}},\dot{\mathbf{r}})(\mathbf{r},\dot{\mathbf{r}}) = 0,$$

since either $\epsilon = 0$ or $(\mathbf{r},\dot{\mathbf{r}}) = 0$. Use the initial conditions to obtain that $(\dot{\mathbf{r}},\dot{\mathbf{r}}) = A^2h^{-4}$ for $A := \Psi[h(\alpha)]^2 \geqslant 0$, from which the two geodesic equations become

$$\ddot{\rho} + A^2h^{-3}h' = 0 \tag{1a}$$

and

$$[h^2\dot{\mathbf{r}}]^{\cdot} = -\epsilon A^2h^{-2}\mathbf{r}. \tag{1b}$$

If $\Psi = 0$, then $A = 0$; thus $\ddot{\rho} = 0$, which implies that $\rho = s + \alpha$, and $(\dot{\mathbf{r}},\dot{\mathbf{r}}) = 0$, which implies that $\ddot{\mathbf{r}} = \mathbf{0}$ and thus $\mathbf{r} = \mathbf{a}$.

Assume then that $\Psi > 0$. Multiply Eq. (1a) by $2\dot{\rho}$ and rearrange:

$$(\dot{\rho}^2)^{\cdot} = A^2(h^{-2})^{\cdot}.$$

Then use the initial conditions to obtain

$$\dot{\rho} = \{A^2h^{-2} + \delta\}^{1/2} = Ah^{-1}\{1 + ch^2\}^{1/2}.$$

Define $\dot{\phi} = \rho_c - \alpha_c$. Now $\rho_c' = h^{-1}\{1 + ch^2\}^{-1/2}$, so $\dot{\phi} = \dot{\rho}_c = \rho_c'\dot{\rho} = Ah^{-2}$. Use the appropriate chain rules to calculate that

$$\frac{d^2\mathbf{r}}{d\phi^2} = \frac{(\dot{\phi}\ddot{\mathbf{r}} - \ddot{\phi}\dot{\mathbf{r}})}{\dot{\phi}^3} = A^{-2}[h^2\dot{\mathbf{r}}]^{\cdot}h^2,$$

so from (1b),

$$\frac{d^2\mathbf{r}}{d\phi^2} = A^{-2}(-\epsilon A^2h^{-2}\mathbf{r})h^2 = -\epsilon\mathbf{r}.$$

The solutions to this equation for the various values of ϵ give the required curves. ■

IV. ALGEBRAIC CONDITIONS FOR PASSIVE CAUSALITY

As in Minkowski space-time, passive causality in $\mathbf{R}$ is defined in terms of nonspacelike geodesics. In the following definition, passive timelike causality is denoted by the symbol $\lessdot$ and passive lightlike causality by the symbol $\lessdot\cdot$.

Definition 4.1: For any two events $(\alpha,\mathbf{a})$ and $(\beta,\mathbf{b})$ in $\mathbf{R}$,

(i) $(\alpha,\mathbf{a}) \lessdot (\beta,\mathbf{b})$ if and only if there exists a timelike geodesic joining $(\alpha,\mathbf{a})$ and $(\beta,\mathbf{b})$ and $\alpha < \beta$.

(ii) $(\alpha,\mathbf{a}) \lessdot\cdot (\beta,\mathbf{b})$ if and only if there exists a null geodesic joining $(\alpha,\mathbf{a})$ and $(\beta,\mathbf{b})$ and $\alpha < \beta$.

The next two theorems give more workable characterizations of passive causality.

Theorem 4.1: For any two distinct points $(\gamma,\mathbf{c})$ and $(\delta,\mathbf{d})$ in $\mathbf{R}$, $(\gamma,\mathbf{c}) \lessdot (\delta,\mathbf{d})$ if and only if $\delta_0 > \gamma_0$ and

$$\begin{cases} \text{for } \mathbf{R} = \mathbf{R}^0(h), & (\delta_0 - \gamma_0)^2 > (\mathbf{d} - \mathbf{c},\mathbf{d} - \mathbf{c}), \\ \text{for } \mathbf{R} = \mathbf{R}^-(h), & \cosh(\delta_0 - \gamma_0) > -(\mathbf{d},\mathbf{c}), \\ \text{for } \mathbf{R} = \mathbf{R}^+(h), & \cos(\delta_0 - \gamma_0) < (\mathbf{d},\mathbf{c}) \\ & \text{or } \delta_0 - \gamma_0 > \pi. \end{cases}$$

Proof: (i) Assume that $\delta_0 > \gamma_0$ and that the appropriate condition holds; then $\delta > \gamma$ [Theorem 2.1 (i)]. If $\mathbf{c} = \mathbf{d}$, then the curve $\mathbf{r} = \mathbf{c}$, $\rho = s + \gamma$ is a timelike geodesic joining the two points, so $(\gamma,\mathbf{c}) < (\delta,\mathbf{d})$. If $\mathbf{c} \neq \mathbf{d}$, then

$$\begin{cases} 0 < (\mathbf{d} - \mathbf{c},\mathbf{d} - \mathbf{c})^{1/2} < \delta_0 - \gamma_0, \\ 0 < \cosh^{-1}\{-(\mathbf{d},\mathbf{c})\} < \delta_0 - \gamma_0, \\ 0 < \cos^{-1}(\mathbf{d},\mathbf{c}) < \delta_0 - \gamma_0. \end{cases}$$

From Theorem 2.1 (iv) there then exists a positive c such that

$$\delta_c - \gamma_c = \begin{cases} (\mathbf{d} - \mathbf{c},\mathbf{d} - \mathbf{c})^{1/2}, \\ \cosh^{-1}\{-(\mathbf{d},\mathbf{c})\}, \\ \cos^{-1}(\mathbf{d},\mathbf{c}). \end{cases}$$

Define

$$\mathbf{b} := \begin{cases} (\delta_c - \gamma_c)^{-1}(\mathbf{d} - \mathbf{c}), \\ -\coth(\delta_c - \gamma_c)\mathbf{c} + \operatorname{csch}(\delta_c - \gamma_c)\mathbf{d}, \\ -\cot(\delta_c - \gamma_c)\mathbf{c} + \csc(\delta_c - \gamma_c)\mathbf{d}. \end{cases}$$

Then it is easily checked that $(\mathbf{b},\mathbf{b}) = 1$, $(\mathbf{b},\mathbf{c}) = 0$ if $\mathbf{R} \neq \mathbf{R}^0(h)$, and

$$\begin{cases} \mathbf{r}(\rho) = (\rho_c - \gamma_c)\mathbf{b} + \mathbf{c}, \\ \mathbf{r}(\rho) = \cosh(\rho_c - \gamma_c)\mathbf{c} + \sinh(\rho_c - \gamma_c)\mathbf{b}, \\ \mathbf{r}(\rho) = \cos(\rho_c - \gamma_c)\mathbf{c} + \sin(\rho_c - \gamma_c)\mathbf{b} \end{cases}$$

is a timelike geodesic joining the two points. Thus $(\gamma,\mathbf{c}) \lessdot (\delta,\mathbf{d})$.

(ii) Assume that $(\gamma,\mathbf{c}) \lessdot (\delta,\mathbf{d})$; then $\delta_0 > \gamma_0$ [Theorem 2.1 (i)] and there exists a timelike geodesic joining the two points. If this geodesic has the form $\mathbf{r} = \mathbf{a}$, $\rho = s + \alpha$, then the appropriate condition holds trivially. Assume otherwise; then the geodesic has the form given above for appropriate $\mathbf{c}$ and $\mathbf{b}$. From $\mathbf{r}(\gamma) = \mathbf{c}$ and $\mathbf{r}(\delta) = \mathbf{d}$ follows

$$\begin{cases} (\mathbf{d} - \mathbf{c},\mathbf{d} - \mathbf{c}) = (\delta_c - \gamma_c)^2, \\ -(\mathbf{d} - \mathbf{c}) = \cosh(\delta_c - \gamma_c), \\ (\mathbf{d},\mathbf{c}) = \cos(\delta_c - \gamma_c), \end{cases}$$

which, since $\delta_c - \gamma_c < \delta_0 - \gamma_0$ [Theorem 2.1(iii)], implies the required relation. ■

Theorem 4.2: For any two points $(\gamma,\mathbf{c})$ and $(\delta,\mathbf{d})$ in $\mathbf{R}$, $(\gamma,\mathbf{c}) \lessdot\cdot (\delta,\mathbf{d})$ if and only if $\delta_0 > \gamma_0$ and

$$\begin{cases} \text{for } \mathbf{R} = \mathbf{R}^0(h), & (\delta_0 - \gamma_0)^2 = (\mathbf{d} - \mathbf{c},\mathbf{d} - \mathbf{c}), \\ \text{for } \mathbf{R} = \mathbf{R}^-(h), & \cosh(\delta_0 - \gamma_0) = -(\mathbf{d},\mathbf{c}), \\ \text{for } \mathbf{R} = \mathbf{R}^+(h), & \cos(\delta_0 - \gamma_0) = (\mathbf{d},\mathbf{c}). \end{cases}$$

Proof: The proof of Theorem 4.2 is analogous to that of Theorem 4.1 with $c = 0$ and will thus be omitted. ■

V. PROOF OF ZEEMAN'S LEMMA

The crux of the proof is to characterize passive lightlike causality ($\prec\cdot$) in terms of passive timelike causality ($\ll\cdot$).

Theorem 5.1: For distinct $(\gamma,\mathbf{c})$ and $(\delta,\mathbf{d})$ in $\mathbf{R}$, $(\gamma,\mathbf{c})\prec\cdot(\delta,\mathbf{d})$ [and $\delta_0-\gamma_0\leqslant\pi$ if $\mathbf{R}=\mathbf{R}^+(h)$] if and only if $(\gamma,\mathbf{c})\not\prec(\delta,\mathbf{d})$ and, for all $(\alpha,\mathbf{a})$ in $\mathbf{R}$, $(\alpha,\mathbf{a})\prec(\gamma,\mathbf{c})$ implies that $(\alpha,\mathbf{a})\prec(\delta,\mathbf{d})$.

Proof for the case $\mathbf{R}=\mathbf{R}^0(h)$: (i) Assume that $(\gamma,\mathbf{c})\prec\cdot(\delta,\mathbf{d})$; then $(\delta_0-\gamma_0)^2=(\mathbf{d}-\mathbf{c},\mathbf{d}-\mathbf{c})$ [so $(\gamma,\mathbf{c})\not\prec(\delta,\mathbf{d})$] and $\delta_0>\gamma_0$. Suppose that $(\alpha,\mathbf{a})\prec(\gamma,\mathbf{c})$ for some $(\alpha,\mathbf{a})$ in $\mathbf{R}$; then $\alpha_0<\gamma_0$ and $(\gamma_0-\alpha_0)^2>(\mathbf{a}-\mathbf{c},\mathbf{a}-\mathbf{c})$. Then

$$(\gamma_0-\alpha_0)^2(\delta_0-\gamma_0)^2>(\mathbf{a}-\mathbf{c},\mathbf{a}-\mathbf{c})(\mathbf{d}-\mathbf{c},\mathbf{d}-\mathbf{c})\geqslant(\mathbf{a}-\mathbf{c},\mathbf{d}-\mathbf{c})^2$$

using the Cauchy–Schwarz inequality. Taking positive square roots, we have

$$(\gamma_0-\alpha_0)(\delta_0-\gamma_0)>|(\mathbf{a}-\mathbf{c},\mathbf{d}-\mathbf{c})|\geqslant(\mathbf{a}-\mathbf{c},\mathbf{d}-\mathbf{c}).$$

Now $\mathbf{d}-\mathbf{a}=(\mathbf{d}-\mathbf{c})-(\mathbf{a}-\mathbf{c})$, so

$$\begin{aligned}(\mathbf{d}-\mathbf{a},\mathbf{d}-\mathbf{a})&=(\mathbf{d}-\mathbf{c},\mathbf{d}-\mathbf{c})-2(\mathbf{d}-\mathbf{c},\mathbf{a}-\mathbf{c})+(\mathbf{a}-\mathbf{c},\mathbf{a}-\mathbf{c})\\&<(\delta_0-\gamma_0)^2-2(\gamma_0-\alpha_0)(\delta_0-\gamma_0)+(\gamma_0-\alpha_0)^2\\&=(\delta_0-\alpha_0)^2.\end{aligned}$$

Thus since $\delta_0>\gamma_0>\alpha_0$, we have that $(\alpha,\mathbf{a})\prec(\delta,\mathbf{d})$.

(ii) Assume that $(\gamma,\mathbf{c})\not\prec(\delta,\mathbf{d})$ and that for all events $(\alpha,\mathbf{a})$ in $\mathbf{R}$, $(\alpha,\mathbf{a})\prec(\gamma,\mathbf{c})$ implies $(\alpha,\mathbf{a})\prec(\delta,\mathbf{d})$. We also assume that $(\gamma,\mathbf{c})\prec\cdot(\delta,\mathbf{d})$ and derive a contradiction. The two subcases $\gamma_0\leqslant\delta_0$ and $\gamma_0>\delta_0$ will be treated separately.

Suppose that $\gamma_0\leqslant\delta_0$. Then [since $(\gamma,\mathbf{c})\neq(\delta,\mathbf{d})$, $(\gamma,\mathbf{c})\not\prec(\delta,\mathbf{d})$, and $(\gamma,\mathbf{c})\prec\cdot(\delta,\mathbf{d})$], $(\delta_0-\gamma_0)^2<(\mathbf{c}-\mathbf{d},\mathbf{c}-\mathbf{d})$, so $\delta_0-(\mathbf{c}-\mathbf{d},\mathbf{c}-\mathbf{d})^{1/2}<\gamma_0$. Choose a number λ in $\mathbf{I}_0$ such that $\delta_0-(\mathbf{c}-\mathbf{d},\mathbf{c}-\mathbf{d})^{1/2}<\lambda<\gamma_0$; then there exists an $\alpha\in\mathbf{I}$ with $\alpha_0=\lambda$. Define $\mathbf{a}:=\mathbf{c}$; then $(\gamma_0-\alpha_0)^2>0=(\mathbf{c}-\mathbf{a},\mathbf{c}-\mathbf{a})$, so $(\alpha,\mathbf{a})\prec(\gamma,\mathbf{c})$. However, $0\leqslant\delta_0-\gamma_0<\delta_0-\alpha_0<(\mathbf{c}-\mathbf{d},\mathbf{c}-\mathbf{d})^{1/2}$, from which $(\delta_0-\alpha_0)^2<(\mathbf{a}-\mathbf{d},\mathbf{a}-\mathbf{d})$. This implies the contradiction $(\alpha,\mathbf{a})\not\prec(\delta,\mathbf{d})$.

Suppose that $\gamma_0>\delta_0$. Choose any λ in $\mathbf{I}_0$ with $\gamma_0>\lambda>\delta_0$; then $\lambda=\alpha_0$ for some α in $\mathbf{I}$. Define $\mathbf{a}:=\mathbf{c}$; then since $\alpha_0<\gamma_0$ and $(\gamma_0-\alpha_0)^2>0=(\mathbf{a}-\mathbf{c},\mathbf{a}-\mathbf{c})$, we obtain $(\alpha,\mathbf{a})\prec(\gamma,\mathbf{c})$. However, $\alpha_0>\delta_0$, so we again have the contradiction that $(\alpha,\mathbf{a})\not\prec(\delta,\mathbf{d})$.

We have thus a contradiction in both subcases, so $(\gamma,\mathbf{c})\prec(\delta,\mathbf{d})$, as required.

Proof for the case $\mathbf{R}=\mathbf{R}^-(h)$: Define $\omega\geqslant0$ by $\cosh\omega:=-(\mathbf{c},\mathbf{d})$.

(i) Assume that $(\gamma,\mathbf{c})\prec\cdot(\delta,\mathbf{d})$. Then $\gamma_0<\delta_0$ and $\cosh(\delta_0-\gamma_0)=-(\mathbf{c},\mathbf{d})$, so $\omega=\delta_0-\gamma_0$. Suppose that for some $(\alpha,\mathbf{a})$ in $\mathbf{R}$, $(\alpha,\mathbf{a})\prec(\gamma,\mathbf{c})$, but $(\alpha,\mathbf{a})\not\prec(\delta,\mathbf{d})$. Define $\theta,\phi\geqslant0$ by $\cosh\theta:=-(\mathbf{a},\mathbf{c})$, $\cosh\phi:=-(\mathbf{a},\mathbf{d})$.

From $(\alpha,\mathbf{a})\prec(\gamma,\mathbf{c})$ follows that $\alpha_0<\gamma_0$ and $\cosh(\gamma_0-\alpha_0)>(\mathbf{a},\mathbf{c})$, so $\cosh(\gamma_0-\alpha_0)>\cosh\theta$, from which $\gamma_0-\alpha_0>\theta$. Since $\delta_0>\gamma_0>\alpha_0$, it follows from $(\alpha,\mathbf{a})\not\prec(\delta,\mathbf{d})$ that $\cosh(\delta_0-\alpha_0)\leqslant-(\mathbf{a},\mathbf{d})=\cosh\phi$, so $\delta_0-\alpha_0\leqslant\phi$. Thus $\theta+\omega=\theta+(\delta_0-\gamma_0)<\delta_0-\alpha_0\leqslant\phi$, so $\cosh(\theta+\omega)<\cosh\phi$.

Now

$$0\geqslant\begin{vmatrix}(\mathbf{a},\mathbf{a})&(\mathbf{a},\mathbf{c})&(\mathbf{a},\mathbf{d})\\(\mathbf{c},\mathbf{a})&(\mathbf{c},\mathbf{c})&(\mathbf{c},\mathbf{d})\\(\mathbf{d},\mathbf{a})&(\mathbf{d},\mathbf{c})&(\mathbf{d},\mathbf{d})\end{vmatrix}=\begin{vmatrix}-1&-\cosh\theta&-\cosh\phi\\-\cosh\theta&-1&-\cosh\omega\\-\cosh\phi&-\cosh\omega&-1\end{vmatrix}.$$

This can be written as

$$\{\cosh(\theta+\omega)-\cosh\phi\}\{\cosh(\theta-\omega)-\cosh\phi\}\leqslant0,$$

so since the first factor has been proven negative, $\cosh(\theta-\omega)\geqslant\cosh\phi$, from which $|\theta-\omega|\geqslant\phi$. However, $\theta+\omega<\phi$, so $\theta+\omega<|\theta-\omega|$, which implies that either θ or ω is negative, a contradiction. Thus no $(\alpha,\mathbf{a})$ with $(\alpha,\mathbf{a})\prec(\gamma,\mathbf{c})$ and $(\alpha,\mathbf{a})\not\prec(\delta,\mathbf{d})$ exists.

(ii) Assume that $(\gamma,\mathbf{c})\not\prec(\delta,\mathbf{d})$ and that for all events $(\alpha,\mathbf{a})$ in $\mathbf{R}$, $(\alpha,\mathbf{a})\prec(\gamma,\mathbf{c})$ implies $(\alpha,\mathbf{a})\prec(\delta,\mathbf{d})$. We compare the values of ω and $\delta_0-\gamma_0$.

If $\delta_0-\gamma_0>\omega\geqslant0$, then $\cosh(\delta_0-\gamma_0)>\cosh\omega=-(\mathbf{c},\mathbf{d})$, so we have the contradiction that $(\gamma,\mathbf{c})\prec(\delta,\mathbf{d})$.

If $\delta_0-\gamma_0<\omega$, then choose a number $\lambda\in\mathbf{I}_0$ with $\delta_0-\omega<\lambda<\gamma_0$. Since $\lambda\in\mathbf{I}_0$, then $\lambda=\alpha_0$ for some $\alpha\in\mathbf{I}$. Define $\mathbf{a}:=\mathbf{c}$; then since $\alpha_0<\delta_0$ and $\cosh(\gamma_0-\alpha_0)>1=-(\mathbf{a},\mathbf{a})=-(\mathbf{a},\mathbf{c})$, we have $(\alpha,\mathbf{a})\prec(\gamma,\mathbf{c})$. However, $\delta_0-\alpha_0<\omega$, so $\cosh(\delta_0-\alpha_0)<\cosh\omega=-(\mathbf{c},\mathbf{d})=-(\mathbf{a},\mathbf{d})$ and we have the contradiction that $(\alpha,\mathbf{a})\not\prec(\delta,\mathbf{d})$.

It follows that $\delta_0-\gamma_0=\omega$. If $\omega=0$, then $\gamma=\delta$ and $(\mathbf{c},\mathbf{d})=-1$, so $\mathbf{c}=\mathbf{d}$ and $(\gamma,\mathbf{c})=(\delta,\mathbf{d})$, another contradiction. Thus $\omega>0$, so $\delta_0>\gamma_0$ and $\cosh(\delta_0-\gamma_0)=\cosh\omega=-(\mathbf{c},\mathbf{d})$, i.e., $(\gamma,\mathbf{c})\prec\cdot(\delta,\mathbf{d})$.

Proof for the case $\mathbf{R}=\mathbf{R}^+(h)$: The proof for this case is analogous to that for the previous case and will thus be omitted (See Ref. 5, Lemma 3.2 for an essentially identical proof.) ■

We now prove Zeeman's lemma. Let $f:\mathbf{R}\to\mathbf{R}$ be a bijective function which preserves passive timelike causality, i.e., such that for all $(\alpha,\mathbf{a})$ and $(\beta,\mathbf{b})$ in $\mathbf{R}$, $(\alpha,\mathbf{a})\prec(\beta,\mathbf{b})$ if and only if $f(\alpha,\mathbf{a})\prec f(\beta,\mathbf{b})$. If $\mathbf{R}\neq\mathbf{R}^+(h)$, it follows immediately from Theorem 5.1 that for all $(\alpha,\mathbf{a})$ and $(\beta,\mathbf{b})$ in $\mathbf{R}$, $(\alpha,\mathbf{a})\prec\cdot(\beta,\mathbf{b})$ if and only if $f(\alpha,\mathbf{a})\prec\cdot f(\beta,\mathbf{b})$, i.e., f preserves passive lightlike causality.

If $\mathbf{R}=\mathbf{R}^+(h)$, the above argument works only when $\beta-\alpha\leqslant\pi$. The following theorem will enable us to extend the theorem to more distant points.

Theorem 5.2: If $(\alpha,\mathbf{a})$, $(\beta,\mathbf{b})$, and $(\gamma,\mathbf{c})$ are points in $\mathbf{R}^+(h)$ with

$$(\alpha,\mathbf{a})\prec\cdot(\beta,\mathbf{b}),\ (\beta,\mathbf{b})\prec\cdot(\gamma,\mathbf{c}),\ (\alpha,\mathbf{a})\prec\cdot(\gamma,\mathbf{c}),$$

then there exists a single null geodesic containing all three points.

Proof: We have that $(\mathbf{a},\mathbf{a})=(\mathbf{b},\mathbf{b})=(\mathbf{c},\mathbf{c})=1$, $\cos(\beta_0-\alpha_0)=(\mathbf{b},\mathbf{a})$, $\cos(\gamma_0-\beta_0)=(\mathbf{c},\mathbf{b})$, and

$\cos(\gamma_0 - \alpha_0) = (\mathbf{c},\mathbf{a})$, from which we calculate that the Gramian determinant

$$\begin{vmatrix} (\mathbf{a},\mathbf{a}) & (\mathbf{a},\mathbf{b}) & (\mathbf{a},\mathbf{c}) \\ (\mathbf{b},\mathbf{a}) & (\mathbf{b},\mathbf{b}) & (\mathbf{b},\mathbf{c}) \\ (\mathbf{c},\mathbf{a}) & (\mathbf{c},\mathbf{b}) & (\mathbf{c},\mathbf{c}) \end{vmatrix}$$

vanishes and so $\mathbf{a}$, $\mathbf{b}$, and $\mathbf{c}$ are linearly dependent. Then there exists a unit vector $\mathbf{d}\in\mathbb{R}^4$ orthogonal to $\mathbf{a}$ and scalars θ and ϕ such that $\mathbf{b} = \cos\theta\mathbf{a} + \sin\theta\,\mathbf{d}$ and $\mathbf{c} = \cos\phi\mathbf{a} + \sin\phi\mathbf{d}$. From the above values for $(\mathbf{a},\mathbf{a})$, $(\mathbf{a},\mathbf{b})$, etc., we calculate that θ and ϕ can be chosen to be $\beta_0 - \alpha_0$ and $\gamma_0 - \alpha_0$, respectively (it may be necessary to replace $\mathbf{d}$ by $-\mathbf{d}$), so all three points lie on the null geodesic with the equation $\mathbf{r}(\rho) = \cos(\rho_0 - \alpha_0)\mathbf{a} + \sin(\rho_0 - \alpha_0)\mathbf{d}$. ■

Now assume that $(\alpha,\mathbf{a})$ and $(\beta,\mathbf{b})$ are arbitrary points in $\mathbf{R}^+(h)$ with $(\alpha,\mathbf{a}) \lessdot (\beta,\mathbf{b})$. Choose consecutive points $(\alpha,\mathbf{a}) = (\gamma_0,\mathbf{c}_0)$, $(\gamma_1,\mathbf{c}_1),\ldots,(\gamma_n,\mathbf{c}_n) = (\beta,\mathbf{b})$ on the null geodesic joining them such that for all $i = 1,2,\ldots,n$, $\gamma_i - \gamma_{i-1} \leqslant \frac{1}{2}\pi$. Then $(\gamma_{i-1},\mathbf{c}_{i-1}) \lessdot (\gamma_i,\mathbf{c}_i)$ for all $i = 1,2,\ldots,n$ and $(\gamma_{i-1},\mathbf{c}_{i-1}) \lessdot (\gamma_{i+1},\mathbf{c}_{i+1})$ for all $i = 1,2,\ldots,n-1$. By Theorem 5.1, since $\gamma_i - \gamma_{i-1} \leqslant \frac{1}{2}\pi < \pi$, $f(\gamma_{i-1},\mathbf{c}_{i-1}) \lessdot f(\gamma_i,\mathbf{c}_i)$ for all $i = 1,2,\ldots,n$ and since $\gamma_{i-1} - \gamma_{i+1} \leqslant \pi$, $f(\gamma_{i-1},\mathbf{c}_{i-1}) \lessdot f(\gamma_{i+1},\mathbf{c}_{i+1})$ for all $i = 1,2,\ldots,n-1$. Then from Theorem 5.2 any three consecutive $f(\gamma_i,\mathbf{c}_i)$'s lie on a common null geodesic. Furthermore, since the ρ coordinates of any two consecutive $f(\gamma_i,\mathbf{c}_i)$'s differ by less than π (this condition is preserved from Theorem 5.1), two consecutive $f(\gamma_i,\mathbf{c}_i)$'s lie on at most one null geodesic. It follows that all $f(\gamma_i,\mathbf{c}_i)$'s lie on the *same* null geodesic, so since the ρ coordinate of $f(\alpha,\mathbf{a}) = f(\gamma_0,\mathbf{c}_0)$ is less than that of $f(\gamma_n,\mathbf{c}_n) = f(\beta,\mathbf{b})$, $f(\alpha,\mathbf{a}) \lessdot f(\beta,\mathbf{b})$.

This concludes our proof of Zeeman's lemma for Robertson–Walker space-times.

[1] E. C. Zeeman, "Causality implies the Lorentz group," J. Math. Phys. **5**, 490 (1964).

[2] S. W. Hawking and G. F. R. Ellis, *The Large-Scale Structure of Space-Time* (Cambridge, U.P., Cambridge, 1973).

[3] A. D. Alexandrov and V. V. Ovchinnikova, "Notes on the foundations of relativity theory," Vestnik Leningrad Univ. Math. **11**, 95 (1953).

[4] J. A. Lester, "Transformations of Robertson–Walker spacetimes preserving separation zero," Aequationes Math. **25**, 216 (1982).

[5] J. A. Lester, "The causal automorphisms of de Sitter and Einstein cylinder space-times," J. Math. Phys. **25**, 113 (1984).

Mode stability of the Kerr black hole

Bernard F. Whiting
Institute of Field Physics, Department of Physics and Astronomy, University of North Carolina, Chapel Hill, North Carolina 27599-3255

(Received 3 January 1989; accepted for publication 8 February 1989)

Separate differential and integral transformations are introduced for the individual radial and angular equations governing the (infinitesimally) gauge invariant Newman–Penrose quantities which represent massless perturbations of the Kerr black hole. Using these new transformations it is shown, without need for numerical investigation or reference to the analytic behavior of the separation constant, that no unstable mode perturbations exist for any physical value of the spin of massless fields on the rotating black hole background.

I. INTRODUCTION

In its most general form, which has been studied at great length over the last 30 years, the problem of black hole stability arises as an initial value problem for the stability of linearized gravitational perturbations on a black hole space-time background. Frequently, the methods used are also applicable to other massless fields of lower spin, so that often a general analysis is able to provide, simultaneously, results for all perturbations of physical or theoretical interest. The most recent results can be categorized[1] as (i) a transition away from mode analysis to obtain pointwise bounds for perturbation of the Schwarzschild space-time; (ii) a demonstration of mode stability for general massless perturbations of the Kerr black hole; and (iii) a specification of the criteria for stability in the boundary value problem for spherical geometries, which arises in the comparatively new context of gravitational thermodynamics. Although result (iii) will not be considered further here, it is perhaps useful to note that in the thermodynamic regime dominated by black hole geometries, the condition for gravitational stability exactly ensures that the various criteria are satisfied for the independently defined question of thermodynamical stability.[2]

With regard to the initial value problem for perturbations on spherical black hole space-times, whether eternal or forming from collapse, the best available results beyond mode stability[3,4] are found in the recent work of Kay and Wald.[5] Kay and Wald[5] have established that the evolution of regular (i.e., smooth and bounded), compact initial data will remain pointwise bounded in time throughout the entire domain of outer communication, including the boundary (horizon). An important point of physical relevance is that the result of Kay and Wald holds even for fields that do not vanish initially on the horizon (nor, in particular, on the bifurcation two-sphere of the global, Kruskal extension of the exterior Schwarzschild space-time).

The other new result pertinent to the initial value problem for black hole stability is the subject of this paper, viz. mode stability for perturbations of the Kerr (rotating) black hole. Although of a different nature than the most recent results for the Schwarzschild space-time, the work reported here represents a major step forward in the study of black hole stability. The present work is also of great astrophysical significance since almost all astronomical bodies rotate, including those that might eventually undergo gravitational collapse. Thus stability of rotating black holes is really a more pressing concern than is the stability of a static, spherically symmetric black hole.

In the remainder of this paper, an introduction to the Kerr metric is first given, after which the difficult nature of the stability problem that it presents is indicated. Then, from a tractable form of the perturbation equations, differential and integral transformations of the equations and their solutions are developed. Finally, the construction of a positive definite "energy" integral is given, permitting the proof of mode stability to be completed. The Appendix gives some mathematical details. The proof given here does not adhere closely to any previous work on this problem.

II. THE KERR BLACK HOLE

The Kerr metric, representing an axisymmetric, black hole solution to the source-free Einstein equations, was discovered by Kerr[6] almost 50 years after the spherically symmetric solution was first written. In subsequent work, Carter was able to establish,[7] rather unexpectedly, that the Hamilton–Jacobi equation for a free particle and the Klein–Gordon equation for a scalar field were separable. (Carter also studied the relationship between this result and the *form* of the Kerr metric[8]). Then, following a method used by Bardeen and Press[9] for (the gauge and tetrad invariant) perturbations of the Weyl tensor in the Schwarzschild geometry, Teukolsky demonstrated[10] that analogous perturbations in the case of the Kerr black hole also obeyed a separable equation. Similar perturbation equations have since been found[11,12] for all fields of physical interest in a background Kerr geometry.

III. PERTURBATION EQUATIONS

Massless Klein–Gordon,[7] Dirac (neutrino),[11] Maxwell,[10] Rarita–Schwinger,[12] and linearized Einstein[10] equations, when written in a decoupled form for the corresponding tetrad and gauge invariant Newman–Penrose[13] (NP) quantities in the Kerr background (and including the Schwarzschild limit), can all be represented in a single (separable) master equation of Teukolsky, which in Boyer–Lindquist coordinates[14] may be written as

 0022-2488/89/061301-05$02.50

$$\left\{\frac{\partial}{\partial r}\Delta\frac{\partial}{\partial r}-\frac{1}{\Delta}\left[(r^2+a^2)\frac{\partial}{\partial t}+a\frac{\partial}{\partial\phi}-(r-M)s\right]^2-4s(r+ia\cos\theta)\frac{\partial}{\partial t}\right.$$
$$\left.+\frac{\partial}{\partial\cos\theta}\sin^2\theta\frac{\partial}{\partial\cos\theta}+\frac{1}{\sin^2\theta}\left[a\sin^2\theta\frac{\partial}{\partial t}+\frac{\partial}{\partial\phi}+i\cos\theta\cdot s\right]^2\right\}\psi_{*_s}=0,$$

where $\Delta=r^2-2Mr+a^2$ (ψ_{*_s} can be given suitably in terms of the relevant NP quantities). In the case of spin-zero fields on the Schwarzschild black hole, this equation immediately admits a conserved "energy" integral with a positive definite integrand, establishing that there are no modes that have an unbounded time derivative, i.e., there are no unstable modes. The radial part of the equation is complex when acting on an individual mode of higher spin, but by a transformation of the radial function for obtaining the Regge–Wheeler[3] (or corresponding[15]) equation, a similar proof of stability can again be given[4] for massless perturbations of spin-2 (and spin-1) in the Schwarzschild space-time.

Except for axisymmetric scalar perturbations, no analogous consideration has previously yielded even this limited proof of stability in the case of Kerr black holes for the following two reasons.

(i) For nonaxisymmetric scalar perturbations the coefficient $(1/\sin^2\theta-a^2/\Delta)$ of $|\partial\psi_{*_s}/\partial\phi|^2$ in the energy integrand is only positive outside the ergosphere, indicative of the fact that there is no Killing vector which is timelike everywhere within the region exterior to the event horizon.

(ii) Prior to this work, although transformations were known[16] that mapped Kerr radial functions to solutions of equations that reduced to Regge–Wheeler-type equations in the Schwarzschild limit, for the rotating black hole these other equations are quite unlike the Regge–Wheeler equation[3] in that they depend on the separation constant (i.e., the unknown angular eigenvalue) in a highly nonlinear way.

Procedures that remedy the above difficulties are given below.

Physical considerations[11] lead one to regard as unstable modes those perturbation solutions that are purely ingoing on the horizon and purely outgoing at (null) infinity. Consequently, in the exterior region, unstable modes have, asymptotically, support only on the *future* horizon and at *future* null infinity. Unstable modes have characteristic frequencies with positive imaginary parts; thus on a spacelike section, they can be regarded as radial eigenfunctions, which become unboundedly large to the future.

In the development of a proof of the mode stability for perturbations of a Kerr black hole, new progress has become possible through a generalization of certain previously known results; these results are detailed as follows.

(i) There are Teukolsky–Starobinsky[17] ordinary differential relations that can be used to change helicity from s to $-s$ for the radial and angular dependence of the separated solutions of Teukolsky's master equation.[10]

(ii) The kernel of an integral equation can be written[18] for radial functions of spin-zero fields in the Schwarzschild background.

In order to proceed, it will be convenient to introduce a notation for exploiting the similarity between the r and $\cos\theta$ dependence of the Teukolsky equation.[10]

The separated radial and angular equations can be written in the form [for $\psi_{*_s}=e^{-i\omega t}e^{im\phi}R_s(r)S_s(\theta)$]

$$\left\{\partial_{xx}-\alpha^2+\frac{\alpha\kappa+\lambda+\frac{1}{2}\kappa^2}{x}+\frac{\frac{1}{4}-\beta^2}{x^2}\right.$$
$$\left.+\frac{\alpha\kappa-\lambda-\frac{1}{2}\kappa^2}{x-1}+\frac{\frac{1}{4}-\gamma^2}{(x-1)^2}\right\}\sqrt{x(x-1)}\,u=0,$$

where for the angular equation, $x=(\cos\theta+1)/2$, $u=S_s$,

$$\alpha=2a\omega,\quad \kappa=s,$$
$$\beta=(s-m)/2,\quad \gamma=(s+m)/2,$$
$$\lambda=\tfrac{1}{2}+\alpha(\gamma-\beta)-\tfrac{1}{2}(\gamma-\beta)^2+(\lambda_T+s)$$
(with α,β,γ given above)

and for the radial equation, $x=(r-r_-)/(r_+-r_-)$, $u=R_s$,

$$\alpha=2iM\omega\epsilon_0,\quad \kappa=s-2iM\omega,$$
$$\beta=(s/2+iM\omega)-(i/\epsilon_0)(M\omega-am/2M),$$
$$\gamma=(s/2+iM\omega)+(i/\epsilon_0)(M\omega-am/2M),$$
$$\lambda=\tfrac{1}{2}+\alpha(\gamma-\beta)-\tfrac{1}{2}(\gamma-\beta)^2+(\lambda_T+s),$$
(with α,β,γ given above).

The separation constant λ_T appears in Teukolsky's radial equation[10]; (λ_T+s) is invariant under $s\to -s$. The relations $\Delta=r^2-2Mr+a^2=(r-r_+)(r-r_-)$ and $\epsilon_0=(r_+-r_-)/(r_++r_-)$ have been used.

$$\left.\begin{array}{l}\text{Near } x=0,\ u\sim x^{\delta'\beta},\\ \text{near } x=1,\ u\sim(x-1)^{\delta''\gamma},\\ \text{near } x=\infty,\ u\sim e^{\delta\alpha x}x^{-\delta\kappa-1},\end{array}\right\}\ \delta,\delta',\delta''=\pm 1.$$

We are particularly interested in transformations of the above equations which leave the singular points (number and type) and (λ_T+s) dependence unchanged.

IV. DIFFERENTIAL TRANSFORMATIONS

A general structure[19] can be shown to underlie the contiguous relations for special functions; familiar examples would be the differential operators which change angular momentum for the spherical Bessel and associated Legendre functions. By examining the Teukolsky–Starobinsky relations[17] in this context one can view them as a particular consequence of the following. For those values of ϵ, ϵ', $\epsilon''(=\pm 1)$ that allow

$$n=\epsilon\kappa+\epsilon'\beta+\epsilon''\gamma$$

to be a positive integer (there can be four such at most), then with

$$\tilde{\alpha}=-\epsilon\alpha,\quad \tilde{\kappa}=\epsilon'\beta+\epsilon''\gamma,$$
$$\tilde{\beta}=n/2-\epsilon'\beta,\quad \tilde{\lambda}=\lambda,$$
$$\tilde{\gamma}=n/2-\epsilon''\gamma$$

the function given by

$$\tilde{u} = e^{\tilde{\alpha}x}x^{\tilde{\beta}}(x-1)^{\tilde{\gamma}}\left(\frac{\partial}{\partial x}\right)^n e^{\epsilon\alpha x}x^{\epsilon'\beta}(x-1)^{\epsilon''\gamma}u$$

satisfies a similar equation to that satisfied by u, where the new parameters are given by the overtilde quantities above. The set of equations that can be obtained by repeated application of this result is finite. A certain degeneracy occurs when the derivatives act on the polynomials that they annihilate. There is an inverse transformation operator

$$e^{\epsilon\alpha x}x^{\epsilon'\beta}(x-1)^{\epsilon''\gamma}\left(\frac{\partial}{\partial x}\right)^n e^{\tilde{\alpha}x}x^{\tilde{\beta}}(x-1)^{\tilde{\gamma}}$$

which maps solutions for $\tilde{u}$ onto solutions for u, with a similar statement as above applying in the case of degeneracy.

The Teukolsky–Starobinsky (helicity flipping) relations[17] are a known example of this kind of transformation, with $n = |2s|$ in the radial and angular transformations. For the radial functions, no other integer values of n are possible for a *general* value of the frequency except on a Schwarzschild space-time, in which case a new expression of known results is obtained.[20] For the angular functions, $n = |s-m|$ and $n = |s+m|$ are always additionally possible; they map angular eigenfunctions to eigenfunctions of two new operators which are related to one another by $x \to 1-x$, i.e., by $\cos\theta \to \cos(\pi-\theta)$. A solution of one of these new operators is given by

$${}_mT_{ls} = (\sin\theta)^{|s-m|}\left(\frac{\epsilon\partial}{\partial\cos\theta} + a\omega + \frac{(s+m\cos\theta)}{\sin^2\theta}\right)^{|s-m|}{}_sS_{lm}(\theta),$$

where $s-m = \epsilon|s-m|$. This new angular function T will be used in a proof of stability: For nonzero $\alpha = 2a\omega$, its construction can never be degenerate.

V. INTEGRAL TRANSFORMATIONS

The known integral equation[18] for scalar wave functions in the Schwarzschild space-time was of the Laplace type, where the "center" of the kernel [i.e., that part of the integral kernel given by the nonseparable functions $H(x,y)$ which occur in the expressions below] is of the form e^{-axy}. For all the radial and angular functions arising from perturbations of the Kerr black hole, I have recently found kernels of integral equations which they satisfy, where these kernels are now more complicated functions of xy: Similarly, Euler-type kernels depending on $(x+y-1)$, etc., have also been found. Moreover, the conditions that disallow the existence of integral kernels simply depending on the functions xy, $x+y-1$ (or their variants) are precisely those conditions that permit the construction of integral transformations[21] to the solutions of *new* equations of the same general type as we need to consider. In the radial case, for the spin-2 (gravitational) perturbations on the Schwarzschild background, one of these new equations turns out to be the Regge–Wheeler equation[3] previously related to the NP perturbations[13] only by a differential transformation: Its simple generalization in the Kerr case will again be useful in a proof of stability.

These integral transforms can be described as follows. Under suitable conditions, the function

$$\tilde{u} = \int_A^B \mathscr{K}(x,y)u(y)dy$$

will satisfy an equation of our given form provided that ϵ, ϵ', $\epsilon''(=\pm 1)$ can be chosen so that $y(y-1)\times W(u(y),\mathscr{K}(x,y))|_A^B$ vanishes identically. (Here W is the Wronskian.) The function $\mathscr{K}(x,y)$ has the general form

$$e^{\tilde{\alpha}x}x^{\tilde{\beta}}(x-1)^{\gamma}H(x,y)e^{\epsilon\alpha y}y^{\epsilon'\beta}(y-1)^{\epsilon''\gamma}$$

and a number of different usable "centers" $H(x,y)$ have been identified, e.g., $e^{-2\epsilon\alpha xy}$, $(x+y-1)^{-\nu-1}$, etc. (where $\nu = \epsilon\kappa + \epsilon'\beta + \epsilon''\gamma$ whether or not it is an integer or real). For the first of these "centers," the quantities

$$\tilde{\alpha} = \epsilon\alpha, \quad \tilde{\kappa} = \epsilon'\beta - \epsilon''\gamma,$$
$$\tilde{\beta} = \tfrac{1}{2}(\epsilon\kappa + \epsilon'\beta + \epsilon''\gamma), \quad \tilde{\lambda} = \lambda,$$
$$\tilde{\gamma} = \tfrac{1}{2}(-\epsilon\kappa + \epsilon'\beta + \epsilon''\gamma),$$

give the parameters in the equation for $\tilde{u}$. (Note that λ is again unchanged.)

We will choose a bounded new radial function ${}_mK_{ls}$ given by an integral transform of ${}_{-|s|}R_{lm}(r)$ over the range (r_+,∞), with ϵ,ϵ', $\epsilon'' = -1$. Since the "center" is of the Laplace type, this integral transformation will never be degenerate. In what follows, stability for negative helicity modes will also assure stability for positive helicity modes because, via the Teukolsky–Starobinsky relations,[17] the modes can be independently transformed into one another. (Only for the algebraically special perturbations are these transformations singular,[22] but then the boundary conditions given above for unstable modes are not satisfied.)

VI. PROOF OF STABILITY

The transformations that have been chosen ensure that the Kerr angular and radial functions corresponding to an unstable mode will map to bounded solutions of the new operators. In addition, this construction implies that the function

$$\Phi_s = e^{-i\omega t}e^{im\phi}K_s(r)T_s(\theta)$$

will satisfy the equation given below, where

$$f(r)\left\{\frac{(r-r_-)^2}{(r-r_+)^2} - \frac{a^2}{M^2}\frac{(r-M)^2}{\Delta}\right\}\frac{\Delta}{\epsilon_0^2}\ (>0 \text{ for } r>r_+):$$

$$\left[\frac{\partial}{\partial r}\Delta\frac{\partial}{\partial r} + \frac{1}{\sin\theta}\frac{\partial}{\partial\theta}\sin\theta\frac{\partial}{\partial\theta} - (f(r)+a^2\cos^2\theta)\frac{\partial^2}{\partial t^2} - 2a\left(\cos\theta + \frac{r-M}{\epsilon_0 M}\right)\frac{\partial^2}{\partial t\,\partial\phi} - s^2\left(\frac{1-\cos\theta}{1+\cos\theta} + \frac{r-r_+}{r-r_-}\right)\right]\Phi_s = 0.$$

Here the operator is totally independent of Teukolsky's separation constant λ_T.[10] (Note that in the metric that can be derived from this equation, $\partial/\partial t$ is globally null.) The conserved quantity that follows from the operator has a positive integrand:

$$\frac{\partial}{\partial t}\frac{1}{2}\int dr\,d\theta\,d\phi\,\sin\theta\left\{(f(r)+a^2\cos^2\theta)\left|\frac{\partial\Phi}{\partial t}\right|^2+\Delta\left|\frac{\partial\Phi}{\partial r}\right|^2+\left|\frac{\partial\Phi}{\partial\theta}\right|^2+s^2\left(\frac{1-\cos\theta}{1+\cos\theta}+\frac{r-r_+}{r-r_-}\right)|\Phi|^2\right\}=0.$$

Since the leading radial dependence of Φ_s near the horizon is $(r-r_+)^{-2iM\omega}$ and near infinity it is $e^{i\omega r}$ and the leading angular dependence near the south pole is $(\sin\theta)^s$, every term in the above integrand is integrable for unstable modes. Hence the value of the conserved "energy" bounds the integral of the time derivative terms, which consequently cannot grow exponentially. Thus there can be no unstable modes for Kerr angular and radial functions since we have now ruled out the solutions of the above equation to which they would be mapped.

Parameters in the transformations depended explicitly on the mode decomposition. Consequently, it seems that any stronger result for the Kerr black hole would require additional understanding concerning mode completeness since the particular methods of Kay and Wald,[5] which might circumvent this, are not directly applicable and no appropriate generalization is known at present.

ACKNOWLEDGMENTS

I wish to acknowledge the continued interest of S. Chandrasekhar from the Enrico Fermi Institute, Chicago. I am grateful to J. W. York, Jr. for comments on the manuscript.

This paper is the culmination of research carried out at various times over several years: in DAMTP, Cambridge, with support from the SERC; at the Centre for Theoretical Physics, Austin, with support from the NSF; and at l'Observatoire de Paris, Meudon, with support from the CNRS and the Bourse Jolliot–Curie. This paper has been prepared for publication in the Institute of Field Physics, Chapel Hill, also with support from the NSF.

APPENDIX

Properties of the transformations referred to in this paper may be sufficiently unfamiliar to some that it would be useful here to sketch a proof of the most important relations. A simple derivation of the differential transformation can be obtained via the use of the following two elementary results:

Result (i):

$$e^{\epsilon\alpha x}x^{\epsilon'\beta}(x-1)^{\epsilon''\gamma}\left\{\partial_{xx}-\alpha^2+\frac{\alpha\kappa+\Lambda}{x}+\frac{\frac{1}{4}-\beta^2}{x^2}+\frac{\alpha\kappa-\Lambda}{x-1}+\frac{\frac{1}{4}-\gamma^2}{(x-1)^2}\right\}\sqrt{x(x-1)}$$
$$=\frac{1}{\sqrt{x(x-1)}}\{x(x-1)\partial_{xx}+(ax^2+bx+c)\partial_x+dx+e\}e^{\epsilon\alpha x}x^{\epsilon'\beta}(x-1)^{\epsilon''\gamma},$$

where

$$a=-2\epsilon\alpha,\quad b=2(\epsilon\alpha-\epsilon'\beta-\epsilon''\gamma+1),\quad c=2\epsilon'\beta-1,$$
$$d=2\epsilon\alpha(\epsilon\kappa+\epsilon'\beta+\epsilon''\gamma-1),\quad e=-\alpha\kappa-\Lambda-(2\epsilon'\beta-1)(\epsilon\alpha+\tfrac{1}{2}-\epsilon''\gamma).$$

Result (ii):

$$\partial_x^N\{x(x-1)\partial_{xx}+(ax^2+bx+c)\partial_x-a(N-1)x+e\}$$
$$=\{x(x-1)\partial_{xx}+(ax^2+(b+2N)x+c-N)\partial_x+a(N+1)x+e+N(b+N-1)\}\partial_x^N,$$

which accounts for the necessity that N be a positive integer.

In the case of the integral transformations, explicit demonstration of the results depends on the form of the integral "center" $H(x,y)$, e.g., for $H(x,y)=e^{-axy}$ we have

$$[x(x-1)\partial_{xx}+(ax^2+bx+c)d_x-avx+e]e^{-axy}=[y(y-1)\partial_{yy}+(ay^2+by+v)\partial_y-acy+e]e^{-axy}.$$

When $c=v$ it is thus possible to construct an integral equation for $u(x)$.

Similarly, for $H(x,y)=(x-y)^v$, say, we can obtain

$$[x(x-1)\partial_{xx}+(ax^2+bx+c)\partial_x-avx+e](x-y)^v$$
$$=[\,y(y-1)\partial_{yy}+\{ay^2+(b+2(v+1))y+c-(v+1)\}\partial_y+a(v+2)x+e+(v+1)(b+v)\,]^\dagger(x-y)^v,$$

where we have indicated by the superscript † the adjoint operator which arises under the integral sign in a verification of the transform properties. Note the similarity of the transformed operator here to the operator arising for the differential transform (with $N\to v+1$). The transformation properties for other similar "centers," e.g., for those depending on $(x-1)(y-1)$ or $(x+y-1)$, can be constructed directly or obtained by the substitution $x\to 1-x$, etc.

Kernels for integral equations can be produced in abundance as a result of the elementary (but nontrivial) observations

$$\{x(x-1)\partial_{xx}+(ax^2+bx+c)\partial_x+dx+e\}-\{y(y-1)\partial_{yy}+(ay^2+by+c)\partial_y+dy+e\}$$
$$=(x-y)[(u\partial_{uu}+(au-c)\partial_u+d+\mu)-(v\partial_{vv}+(-av+a+b+c)\partial_v+\mu)]$$
$$=(x-y)\{(f\partial_{ff}+(af+a+b)\partial_f+d+\sigma/f)-[(1-g)/f]$$
$$\times[g(1-g)\partial_{gg}+(-(2+c)g+a+b+c)\partial_g+\sigma/(1-g)]\},$$

where $u = xy$, $v = (x-1)(y-1)$, $f = x + y - 1$, and $g = (x-1)(y-1)/xy$. Here, μ and σ would be new constants of separation.

[1]B. F. Whiting, "The black-hole stability problem," in *Proceedings of the Fifth Marcel Grossman Meeting*, Perth, Western Australia, 1988, edited by D. G. Blair and M. J. Buckingham (Cambridge U.P., Cambridge, 1989).

[2]J. W. York, Phys. Rev. D **33**, 2092 (1986); B. F. Whiting and J. W. York, Phys. Rev. Lett. **61**, 1336 (1988); B. F. Whiting, "Black holes and Thermodynamics," in *Proceedings of the Ninth IAMP Congress*, Swansea, Wales, 1988, edited by B. Simon, I. M. Davies, and A. Truman (Hilger, Bristol, 1989).

[3]T. Regge and J. A. Wheeler, Phys. Rev. **108**, 1063 (1957).

[4]C. V. Vishveshwara, Phys. Rev. D **10**, 2870 (1970).

[5]R. M. Wald, J. Math. Phys. **20**, 1056 (1979); J. Dimock and B. S. Kay, Ann. Phys. **175**, 366 (1987), footnote 13 and Lemma 5.10; B. S. Kay and R. M. Wald, Class. Quant. Grav. **4**, 893 (1987).

[6]R. P. Kerr, Phys. Rev. Lett. **11**, 237 (1963).

[7]B. Carter, Commun. Math. Phys. **10**, 280 (1968).

[8]B. Carter, "Black hole equilibrium states," in *Black Holes*, edited by C. DeWitt and B. S. DeWitt (Gordon and Breach, New York, 1973).

[9]J. M. Bardeen and W. H. Press, J. Math. Phys. **14**, 7 (1973).

[10]S. A. Teukolsky, Phys. Rev. Lett. **29**, 1114 (1972).

[11]S. A. Teukolsky, Astrophys. J. **185**, 635 (1973).

[12]R. Güven, Phys. Rev. D **22**, 2327 (1980).

[13]E. T. Newman and R. Penrose, J. Math. Phys. **3**, 566 (1962).

[14]R. H. Boyer and R. W. Lindquist, J. Math. Phys. **8**, 265 (1967).

[15]J. A. Wheeler, Phys. Rev. **97**, 511 (1955).

[16]S. Chandrasekhar and S. Detweiler, Proc. R. Soc. London Ser. A **350**, 165 (1976).

[17]W. H. Press and S. A. Teukolsky, Astrophys. J. **185**, 649 (1973).

[18]S. Persides, J. Math. Phys. **14**, 1017 (1973).

[19]B. F. Whiting, "The relation of solutions of different ODE's is a commutation relation," in *Differential Equations, The Proceedings of the International Conference on Differential Equations*, Birmingham, AL, 1983, edited by I. W. Knowles and R. T. Lewis (North-Holland, Amsterdam, 1984), pp. 561–570.

[20]S. Chandrasekhar, Proc. R. Soc. London Ser. A **343**, 289 (1975).

[21]E. L. Ince, *Ordinary Differential Equations* (Dover, New York, 1956), Chap. VIII, especially Sec. 8.601.

[22]R. M. Wald, J. Math. Phys. **14**, 1453 (1973).

Duality and conformal structure

Tevian Dray[a)]
Raman Research Institute, Bangalore 560080, India
and Department of Mathematics, Oregon State University, Corvallis, Oregon 97331

Ravi Kulkarni[b)] and Joseph Samuel[c)]
Raman Research Institute, Bangalore 560080, India

(Received 22 November 1988; accepted for publication 15 February 1989)

In four dimensions, two metrics that are conformally related define the same Hodge dual operator on the space of two-forms. The converse, namely, that two metrics that have the same Hodge dual are conformally related, is established. This is true for metrics of arbitrary (nondegenerate) signature. For Euclidean signature a stronger result, namely, that the conformal class of the metric is completely determined by choosing a dual operator on two-forms satisfying certain conditions, is proved.

I. INTRODUCTION

Self-dual fields have played a major role in many of the recent developments in general relativity. Foremost among these is Penrose's twistor program (see, e.g., Refs. 1 and 2). Such fields seem to be fundamentally involved in attempts to quantize gravity (see, e.g. Ref. 3), notably Ashtekar's new variables (for a review see Ref. 4). The material discussed here was motivated by the attempts of one of us (JS) to better understand Ashtekar's new variables, but the presentation is entirely self-contained.

In four dimensions the Hodge dual operator on two-forms is manifestly conformally invariant. Thus, if two metrics are conformally related, they have the same Hodge dual. We show that the converse is also true: Two metrics of arbitrary (nondegenerate) signature that have the same Hodge dual are conformally related. For Riemannian manifolds (signature $+\ +\ +\ +$), we are able to establish a much stronger result: Any choice of a three-dimensional, positive-definite subspace of the space of two-forms determines a dual operator, which in turn determines a conformal class of metrics whose Hodge dual agrees with the original choice. Thus there is a one–one correspondence between conformal classes of metrics and dual operators.

Our presentation is organized as follows. After establishing the notation in Sec. II we show that the equality of Hodge duals implies that the metrics are conformally related. In Sec. III we consider the Riemannian case and establish the stronger result described above. Finally, in Sec. IV we discuss our results.

II. CONVERSE

Let M be an oriented four-dimensional manifold with (nondegenerate) metric g_{ab}. The volume element (Levi-Civita tensor) is the four-form $\epsilon_{abcd} = \epsilon_{[abcd]}$, which agrees with the orientation and whose nonzero components are $\pm\sqrt{|g|}$, where $g = \det(g_{ab})$. Denote by Λ^2, the space of two-forms $F_{ab} = F_{[ab]}$ on M. Then the Hodge dual operator $*$, defined by g_{ab}, is a map from $\Lambda^2 \to \Lambda^2$ given by

$$(*F)_{ab} = \tfrac{1}{2}\epsilon_{ab}{}^{cd}F_{cd}. \tag{1}$$

It is straightforward to check that

$$\hat{g}_{ab} = \Omega g_{ab} \Rightarrow * \equiv *. \tag{2}$$

One also has

$$** = \pm I, \tag{3}$$

where I is the identity operator and

$$\epsilon_{abmn}\epsilon_{cd}{}^{mn} = \pm 4g_{a[c}g_{d]b}, \tag{4}$$

where the $-$ sign holds for the Lorentzian signature $[(-\ +\ +\ +)$ or $(+\ -\ -\ -)]$ and the $+$ sign holds for all other signatures. We are now ready to prove the converse to (2).

Theorem 1: Let g_{ab} and $\hat{g}_{ab}$ be (real, nondegenerate) metrics of arbitrary signature on a four-dimensional manifold M, such that for all two-forms F on M,

$$*F \equiv \hat{*}F. \tag{5}$$

Then

$$\hat{g}_{ab} = \pm \Omega g_{ab}, \tag{6}$$

where $\Omega = |\hat{g}/g|^{1/4}$.

Proof:

Step 1: Equation (1) implies

$$\hat{\epsilon}_{cd}{}^{mn}F_{mn} = \epsilon_{cd}{}^{mn}F_{mn}, \quad \forall F \in \Lambda^2, \tag{7a}$$

which implies

$$\hat{\epsilon}_{cd}{}^{mn} = \epsilon_{cd}{}^{mn}. \tag{7b}$$

But from the definition of the volume element,

$$\hat{\epsilon}_{abmn} = \Omega^2\epsilon_{abmn}. \tag{8}$$

Contracting (8) with (7b), using (4), yields

$$\hat{g}_{a[c}\hat{g}_{d]b} = \Omega^2 g_{a[c}g_{d]b}. \tag{9}$$

Step 2: It is sufficient to establish (6) at each point $p \in M$. Choose coordinates x^i on a neighborhood of p so that $\hat{g}_{ij}|_p$ is diagonal. Then

$$\hat{g}_{i[j}\hat{g}_{k]l} = 0, \quad \text{unless } (i,k) = (j,l) \text{ or } (i,j) = (k,l). \tag{10}$$

[a)] Permanent address: Department of Mathematics, Oregon State University, Corvallis, Oregon 97331.
[b)] Permanent address: Department of Mathematics, University of Poona, Pune 411007, India.
[c)] Present address: Department of Physics, University of Utah, Salt Lake City, Utah 84112.

 0022-2488/89/061306-04$02.50

In particular, (10) holds if j, k, and l are all different. (Here and for the remainder of the proof, all quantities are to be evaluated at p.)

Using (9) and (10) we have

$$\begin{aligned}(g_{ii}g_{jj} - g_{ij}^2)g_{kl}^2 &= (g_{ii}g_{kl})(g_{jj}g_{kl}) - (g_{ij}g_{kl})(g_{ij}g_{lk}) \\ &= (g_{ik}g_{il})(g_{jk}g_{jl}) - (g_{ik}g_{jl})(g_{il}g_{jk}) \\ &= 0 \quad (\text{for } \epsilon_{ijkl} \neq 0). \end{aligned} \tag{11}$$

But since $\hat{g}_{ii}\hat{g}_{jj} - \hat{g}_{ij}^2 \neq 0$ by assumption, one final use of (9) yields

$$g_{kl} = 0, \tag{12}$$

so that g_{ij} is also diagonal at p.

Step 3: Inserting the diagonality of both $\hat{g}_{ij}$ and g_{ij} into (9) yields

$$\hat{g}_{ii}\hat{g}_{jj} = \Omega^2 g_{ii}g_{jj}, \quad \text{for } i \neq j, \tag{13}$$

which implies the result (6). Q.E.D.

III. EUCLIDEAN SIGNATURE

We now turn to the special case of an oriented Riemannian manifold [signature $(+\,+\,+\,+)$ or $(-\,-\,-\,-)$] with volume element ϵ_{abcd}. First we need some results about the vector space Λ_p^2 of two-forms at a point $p \in M$.

There is a natural product (symmetric bilinear form) on Λ^2, given by the wedge product of forms, namely,

$$\langle F_{ab}, G_{cd} \rangle = \epsilon^{abcd} F_{ab} G_{cd}, \tag{14a}$$

or equivalently

$$\langle F, G \rangle \epsilon = F \wedge G, \tag{14b}$$

where $\epsilon^{abcd} = \epsilon^{[abcd]}$ is defined by

$$\epsilon_{abcd}\epsilon^{abcd} = 4! \ . \tag{15}$$

Note that the metric has not been used in defining (14) and that the inner product is not positive definite. If one chooses a basis α^i of the space Λ_p^1 of one-forms at p, then the independent $\alpha^i \wedge \alpha^j$ form a basis for Λ_p^2. In four dimensions there are six such two-forms, so $\dim \Lambda_p^2 = 6$. Furthermore, by choosing appropriate linear combinations, one easily sees that the signature of the wedge product (14) on Λ_p^2 is $(+\,+\,+\,-\,-\,-)$.

Lemma 1: Given a vector space V with a symmetric bilinear form $\omega: V \times V \to \mathbb{R}$ and a subspace $W^+ \subset V$, such that (W^+, ω) is an inner product space (i.e., $\omega|_{W^+}$ is positive definite), there exists an operator $\#: V \to V$, such that

$$V = W^+ \oplus W^-, \tag{16a}$$

with

$$W^{\pm} = \tfrac{1}{2}(I \pm \#)V. \tag{16b}$$

Proof: Pick an orthonormal basis w_i of W^+. Then the projection operator P from V to W^+ is given by

$$\begin{aligned} &P: V \to W^+, \\ &\quad v \mapsto \langle v, w_i \rangle w_i. \end{aligned} \tag{17}$$

Define $\#$ by

$$\#v = 2Pv - v. \tag{18}$$

Then

$$P = \tfrac{1}{2}(I + \#), \tag{19}$$

and the result follows. Q.E.D.

Corollary: It is an immediate consequence of (18) that

$$\#\# = I. \tag{20}$$

Furthermore, since $(I + \#)(I - \#) = 0$, we have

$$\langle w^+, w^- \rangle = 0, \quad \forall w^{\pm} \in W^{\pm}, \tag{21a}$$

or equivalently

$$\langle v_1, \#v_2 \rangle = \langle \#v_1, v_2 \rangle, \quad \forall v \in V. \tag{21b}$$

Let T_pM denote the tangent space to M at p and for $T \in T_pM$ let $\mathrm{Ker}(T) = \{\alpha \in \Lambda_p^1 : \alpha(T) := \alpha_a T^a = 0\}$ denote the kernel.

Lemma 2: Let Λ^+ be a three-dimensional subspace of Λ_p^2, such that

$$0 \neq F \in \Lambda^+ \ \Rightarrow \ F \wedge F \neq 0. \tag{22}$$

Then for each vector $T \in T_pM$, the map

$$\begin{aligned} &\phi_T: \Lambda^+ \to \mathrm{Ker}\ T, \\ &\quad F \mapsto F(T) := F_{ab} T^b, \end{aligned} \tag{23}$$

is an ismorphism.

Proof: Since $\dim(\mathrm{Ker}\ T) = 3$ is suffices to show that ϕ_T is one–one, i.e., that $F(T) = 0$ implies $F = 0$. But by choosing a basis of $\mathrm{Ker}\ T$ and extending it to a basis α^i of Λ_p^1 and then forming the associated basis of Λ_p^2, one sees that

$$F(T) = 0 \ \Rightarrow \ F \wedge F = 0. \tag{24}$$

Using (22) now proves the assertion. Q.E.D.

For the remainder of this section, we assume that Λ^+ is as in Lemma 2, and that the wedge product (14) is positive definite on Λ^+. Using Lemma 1, this is equivalent to giving a dual operator $\#$ on Λ_p^2; Λ^+ is the space of *self-dual* two-forms, i.e.,

$$F \in \Lambda^+ \ \Leftrightarrow \ \#F = F. \tag{25}$$

We now show how to construct a conformal metric $h_{\#}$ from Λ^+. Fix any $\eta \in \Lambda_p^1$. Then for any $\alpha, \beta \in \Lambda_p^1$, choose $T \in \mathrm{Ker}\ \alpha \cap \mathrm{Ker}\ \beta \cap \mathrm{Ker}\ \eta$. By Lemma 2, there exist unique $F_\alpha, F_\beta, F_\eta \in \Lambda^+$, so that $F_\alpha(T) = \alpha$, $F_\beta(T) = \beta$, and $F_\eta(T) = \eta$. Define

$$h_{\#}(\alpha, \beta)/h_{\#}(\eta, \eta) = \langle F_\alpha, F_\beta \rangle / \langle F_\eta, F_\eta \rangle. \tag{26}$$

This defines $h_{\#}$ up to the single choice of the scale $h_{\#}(\eta, \eta)$, i.e., $h_{\#}$ is determined up to a conformal factor.

We now establish that $h_{\#}$ is well defined, i.e., that the right-hand side of (26) is unchanged under the transformation $T \mapsto T'$ with both $T, T' \in \mathrm{Ker}\ \alpha \cap \mathrm{Ker}\ \beta \cap \mathrm{Ker}\ \eta$. This is obvious if T' is a multiple of T, so we will assume that T and T' are linearly independent. First we introduce some notation.

Extend T and T' to a basis $\{e_0 = T, e_1 = T', e_2, e_3\}$ of T_pM and let $\{\omega^0, \omega^1, \omega^2, \omega^3\}$ be the dual basis of Λ_p^1. Let F_2, F_3, F_2', and F_3' be the unique elements of Λ^+ obtained using Lemma 2 that satisfy

$$F_2(T) = \omega^2 = F_2'(T'), \quad F_3(T) = \omega^3 = F_3'(T'). \tag{27}$$

But since (27) implies

$$F_2(T',T) = 0 = F_2(T',T'), \tag{28}$$

we have

$$F_2(T'') = A_{22}\omega^2 + A_{23}\omega^3, \tag{29a}$$

for some constants A_{22} and A_{23} so that, again using Lemma 2,

$$F_2 \equiv A_{22}F_2' + A_{23}F_3'. \tag{29b}$$

Similarly,

$$F_3 \equiv A_{32}F_2' + A_{33}F_3'. \tag{29c}$$

Since this argument can be reversed to express F_2', and F_3' in terms of F_2 and F_3, we must have

$$\lambda := \det A \equiv A_{22}A_{33} - A_{23}A_{32} \neq 0. \tag{30}$$

Now consider the transformation $T_pM \to T_pM$, defined by

$$\begin{aligned} &e_0 \mapsto e_1, \quad e_1 \mapsto \lambda_0 + be_1, \\ &e_2 \mapsto e_2 + ce_1, \quad e_3 \mapsto e_3 + de_1, \end{aligned} \tag{31a}$$

where

$$b := \operatorname{tr} A = A_{22} + A_{33}, \tag{31b}$$

and c,d are constants to be determined. The induced transformation on Λ_p^1 is

$$\begin{aligned} &\omega^0 \mapsto -(b/\lambda)\omega^0 + \omega^1 - c\omega^2 - d\omega^3, \\ &\omega^1 \mapsto (1/\lambda)\omega^0, \quad \omega^2 \mapsto \omega^2, \quad \omega^3 \mapsto \omega^3, \end{aligned} \tag{32a}$$

which we will also write as

$$\omega^i \mapsto B^i_j\omega^j. \tag{32b}$$

Lemma 3: Let $\gamma \in \Lambda_p^1$ satisfy

$$\gamma(T) = 0 = \gamma(T') \tag{33}$$

and let $F,F' \in \Lambda^+$ be determined by Lemma 2, so that

$$F(T) = \gamma = F'(T'). \tag{34}$$

Then (for an appropriate, γ-independent choice of c,d),

$$F' \equiv B^tFB, \tag{35a}$$

i.e.,

$$F'_{ab} \equiv F_{mn}B^m{}_aB^n{}_b. \tag{35b}$$

Proof: One has immediately that

$$B^tFB(T') = B^tF(T) = B^t\gamma = \gamma = F'(T). \tag{36}$$

In order to invoke Lemma 2 to conclude that (35) holds by uniqueness we must show that c,d can be chosen, so that B^tFB is in Λ^+.

But since

$$\gamma \equiv \gamma_2\omega^2 + \gamma_3\omega^3, \tag{37a}$$

we see that

$$F = \gamma_2F_2 + \gamma_3F_3, \quad F' = \gamma_2F_2' + \gamma_3F_3', \tag{37b}$$

so that it is enough to show that B^tF_2B and B^tF_3B are in Λ^+. Direct calculation using (29), (34), and (37) shows that the first of these reduces to a linear equation involving d only, while the second determines c. We note in passing that (32) is *not* the only linear transformation that satisfies (35). Q.E.D.

Lemma 4: $h_\#$ is well-defined.

Proof: Assume, as above, that T and T' are linearly independent and let F_α', F_β', and F_η' be the unique elements of Λ^+ determined by Lemma 2 which satisfy $F_\alpha'(T') = \alpha$, $F_\beta'(T') = \beta, F_\eta'(T') = \eta$. Then

$$\begin{aligned} \langle F_\alpha', F_\beta' \rangle &= \epsilon^{abcd}F_\alpha'{}_{ab}F_\beta'{}_{cd} \\ &= \epsilon^{abcd}F_{\alpha mn}F_{\beta pq}B^m{}_aB^n{}_bB^p{}_cB^q{}_d \\ &\equiv \epsilon^{mnpq}F_{\alpha mn}F_{\beta pq}\,(\det B)\,f \\ &= \langle F_\alpha, F_\beta \rangle (\det B)\,f, \end{aligned}$$

where f is a constant that depends on the volume element. Therefore the two factors of $(f \det B)$ in the primed version of (26) cancel so (26) is independent of the choice of T.

Theorem 2: Let $*$ be the Hodge dual defined by the metric g_{ab}. Then h_* and g are conformally related.

Proof: For $F = *F$, $G = *G$, we have

$$\begin{aligned} \langle F,G \rangle &:= \epsilon^{abcd}F_{ab}G_{cd} \\ &= 2F^{ab}(*G)_{ab} = 2F^{ab}G_{ab}. \end{aligned} \tag{38}$$

But for any $T \in T_pM$, we have

$$\begin{aligned} F_{am}T^mG^{an}T_n &= \tfrac{1}{4}\epsilon_{ampq}F^{pq}T^m\epsilon^{anrs}G_{rs}T_n \\ &= \tfrac{3}{2}\delta_{[m}{}^n\delta_p{}^r\delta_{q]}{}^sF^{pq}G_{rs}T^mT_n \\ &= \tfrac{1}{2}F^{pq}G_{pq}T^mT_m + F^{pq}G_{qm}T^mT_p, \end{aligned} \tag{39}$$

so that

$$4F_{am}T^mG^{an}T_n \equiv F^{pq}G_{pq}T^mT_m, \tag{40a}$$

or in other words,

$$g(F(T),G(T)) = \langle F,G \rangle (g(T,T)/8), \tag{40b}$$

so that

$$g(F(T),G(T))/g(H(T),H(T)) = \langle F,G \rangle / \langle H,H \rangle. \tag{41}$$

Comparison with (26) shows that g is in the same conformal class as h_*. Q.E.D.

This shows that our definition (26) reproduces the given metric from its Hodge dual. We now show the converse.

Theorem 3: Let $h_\#$ be defined by (26) and denote its Hodge dual by $*$. Then $* = \#$.

Proof: Choose an orthonormal (with respect to $h_\#$) basis ω^α of Λ_p^1 satisfying

$$\omega^0 \wedge \omega^1 \wedge \omega^2 \wedge \omega^3 = \epsilon. \tag{42}$$

Let F^i, $i = 1, 2, 3$, be the self-dual (with respect to $\#$) two-forms defined by $F^i(X_0) = \omega^i$, where X_α is the basis of T_pM dual to ω^α. Then, e.g.,

$$F^1 = \omega^0 \wedge \omega^1 + a\omega^2 \wedge \omega^3 + b\omega^1 \wedge \omega^2 + c\omega^3 \wedge \omega^1, \tag{43}$$

with $a > 0, b, c$, to be determined. But using the definition (26), we have

$$\langle F^i,\ F^j \rangle = \delta^{ij}\langle F^1,\ F^1 \rangle, \tag{44}$$

which implies

$$F^2 = \omega^0 \wedge \omega^2 + a\omega^3 \wedge \omega^1 + c\omega^2 \wedge \omega^3 + d\omega^1 \wedge \omega^2,$$
$$F^3 = \omega^0 \wedge \omega^3 + a\omega^1 \wedge \omega^2 + d\omega^3 \wedge \omega^1 + b\omega^2 \wedge \omega^3.$$

Repeating this procedure for X_1, e.g., constructing $G^0(X_1) = \omega^0$, etc., and using the fact that the F^i form a basis for self-dual two-forms, yields a set of linear equations that can be solved to give

$$b = c = 0. \tag{45}$$

Finally, using X_2 gives

$$d = 0, \quad a^2 = 1, \tag{46}$$

so that

$$\begin{aligned} F^1 &= \omega^0 \wedge \omega^1 + \omega^2 \wedge \omega^3, \\ F^2 &= \omega^0 \wedge \omega^2 + \omega^3 \wedge \omega^1, \\ F^3 &= \omega^0 \wedge \omega^3 + \omega^1 \wedge \omega^2. \end{aligned} \tag{47}$$

But this is just the standard basis for self-dual two-forms with respect to $*$! Q.E.D.

IV. DISCUSSION

For Euclidean signature, let $\mathscr{M}$ denote the manifold of classes of conformal metrics at a point $p \in M$ and $\mathscr{H}$ denote the manifold of dual operators on Λ_p^2. We have the following situation:

$$\mathscr{M} \underset{B}{\overset{A}{\rightleftarrows}} \mathscr{H},$$

where A takes a metric to its Hodge dual, and B is given by (26). Theorem 2 says that $B \circ A = I$, while Theorem 3 says that $A \circ B = I$. Thus both A and B are one–one and onto, and are therefore isomorphisms. The manifold $\mathscr{M}$ is nine-dimensional (10 metric components $-$ 1 constraint), and the manifold $\mathscr{H} \simeq \mathrm{SO}(3,3)/[\mathrm{SO}(3) + \mathrm{SO}(3)]$, so $\dim \mathscr{H} = 15 - 6 = 9$.

All of our results have been obtained at a point $p \in M$. Suitably smooth metric tensors and dual operators are obtained by working throughout with suitably smooth tensor fields in a neighborhood of p.

We believe that a result similar to Theorems 2 and 3 holds for other signatures. However, our attempts to modify the argument in Sec. III have so far failed, primarily because of the failure of Lemma 2 if T is null. In the Lorentzian case, one can define $\alpha \in \Lambda_p^1$ to be null if there exists a (real) two-form F and a vector T such that

$$F(T) = \alpha, \quad \#F(T) = 0, \quad F \wedge F = 0 = F \wedge \#F. \tag{48}$$

Although this definition is correct if $\#$ is the Hodge dual of a Lorentzian metric, we have been unable to use it to actually construct a conformal metric.

ACKNOWLEDGMENTS

TD gratefully acknowledges an Indo-American fellowship, funded jointly by the NSF and the USIA in the United States and by the UGC in India.

[1] R. Penrose, "The nonlinear graviton," Gen. Relativ. Gravit. **7**, 171 (1976); "Nonlinear gravitons and curved twistor theory" Gen. Relativ. Gravit. **7**, 31 (1976).
[2] M. F. Atiyah, N. J. Hitchin, and I. M. Singer, "Self-duality in four-dimensional Riemannian geometry," Proc. R. Soc. London Ser. A **362**, 425 (1978).
[3] A. Ashtekar, "A note on helicity and self-duality," J. Math. Phys. **27**, 824 (1986).
[4] A. Ashtekar, *New Perspectives in Canonical Gravity* (Bibliopolis, Naples, 1988).

Magnetic generalizations of the Carter metrics

Alberto García Díaz and Nora Bretón Baez
Centro de Investigación y de Estudios Avanzados del IPN, Departamento de Física, Apartado Postal 14-740, 07000 México D. F., México

(Received 20 February 1987; accepted for publication 14 December 1988)

Magnetic generalizations of all Carter metrics—Carter A, $B(+)$, $B(-)$, and D solutions—are obtained by applying a Harrison transformation to the Ernst potentials determined from a linear combination of two Killing vectors. The most general magnetized Carter-A metric contains, among other solutions, the magnetized Kerr–Newman metrics and, modulo limiting transitions, the magnetized branches $B(+)$, $B(-)$, and D.

I. INTRODUCTION

The main purpose of this work is to give the explicit expressions of the metric and electromagnetic field for a general magnetic generalization of the Carter-A solution,[1] C-A for short. Since the seed C-A metric represents the axisymmetric gravitational field of rotating charged masses and the magnetizing process used (Harrison transformation[2]) incorporates into the seed solution two electromagnetic field parameters and two kinematical constants (corresponding to the choice of the Killing vectors), the magnetized Carter-A metric obtained, MC-A for short, contains (besides the parameters of the C-A metric) four additional parameters.

The MC-A metric contains as subbranches of solutions the magnetized Kerr–Newman metrices (MKN), corresponding to magnetic generalizations of the seed KN solution for the both choices of the Killing vectors ∂_ϕ and ∂_t.

The magnetizing procedure applied to the Carter-$B(\pm)$ metrices (Ref. 1) yield the magnetized Carter-$B(\pm)$ solutions, MC-$B(\pm)$ for short (see Secs. III and IV). Since the MC-$B(\pm)$ metrics are interrelated by complex transformations [the same fact takes place in the case of "proper" C-$B(\pm)$ solutions], one can consider the MC-B metric as a complexified space-time and the real MC-$B(+)$ and MC-$B(-)$ solutions as two different real cuts of the same complex structure. The MC-$B(\pm)$ metrics contain as particular solutions the magnetized Reissner–Nordström metrics, which we shall denote by MRN$(\pm)$.

Section V deals with the determination of the magnetized Carter-D metric, which in all respects is equivalent to the magnetized Bertotti–Robinson metric.[3] We propose to call this class of solutions the magnetized Bertotti–Robinson metric and denote it by MBR.

In Sec. VI we present limiting transitions of the MC-A metric to the MC-$B(\pm)$ solutions. From the MC-$B(\pm)$ metrics, for vanishing parameter l, one arrives at the MRN$(\pm)$ solutions. By limiting transitions of the MRN$(\pm)$ metrics one obtains the MBR solution, which degenerates into a magnetic flat space-time class of solutions (MF) containing as a particular case the Melvin magnetic universe.[4] Hence the obtained solutions satisfy the following scheme of transitions:

$$\text{MC-}A \rightarrow \left.\begin{matrix}\text{MC-}B(+)\rightarrow\text{MRN}(+)\\ \text{MC-}B(-)\rightarrow\text{MRN}(-)\end{matrix}\right\}\rightarrow\text{MBR}\rightarrow\text{MF}.$$

It should be pointed out that we are dealing with metric structures without the cosmological constant, i.e., with magnetized Carter metrics (and their limiting contractions) with vanishing λ.

II. MAGNETIZED CARTER-A METRIC

We shall briefly give the steps followed to generate the MC-A solution starting from the seed C-A metric. Let the C-A solution be given in the form[5]

$$g=\frac{\Delta}{P}dp^2+\frac{P}{\Delta}(d\tau+q^2\,d\sigma)^2+\frac{\Delta}{Q}dq^2 - \frac{Q}{\Delta}(d\tau-p^2\,d\sigma)^2, \tag{2.1}$$

with the electromagnetic field described by the two-form

$$\omega=\tfrac{1}{2}(f_{\mu\nu}+\check{f}_{\mu\nu})dx^\mu\wedge dx^\nu = -d[(e+ig)/(q+ip)](d\tau-ipq\,d\sigma). \tag{2.2}$$

The structural functions P, Q, and Δ are

$$\begin{aligned} P&=\gamma-\nu+2np-\kappa p^2,\\ Q&=\gamma+\nu-2mq+\kappa q^2,\\ \Delta&=p^2+q^2,\\ 2\nu&:=e^2+g^2, \end{aligned} \tag{2.3}$$

where m is the mass, n represents the Newman–Unti–Tamburino (NUT) parameter (magnetic mass), e and g are the electric and magnetic charges, γ is related to the rotation (Kerr) parameter, and κ is a free parameter which can be considered to assume the discrete values $\{1, 0, -1\}$.

Choosing the Killing vector as a linear combination of the Killing directions ∂_τ and ∂_σ according to

$$K^\mu=\alpha\delta^\mu_\tau+\beta\delta^\mu_\sigma, \tag{2.4}$$

one brings the metric (2.1) to the form

$$g=f^{-1}\{f[(\Delta/P)dp^2+(\Delta/Q)dq^2] + PQ(\beta\,d\tau-\alpha\,d\sigma)^2\}-f\{-(\alpha\,d\tau+\beta\,d\sigma)/(\alpha^2+\beta^2)+W_i(\beta\,d\tau-\alpha\,d\sigma)\}^2, \tag{2.5}$$

where

 0022-2488/89/061310-06$02.50

$$-f = K_\mu K^\mu = \Delta^{-1}\{P(\alpha+\beta q^2)^2 - Q(\alpha-\beta p^2)^2\} =: \Delta^{-1}D, \tag{2.6}$$

and

$$W_i = -[D^{-1}/(\alpha^2+\beta^2)]\{(\alpha+\beta q^2)(\beta-\alpha q^2)P - (\alpha-\beta p^2)(\beta+\alpha p^2)Q\}. \tag{2.7}$$

The subindex i in W stands for "initial."

The Ernst potentials (ϕ and $\mathscr{E}$) of the metric (2.5) can be evaluated from the relations

$$\begin{aligned} d\phi &= -iK \lrcorner \omega, \\ d\mathscr{E} &= iK(dK + *\,dK) - 2\bar{\phi}\, d\phi, \end{aligned} \tag{2.8}$$

where $K = K_\mu dx^\mu$, $\lrcorner$ denotes the step product, and $*$ is Hodge's star operation (see details in Ref. 6).

The Ernst potentials for the studied metric is

$$\begin{aligned} \phi &= -[(e+ig)/(q+ip)](\alpha - i\beta pq) + \phi_0, \\ \phi_0 &= F + iG, \end{aligned} \tag{2.9}$$

and

$$\begin{aligned} \mathscr{E} = {} & f - 2\nu\Delta^{-1}(\alpha^2+\beta^2p^2q^2) \\ & + 2i\{mp[\Delta^{-1}(\alpha-\beta p^2)^2 + \beta(3\alpha-\beta p^2)] \\ & - nq[\Delta^{-1}(\alpha+\beta q^2)^2 - \beta(3\alpha+\beta q^2)] \\ & - \alpha\beta\kappa pq + \beta^2\gamma pq - \beta\nu\Delta^{-1}pq\,[2\alpha+\beta(q^2-p^2)]\} \\ & + 2\bar{\phi}_0[(e+ig)/(q+ip)](\alpha - i\beta pq) + \mathscr{E}_0, \end{aligned} \tag{2.10}$$

where F, G are real constants and $\mathscr{E}_0$ is an arbitrary complex constant.

To magnetize a given metric one applies the Harrison transformation to the Ernst potentials according to the rules

$$\begin{aligned} &\tilde{\mathscr{E}} = \mathscr{E}\psi^{-1}, \quad \tilde{\phi} = \psi^{-1}(\phi + (E+iB)\mathscr{E}), \\ &\psi = 1 - 2(E-iB)\phi - \delta\mathscr{E}, \quad \delta := E^2+B^2, \\ &E = \text{const}, \quad B = \text{const} \end{aligned} \tag{2.11}$$

(the tilde is used to denote the new quantities).

The new f function becomes

$$\tilde{f} = |\psi|^{-2} f, \tag{2.12}$$

while the new W function ought to fulfill the equation

$$d\tilde{W} = \psi\bar{\psi}\, dW_i - if^{-1}\{(\psi\bar{\psi}_{,p} - \bar{\psi}\psi_{,p})P\, dq - (\psi\bar{\psi}_{,q} - \bar{\psi}\psi_{,q})Q\, dp\}. \tag{2.13}$$

The generated magnetized Carter-A metric, MC-A metric for short, amounts to

$$g = |\psi|^2\left\{\frac{\Delta}{P}dp^2 + \frac{\Delta}{Q}dq^2 + f^{-1}PQ(\beta\, d\tau - \alpha\, d\sigma)^2\right\} - |\psi|^{-2}f\left\{-\frac{\alpha\, d\tau + \beta\, d\sigma}{\alpha^2+\beta^2} + \tilde{W}(\beta\, d\tau - \alpha\, d\sigma)\right\}^2, \tag{2.14}$$

with the electromagnetic field given by the two-form

$$\begin{aligned} -\omega = {} & d\tilde{\phi} \wedge \left[-\frac{\alpha\, d\tau + \beta\, d\sigma}{\alpha^2+\beta^2} + \tilde{W}(\beta\, d\tau - \alpha\, d\sigma)\right] \\ & + *\left\{d\tilde{\phi} \wedge \left[-\frac{\alpha\, d\tau + \beta\, d\sigma}{\alpha^2+\beta^2} + \tilde{W}(\beta\, d\tau - a\, d\sigma)\right]\right\}. \end{aligned} \tag{2.15}$$

The structural functions P, Q, and Δ are given by formulas (2.3), the function f is defined as in formula (2.6). The complex factor function ψ is given, in terms of the Ernst potentials ϕ and $\mathscr{E}$ from (2.9) and (2.10), according to formula (2.11).

The $\tilde{W}$ function, which from now on we shall denote as $W(E,B\,|\alpha,\beta)$, is given by

$$W(E,B\,|\alpha,\beta) = W(0,0|\alpha,\beta) + \Omega(E,B\,|\alpha,\beta) + \Omega(\phi_0\cdot\mathscr{E}_0) + W_0, \tag{2.16}$$

where

$$W(0,0|\alpha,\beta) = W_i \text{ [from (2.7)]}, \quad W_0 = \text{const}, \tag{2.17}$$

the function $\Omega(E,B\,|\alpha,\beta)$ amounts to

$$\begin{aligned} D\Omega(E,B\,|\alpha,\beta) = {} & 4E(+)(a+\beta q^2)qP + 4E(-)(\alpha-\beta p^2)pQ \\ & + 6\delta\nu[(\alpha^2-\beta^2q^4)P + (\alpha^2-\beta^2p^4)Q] \\ & + 4\delta E(+)\mathscr{M} + 4\delta E(-)\mathscr{N} \\ & + \delta^2\{\mathscr{A}P + \mathscr{B}Q + \mathscr{H}PQ\}, \end{aligned} \tag{2.18}$$

where the polynominals $\mathscr{A}$, $\mathscr{B}$, $\mathscr{H}$, $\mathscr{M}$, and $\mathscr{N}$ are

$$\begin{aligned} \mathscr{A} = {} & \alpha^4(2m^2+2n^2-4m\kappa q+\kappa^2q^2) - \alpha^3\beta q^2(8n^2+4\kappa\gamma+\kappa^2q^2) + 2\alpha^2\beta^2q^2[(\gamma-\nu)(2(\gamma+\nu)-6mq \\ & + \kappa q^2) - (m^2+3n^2)q^2] + \alpha\beta^3q^4[2(\gamma-\nu)\nu + 4n^2q^2 - (\gamma-\nu)(6m-\kappa q)q] - \beta^4(\gamma-\nu)^2q^6, \\ \mathscr{B} = {} & \alpha^4(2m^2+2n^2-4n\kappa p+\kappa^2p^2) + \alpha^3\beta p^2(8m^2-4\kappa\gamma+\kappa^2p^2) + 2\alpha^2\beta^2p^2[(\gamma+\nu)(2(\gamma-\nu)+6np-\kappa p^2) \\ & - (n^2+3m^2)p^2] + \alpha\beta^3p^4[2(\gamma+\nu)\nu - 4m^2p^2 - (\gamma+\nu)(6n-\kappa p)p] - \beta^4(\gamma+\nu)^2p^6, \\ \mathscr{H} = {} & 8\alpha^3\beta(mq+np) + 12\alpha^2\beta^2(nq-mp)pq + 3\alpha\beta^3pq[q(2pQ-4\kappa pq^2-2nq^2+3\kappa p^3) - p(2qP+4\kappa qp^2 \\ & + 2mp^2-3\kappa q^3)] - \beta^4p^2q^2[3(\gamma+\nu)p^2 + 3(\gamma-\nu)q^2 - 2mqp^2 + 2npq^2], \\ \mathscr{M} = {} & (\alpha+\beta q^2)[\alpha m(\alpha-\beta q^2) - \beta^2(\gamma-\nu)q^3 - \alpha^2\kappa q]P + \alpha m(\alpha-\beta p^2)(\alpha+3\beta p^2)Q \\ & + \beta q[2\alpha(\alpha-\beta p^2) - \beta p^2(\alpha+\beta q^2)]PQ, \\ \mathscr{N} = {} & (\alpha-\beta p^2)[\alpha n(\alpha+\beta p^2) + \beta^2(\gamma+\nu)p^3 - \alpha^2\kappa p]Q + \alpha n(\alpha+\beta q^2)(\alpha-3\beta q^2)P \\ & + \beta p[2\alpha(\alpha+\beta q^2) + \beta q^2(\alpha-\beta p^2)]PQ, \end{aligned} \tag{2.19}$$

the symbols $E(+)$ and $E(-)$ in the expression above are constants defined as

$$E(+) := Ee + Bg, \quad E(-) := Eg - Be. \tag{2.20}$$

The function $\Omega(\phi_0, \mathscr{E}_0)$, with $\phi_0 = F + iG = \text{const}$ and $\mathscr{E}_0 = -(F^2 + G^2) - i\epsilon = \text{const}$, which represents the contributions to $W(E, B\,|\alpha, \beta)$ of the additive constants in the Ernst potentials, can be given as

$$\begin{aligned} D\Omega(\phi_0, \mathscr{E}_0) = & -\{4(EF+BG) - 6\delta(F^2+G^2) + 4\delta[(EF+BG)(F^2+G^2) + (EG-BF)\epsilon] \\ & - \delta^2[\epsilon^2 + (F^2+G^2)^2]\}DW(0,0|\alpha,\beta) - 12\delta[(Fe+Gg)(\alpha+\beta q^2)qP + (Fg-Ge)(\alpha-\beta p^2)pQ] \\ & + 4\delta\{B(F^2+G^2) - E\epsilon - 4G(EF+BG) + \delta[G(F^2+G^2) + F\epsilon]\}[e(\alpha-\beta p^2)pQ - g(\alpha+\beta q^2)qP] \\ & - 4\delta\{E(F^2+G^2) + B\epsilon - 4F(EF+BG) + \delta[F(F^2+G^2) - G\epsilon]\}[g(\alpha-\beta p^2)pQ \\ & + e(\alpha+\beta q^2)qP] + 4\delta[2(EG-BF) - \delta\epsilon][m(\alpha^2-\beta^2p^4)pQ - n(\alpha^2-\beta^2q^4)qP - \beta^2\Delta pqPQ] \\ & - 6\delta\nu[2(EF+BG) - \delta(F^2+G^2)][(\alpha^2-\beta^2q^4)P + (\alpha^2-\beta^2p^4)Q] \\ & - 4\delta^2(Fe+Gg)\mathscr{M} + 4\delta^2(Ge-Fg)\mathscr{N}. \end{aligned} \tag{2.21}$$

The function $\Omega(\phi_0, \mathscr{E}_0)$, included here for completeness of the obtained solution, is useful when one accomplishes limiting transitions to derive other subclasses of solutions (see Sec. V).

The magnetized Carter-A metric, given by Eqs. (2.18)–(2.21), is endowed with the following set of parameters:

$$\mathscr{C} := \{m, n, \kappa, \gamma, e, g, E, B, \alpha, \beta\}, \tag{2.22}$$

where m is the mass, n is the NUT parameter, κ and γ are related to the rotation parameter, e and g correspond to the electric and magnetic charges, E and B are the "external" field parameters, and α and β are the selecting Killing vectors parameters.

By assigning to these parameters some particular values one obtains subbranches of solutions of the MC-A metric. For instance, choosing

$$\mathscr{C} = \{m, 0, 1, \tfrac{1}{2}e^2 + a^2, -e, 0, E, B, \beta' a - \alpha', a^{-1}\beta'\}, \tag{2.23}$$

and accomplishing in the MC-A metric (2.18)–(2.21) the coordinate transformations

$$P = a\cos\theta, \quad q = r, \quad \tau = a\phi - \tau, \quad \sigma = a^{-1}\phi, \quad a = \text{const}, \tag{2.24}$$

one arrives just at the most general magnetic generalization of the Kerr–Newman metric (MK-N), which in particular contains the magnetized Kerr–Newman metric obtainable from the Killing vector ∂_ϕ,[7,8] and also class of magnetized Reissner–Nordström metrics (MRN); at least three generalizations for two out of the four Killing vectors which the RN solution possesses.

Other choices of $\mathscr{C}$ will give rise to different subclasses of magnetized metrics, among them, those corresponding to the seed vacuum metrics.

III. MAGNETIZED CARTER-$B(+)$ METRIC

Starting from the seed Carter-$B(+)$ metric given in the coordinate chart $\{p, q, \tau, \sigma\}$,[5] and applying the magnetizing procedure outlined in the previous section, one arrives at the magnetized Carter-$B(+)$ solution [MC-$B(+)$] determined by the metric line element

$$\begin{aligned} g = & |\psi|^2\{(\Delta/P)dp^2 + (\Delta/Q)dq^2 \\ & + f^{-1}PQ(\beta\,d\tau - \alpha\,d\sigma)^2\} \\ & - |\psi|^{-2}f\{-[1/(\alpha^2+\beta^2)](\alpha\,d\tau + \beta\,d\sigma) \\ & + W(\beta\,d\tau - \alpha\,d\sigma)\}^2, \end{aligned} \tag{3.1}$$

and the electromagnetic two-form

$$\begin{aligned} -\omega = & d\tilde{\phi}[-(1/(\alpha^2+\beta^2))(\alpha\,d\tau + \beta\,d\sigma) \\ & + W(\beta\,d\tau - \alpha\,d\sigma)] \\ & + *\{d\tilde{\phi}[-(1/(\alpha^2+\beta^2))(\alpha\,d\tau + \beta\,d\sigma) \\ & + W(\beta\,d\tau - \alpha\,d\sigma)]\}. \end{aligned} \tag{3.2}$$

Notice that the metric g and the electromagnetic field ω for the MC-A solution are in exactly the same form as for the MC-$B(+)$ solution.

The structural functions P, Q, and Δ for the studied MC-$B(+)$ metric are

$$\begin{aligned} P &= a + 2bp - \kappa p^2, \\ Q &= 2\nu - l^2 - 2mq + \kappa q^2, \\ \Delta &= q^2 + l^2, \\ 2\nu &= e^2 + g^2, \end{aligned} \tag{3.3}$$

where m represents the mass, l is the NUT parameter, e and g are, correspondingly, the electric and magnetic charges, a and b are "kinematical" parameters related to the choice of the coordinate chart (by shifting the coordinate p one can cancel out the parameter b), and κ is a free parameter, which can be considered to assume the discrete values $(1, 0, -1)$.

The function f amounts to

$$-f = \Delta^{-1}\{\beta^2(q^2+l^2)^2P - (\alpha - 2\beta lp)^2Q\} =: D/\Delta, \tag{3.4}$$

where α and β are arbitrary parameters related with the Killing vector defined by Eq. (2.4), $K^\mu = \alpha\delta^\mu_\tau + \beta\delta^\mu_\delta$.

The Ernst potentials $\mathscr{E}$ and ϕ, which determine the factor function ψ and the new potentials $\tilde{\mathscr{E}}$ and $\tilde{\phi}$ [see Eq. (2.11)], are given by

$$\phi = -[(e+ig)/(q+il)][\alpha - \beta p(l+iq)] \tag{3.5}$$

and

$$\mathscr{E} = f - 2\nu\Delta^{-1}[\alpha(\alpha - 2\beta lp) + \beta^2\Delta p^2] + 2i\{m[\Delta^{-1}l(\alpha - 2l\beta p)^2 + 3\beta p(\alpha - \beta lp)] - \kappa q[\alpha^2 l + \beta p(a - \beta lp)(q^2 - 3l^2)]\Delta^{-1} + \beta q(\alpha b + a\beta l) - 2\nu\beta pq(\alpha - 2\beta lp)\}. \quad (3.6)$$

For simplicity, we have omitted here the additive constants ϕ_0 and $\mathscr{E}_0$; one can include them, remembering that if $\phi \to \phi(3.5) + \phi_0$, then $\mathscr{E} \to \mathscr{E}(3.6) - 2\bar{\phi}_0\phi(3.5) + \mathscr{E}_0$, $\mathscr{E}_0 = -\phi_0\bar{\phi}_0 - i\epsilon^0$.

According to Eq. (2.13), the function W amounts to

$$W(E,B\,|\alpha,\beta) = W(0,0|\alpha,\beta) + \Omega(E,B\,|\alpha,\beta) + W_0, \quad (3.7)$$

where

$$W(0,0|\alpha,\beta) = (D^{-1}/(\alpha^2+\beta^2))[\alpha\beta(q^2+l^2)^2P + (a - 2\beta lp)(\beta + 2\alpha lp)Q], \quad (3.8)$$

$W_0 = \text{const}$,

and

$$\begin{aligned} D\Omega(E,B\,|\alpha,\beta) &= 4\beta E(+)q\Delta P + 2E(-)[\beta lP + (\alpha - 2\beta lp)pQ] + 12\delta\nu\beta[\alpha\Delta P + (\alpha - 2\beta lp)p^2Q] \\ &+ 4\delta\beta E(+)\{mp^2(a - 2\beta lp)(2a - \beta lp)p^2Q + \Delta[\alpha^2(2m - \kappa q) + 2\alpha\beta blq + a\beta^2 q(l^2 - q^2)]P \\ &+ q[2(\alpha - 2\beta lp)(\alpha - \beta lp) - \beta^2p^2\Delta]PQ\} + 4\delta E(-)\{(\alpha - 2\beta lp)[2\nu\beta^2p^2 \\ &- \kappa(\alpha - 2\beta lp)(\alpha - \beta lp)]pQ + \beta\Delta[\alpha^2\kappa l - 2\beta(\alpha b + a\beta l)q^2]P \\ &+ \beta[(\alpha - 2\beta lp)^2 + 3\Delta(\alpha - \beta lp)p]PQ\} + \delta^2\{\mathscr{A}P + \mathscr{B}Q + \mathscr{H}PQ\}, \end{aligned}$$

$$\mathscr{A} = \beta\Delta\{\alpha^3[\kappa^2l^2 + (2m - \kappa q)^2] + 8\alpha\beta l(\alpha b + a\beta l)(m - \kappa q)q + 4\Delta[b\beta^2(\alpha b + a\beta l)q - a\alpha\beta^2 m]q\},$$

$$\begin{aligned} \mathscr{B} &= 2(a - 2\beta lp)\{18\beta m^2p^2(\alpha - \beta lp)^2 + 2\nu^2\beta^3p^4 - \kappa^2lp(\alpha - 2\beta lp)^3 - 2\beta p(a - 2\beta lp)[\alpha(\kappa\nu + \kappa^2l^2 + 6m^2)p \\ &+ \beta l(2\nu - \kappa l^2)(a + bp) - \beta lp^2(4m^2 + 2\nu\kappa + \kappa^2l^2)]\}, \end{aligned} \quad (3.9)$$

$$\begin{aligned} \mathscr{H} &= \beta(\alpha - 2\beta lp)\{4\nu\alpha(\alpha - 2\beta lp) - 2(\alpha - 2\beta lp)^2Q + \beta\Delta^2P + 4q[\alpha^2m + \alpha\beta l(bq - \kappa pq + 2mp) \\ &+ \beta^2l^2(aq + \kappa p^2q - 2mp^2)] + 4\beta^2p\Delta\{\alpha^2\kappa l + \alpha\beta(3\nu p - 2\kappa l^2p - 3mpq - 2bq^2 + \kappa pq^2) \\ &+ \beta^2l[a(l^2 - q^2) + bp\Delta + p^2(\kappa l^2 - 4\nu + 2mq - \kappa q^2)]\}, \end{aligned}$$

where

$$E(+) = Ee + Bg \quad \text{and} \quad E(-) = Eg - Be. \quad (3.10)$$

The magnetized MC-$B(+)$ metric is endowed with the set of parameters

$$\mathscr{C} = \{m,l,\kappa,a,b,e,g,E,B,\alpha,\beta\}. \quad (3.11)$$

Without any loss of generality the parameter b can be assumed to be zero, and κ can be assumed to be equal to the discrete values $\{1,0,-1\}$.

Since the "seed" Carter-$B(+)$ metric contains as a special case for $l = 0$ the Reissner–Nordström metric RN$(+)$, the MC-$B(+)$ solution described above contains, for l equal to zero, the magnetized RN$(+)$ metric, MRN$(+)$ for short. In particular, for

$$\mathscr{C} = \{m,0,1,1,0,e,0,0,B,0,1\}, \quad (3.12)$$

one obtains a solution due to Ernst.[9]

IV. MAGNETIZED CARTER-$B(-)$ METRIC

By applying the previously used magnetizing process to the Carter-$B(-)$ metric, one generates the magnetized Carter-$B(-)$ solution [MC-$B(-)$]. However, one arrives at the same result if one executes in the MC-$B(+)$ metric—understood as a complex solution of the complex Einstein–Maxwell equation—complex coordinate transformations and complex redefinitions of constants in such a manner that the resulting (real cross-section) solution will have the correct signature $(+\,+\,+\,-)$. Therefore the magnetized Carter-$B(\pm)$ metrics can be thought of as two different real slices of the same complex Einstein–Maxwell structure.

Executing in expressions (3.1)–(3.10), which determine the MC-$B(+)$ metric, the coordinate transformations

$$p \to q, \quad q \to p, \quad \tau \to i\tau, \quad \sigma \to i\sigma \quad (4.1)$$

and redefining the constants according to

$$\begin{aligned} &m \to -n, \quad l \to -l, \quad \kappa \to \kappa, \quad a \to a, \quad b \to b, \\ &e \to -ig, \quad g \to ie, \quad E \to E, \quad B \to B, \quad \alpha \to i\alpha, \quad \beta \to i\beta, \\ &[P \to Q, \quad Q \to P, \quad E(+) \to -iE(-), \\ &E(-) \to iE(+)], \end{aligned} \quad (4.2)$$

one obtains the magnetized Carter-$B(-)$ solution characterized by the set of constants

$$\mathscr{C} = \{n,l,\kappa,a,b,e,g,E,B,\alpha,\beta\}. \quad (4.3)$$

The explicit expression of the metric g, which can be very easily obtained according to the procedure outlined above, is not included in this text because of its length.

The MC-$B(-)$ metric contains several subfamilies of solutions to the Einstein–Maxwell equations, among them (for $l = 0$) the magnetic generalization of the anti-Reissner–Nordström solution [MRN$(-)$ solution].

V. THE MAGNETIZED CARTER-*D* OR BERTOTTI-ROBINSON METRIC

This class of solutions is determined by the metric g from (2.14) and the two-form ω from (2.15), with the following set of structural functions:

$$P = a + bp - 2\nu p^2, \quad Q = h + kq + 2\nu q^2, \quad 2\nu := e^2 + g^2,$$
$$\Delta = 1, \quad f = \alpha^2 Q - \beta^2 P \equiv D,$$
$$\psi = 1 - 2(E - iB)\Phi - \delta\mathscr{E}, \quad \delta := E^2 + B^2,$$
$$\phi = (e + ig)(\alpha q + i\beta p),$$
$$\mathscr{E} = f - 2\nu(\alpha^2 q^2 + \beta^2 p^2) + i\alpha\beta(bq + kp),$$
$$\tilde{W} \equiv W(E,B\,|\alpha,\beta) = W_i + D^{-1}\Omega(E,B) + W_0,$$
$$W_i \equiv W(0,0|\alpha,\beta) = \frac{\alpha\beta}{\alpha^2+\beta^2}\frac{(P+Q)}{D}, \quad W_0 = \text{const},$$
$$\begin{aligned}\Omega(E,B) = &-4E(+)\beta qP + 4E(-)\alpha pQ \\ &+ 12\delta\nu\alpha\beta(p^2Q - q^2P) + 4\delta\alpha E(-) \\ &\times[\beta^2(pPQ + bq^2P + 2\nu p^3Q) - \alpha^2 hpQ] \\ &+ 4\delta\beta E(+)[\alpha^2(qPQ + kp^2Q - 2\nu q^3P) \\ &- \beta^2 aqP] + \delta^2\alpha\beta\{\alpha^2(h + kq)^2P \\ &+ \beta^2(a + bp)^2Q + \alpha^2(k^2p^2 - 2bhp - ah)Q \\ &+ \beta^2(b^2q^2 - 2ak - ah)P\},\end{aligned}$$
$$E(+): Ee + Bg, \quad E(-) := Eg - Be, \tag{5.1}$$

where

$$\mathscr{C} = \{a,b,h,k,e,g,E,B,\alpha,\beta\}$$

are arbitrary constants. Without any loss of generality, the parameters b and k can be equated to zero.

This metric, for vanishing charges e and g, degenerates into a magnetic subclass of solutions—magnetized flat space-time metric (MF)—which contains as a particular case the Melvin magnetic universe. In fact, the metric (2.14), with structural functions from (5.1) specialized to the values

$$\mathscr{C} = \{a = 0, b = 4, h = 1, k = 0, E = 0, B = B_0/4, \alpha = 0, \beta = 1\},$$

are subjected to the coordinate transformations

$$p = \rho^2, \quad q = z, \quad \tau = T, \quad \sigma = \tfrac{1}{2}\phi,$$

which reduces to the Melvin metric.[4]

VI. LIMITING TRANSITIONS

A limiting transition of a given metric structure (g,ω) consists in taking its limit with respect to a contraction parameter ϵ when the coordinates are subjected to transformations depending on new coordinates and the parameter ϵ, $x^\mu = x^\mu(x'^\mu,\epsilon)$, and simultaneously the constants $\mathscr{C}$ (which characterize the studied structure) are replaced by functions of new constants and the contraction parameter ϵ, $\mathscr{C} = \mathscr{C}(\mathscr{C}',\epsilon)$. If there exists a finite limit, $\lim_{\epsilon\to 0}(g,\omega) = (g',\omega')$, this limit represents some solution. Notice that this process does not require finiteness of all $\lim_{\epsilon\to 0}\mathscr{C}(\epsilon)$ or $\lim_{\epsilon\to 0}x^\mu(x'^\mu,\epsilon)$.

The purpose of this section is to show that the MC-$B(\pm)$ solutions can be derived from the MC-A metric via corresponding limiting transitions. From the MC-$B(\pm)$ metrics, by letting the parameter l tend to zero, one arrives at the MRN$(\pm)$ solutions. Accomplishing limiting transitions in the MRN$(\pm)$ metrics one obtains the MBR solution (without the cosmological constant), which degenerates, by equating to zero the "proper" charges, into the MF metric. In this way, one establishes the limiting transition scheme given in Sec. I.

It should be noticed that all these solutions [MC-A, MC-$B(\pm)$, MRN$(\pm)$, MBR, and MF] to the Einstein–Maxwell equations arise as particular branches of solutions from the canonical metric structure

$$\begin{aligned}g = &|\psi|^2\{(\Delta/P)dp^2 + (\Delta/Q)dq^2 \\ &+ f^{-1}PQ(\beta\,d\tau - \alpha\,d\sigma)^2\} - |\psi|^{-2}f\chi^2, \\ -\omega = &\,d\tilde{\phi} \wedge \chi + *(d\tilde{\phi} \wedge \chi), \\ \chi := &-\frac{\alpha\,d\tau + \beta\,d\sigma}{\alpha^2+\beta^2} + W(E,B\,|\alpha,\beta)(\beta\,d\tau - \alpha\,d\sigma),\end{aligned} \tag{6.1}$$

for specific (to each branch) structural functions $\psi, \phi, \mathscr{E}, f, P, Q, \Delta$, and $W(E,B\,|\alpha,\beta)$.

In the general case with α and β different from zero, we found that it is more easy to accomplish limiting transitions from an alternative form of the metric structure (5.1) concerned with a new representation of the metric term χ. Representing the constants W_0 in $W(E,B\,|\alpha,\beta)$ as

$$W_0 \to W_0 + \frac{\alpha}{\beta}\frac{1}{\alpha^2+\beta^2} - \frac{\alpha}{\beta}\frac{1}{\alpha^2+\beta^2}, \tag{6.2}$$

the χ acquires the form

$$\chi = -\frac{d\sigma}{\beta} + W(E,B\,|\alpha,\beta)(\beta\,d\tau - \alpha\,d\sigma), \tag{6.3}$$

with $W(E,B\,|\alpha,\beta)$ differing from the previous one in the $W(0,0|\alpha,\beta)$ term; now this term should be replaced by

$$W(0,0|\alpha,\beta) \to W(0,0|\alpha,\beta) - \frac{\alpha}{\beta}\frac{1}{\alpha^2+\beta^2}. \tag{6.4}$$

For the MC-A metric, the $W(0,0|\alpha,\beta)$ function, modified according to (6.4), is given by

$$W(0,0|\alpha,\beta) = -(1/\beta D)[P(\alpha + \beta q^2) - Q(\alpha - \beta p^2)] \tag{6.5}$$

[with P and Q from (2.3) and D from (2.6)], while for the MC-$B(+)$ metric the modified $W(0,0|\alpha,\beta)$ is

$$W(0,0|\alpha,\beta) = (1/\beta D)(\alpha - 2\beta lp)Q \tag{6.6}$$

[with Q from (3.3) and D from (3.4)].

Accomplishing the MC-A metric in [with χ from (6.3), and $W(0,0|\alpha,\beta)$ from (6.5)] the coordinate transformations

$$p \to l + p, \quad q \to q, \quad \tau \to \tau + l^2\epsilon^{-1}\sigma, \quad \sigma \to \epsilon^{-1}\sigma, \tag{6.7}$$

accompanied by the redefinitions

$$\begin{aligned}&m \to m, \quad n \to \kappa l + b\epsilon, \quad \kappa \to \kappa, \quad \gamma \to \nu - \kappa l^2 - 2\epsilon bl + a\epsilon^2, \\ &e \to e, \quad g \to g, \quad E \to E, \quad B \to B, \quad E \to E + l^2\epsilon^{-1}\beta, \\ &\beta \to \beta\epsilon^{-1},\end{aligned} \tag{6.8}$$

and assigning to ϕ_0, ϵ, and W_0 the values

$$\begin{aligned}&\phi_0 = F + iG \to -i\beta l(e+ig)\epsilon^{-1},\\&\epsilon \to 2m\beta l(3\alpha + 2\beta l^2\epsilon^{-1})\epsilon^{-1},\\&W_0 \to W_0\epsilon^{-1},\end{aligned} \tag{6.9}$$

one observes that

$$\begin{aligned}\lim_{\epsilon\to 0}\{&\epsilon^{-2}P,Q,\Delta,D,f,\phi,\mathscr{E},\psi,\epsilon^{-1}W(E,B\,|\alpha,\beta)\}_{\text{MC-}A}\\&=\{P,Q,\Delta,D,f,\phi,\mathscr{E},\psi,W(E,B\,|\alpha,\beta)\}_{\text{MC-}B(+)},\end{aligned} \tag{6.10}$$

which implies that the MC-A metric in the limit $\epsilon\to 0$ reduces just to the MC-$B(+)$ metric.

In a similar manner one obtains the MC-$B(-)$ solution as limiting transition of the MC-A metric. The explicit form of the coordinate transformations and redefinitions of the constants can be derived by executing the formal complex transformations quoted in Sec. IV.

As it was mentioned in Sec. III by canceling out the parameter l in the MC-$B(+)$ solution one arrives at the MRN$(+)$ metric, which now we shall use in its formulation with χ from (6.3), $W(0,0|\alpha,\beta)$ given by

$$W(0,0|\alpha,\beta) = (\alpha/\beta)(Q/D), \tag{6.11}$$

and the parameter b equated to zero. Now, let us subject the coordinate in MRN$(+)$ to the transformations

$$p\to p,\quad q\to 1+\epsilon q,\quad \tau\to\epsilon^{-1}\tau,\quad \sigma\to\sigma, \tag{6.12}$$

accompanied by

$$\begin{aligned}&a\to a,\quad m\to 2\nu-\epsilon^2(h/2),\quad \kappa\to 2\nu,\quad e\to e,\quad g\to g,\\&E\to E,\quad B\to B,\quad \alpha\to\epsilon^{-1}\alpha,\quad \beta\to\beta,\\&F\to e\alpha\epsilon^{-2},\quad G\to\alpha g\epsilon^{-2},\quad \epsilon\to 0.\end{aligned} \tag{6.13}$$

In the limit $\epsilon\to 0$ one has

$$\begin{aligned}\lim_{\epsilon\to 0}\{&P,Q\epsilon^{-2},\Delta,D,f,\psi,\mathscr{E},\phi,\epsilon^{-1},W(E,B\,|\alpha,\beta)\}_{\text{MRN}(+)}\\&=\{P,Q,\Delta,D,f,\psi,\mathscr{E},\phi,W(E,B\,|\alpha,\beta)\}_{\text{MBR}},\end{aligned} \tag{6.14}$$

and consequently the MRN$(+)$ metric becomes the MBR solution.

Starting from the MNR$(-)$ metric, by a similar transition process, one obtains the MBR metric.

ACKNOWLEDGMENTS

We acknowledge the Instituto de Investigaciones en Matemáticas Aplicadas y en Sistemas de la UNAM for allowing us to use its computer system.

This work was partially supported by COSNET-SEP.

[1] B. Carter, Commun. Math. Phys. **10**, 280 (1968).
[2] B. K. Harrison, J. Math. Phys. (NY) **9**, 1744 (1968).
[3] N. Bretón B. and A. García D., Soc. Mex. Fis. **7**, 222 (1984).
[4] M. A. Melvin, Phys. Lett. **8**, 65 (1964).
[5] J. F. Plebański, Ann. Phys. (NY) **90**, 196 (1975).
[6] F. J. Ernst and J. F. Plebański, Ann. Phys. (NY) **107**, 226 (1977).
[7] F. J. Ernst and W. J. Wild, J. Math. Phys. **17**, 182 (1976).
[8] A. García D., J. Math. Phys. **26**, 155 (1985).
[9] F. J. Ernst, J. Math. Phys. **17**, 54 (1976).

Relativistic fluids and metric symmetries

K. L. Duggal
Department of Mathematics and Statistics, University of Windsor, Windsor, Ontario, N9B 3P4 Canada

(Received 22 June 1988; accepted for publication 11 January 1989)

This paper deals with fluid space-times within the framework of general relativity and subject to a metric symmetry (self-similarity and conformal or conformal collineation). Via the investigation of the kinematic properties of such fluids, it is shown that the conformal collineation symmetry plays the important role of smooth transition at the boundary of a change in the fluid pattern (such as shear-free and isotropic to distorted fluids). As applications, it is shown that the study is relevant to the radiationlike viscous fluid Friedman–Robertson–Walker model, with the conformal collineation symmetry vector parallel to its tilted velocity vector. Also, the results lead to physically significant material curves (a curve that moves with the fluid as the fluid evolves) in the fluid.

I. INTRODUCTION

According to the theory of general relativity, the Einstein field equations for a space-time manifold are

$$R_{ab} - \tfrac{1}{2}Rg_{ab} = 8\pi G c^{-4} T_{ab} \quad (a,b = 0,...,3), \tag{1.1}$$

where R_{ab}, R, and T_{ab} are the Ricci tensor (computed from the metric tensor g_{ab}), the scalar curvature, and the matter tensor, respectively; G is the Newtonian constant of gravitation; and c is the velocity of light. Here T_{ab} is subject to the conservation law

$$T^{ab};b = 0, \tag{1.2}$$

where the semicolon denotes the covariant derivative with respect to g_{ab}.

The basic problem of relativistic fluids is the study of fluid space-times subject to certain physically meaningful prescribed matter tensors. To illustrate this point, we consider the case of a thermodynamic perfect fluid for which the matter tensor is prescribed by[1-3]

$$T_{ab} = (\mu + pc^{-2})u_a u_b + pg_{ab}, \tag{1.3}$$

where $u^a (u^a u_a = -c^2)$, μ, and p are the fluid four-velocity vector, the density, and the pressure, respectively. The fluid is specified by a barotropic equation of state $p = p(\mu), \mu > 0$, and $p \geq 0$. The continuity equation is

$$(\mu u^a);a = 0. \tag{1.4}$$

If the only matter creating the gravitational field is the fluid whose motion is to be studied, then we call this the "self-attracting problem," for which the unknown functions are ten components of g_{ab}, three independent components of u^a, and the two scalars μ, p satisfying Eqs. (1.1)–(1.4). This problem has been solved in a variety of ways. For example, Taub[1] has shown that for irrotational (vorticity zero) and isentropic (constant entropy) flows, one can choose a simple comoving coordinate system in which the matter tensor depends on g_{44} only. Therefore, the field equations involve a single set of dependent variables, which simplifies the problem; see, also, Refs. 2 and 3.

Since the Einstein field equations are a complicated set of nonlinear partial differential equations (PDE's), one of the most widely used techniques (to simplify the problem) is to impose some *metric symmetry*. For example, plane, cylindrical, or spherical symmetries reduce from ten to three nonzero components of g_{ab}. Well-known physical models are the Schwarzschild and Friedman space-times.[4]

The objective of this paper is to present a few fresh ideas on the role of metric symmetries in the study of fluid space-times. The main emphasis is on the use of a symmetry called *conformal collineation*,[5-8] particularly in dealing with a mixture of fluids and/or a sudden change in the fluid pattern (such as nonviscous to viscous flows). We also discuss some applications in relativistic fluids.

The space-time manifold will be denoted by (M,g), where g is the Lorentz metric of signature $(-+++)$. All structures on M will be assumed smooth.

In Sec. II we review the relevant aspects of self-similarity and conformal symmetry and state their limited use in the study of fluids. In Sec. III we discuss the kinematics of conformal collineation symmetry,[5-8] which takes over when the conformal symmetry breaks and shear appears: This causes distortion in the fluid. We show that non-null conformal collineations play the role of smooth transition at the boundary of the shear-free and distorted regions: This is an extension of an earlier result[8] for a timelike symmetry vector parallel to the four-velocity vector. Section IV is devoted to two applications. First, we show that our study is relevant to the radiationlike viscous fluid solution of the Friedman–Robertson–Walker (FRW) models, with the symmetry vector parallel to its tilted velocity vector and second, our results lead to physically significant material curves (a curve that moves with the fluid as the fluid moves).

II. SELF-SIMILARITY AND CONFORMAL SYMMETRY

A. Self-similarity

Consider a space-time (M,g). We denote by Ω any physical field (such as metric, matter field, scalar, etc.) and use natural units so that $G = c = 1$. Let l be the unit of length. Each Ω can then be assigned a dimension q (usually an integer) such that under the scale transformation

$$l' = e^k l, \quad k = \text{const}, \tag{2.1}$$

Ω transforms as

$$\Omega' = e^{qk}\Omega. \tag{2.2}$$

 0022-2488/89/061316-07$02.50

Since any two physical fields are related by the raising or lowering of the indices with the metric tensor g, it is always better to let g carry the dimension. Thus if we choose $q = 2$ for g, then $q = 1$ and 0 for the space-time interval ds and the coordinates x^a, respectively.

Definition: If there exists a smooth map $M \to M$ such that the metric g transforms under a constant scale factor, i.e.,

$$g \to g' = e^{2k} g, \tag{2.3}$$

then (M,g) is called a self-similar space-time.[9]

Therefore, it follows that the geometry and physics at different points of a region (where self-similarity holds) of a space-time differ only by a change in the overall length scale.

The research on self-similarity in continuum mechanics has its roots in a classical procedure which reduces the PDE's (which characterize a given problem) into ordinary differential equations. Precisely, one assumes a specific form for the solution in which the dependent variables are functions of a single independent variable. For example, in a spherical symmetric problem where the independent variables are a distance from the center of symmetry r and the time t, the dependent variables are assumed to be essentially functions of the variable z.

Cahil and Taub[10] first formulated the relativistic version of self-similar solutions of classical hydrodynamics, followed by several others.[11–16] In general, self-similarity has been extensively studied in several related areas. As examples, see Eardley[9] for the geometry and dynamics and Weinwright[17] for the cosmology of self-similarity.

To understand the primary role of this symmetry in the study of fluids, we consider two regions of the space-time such that a similarity is described in the first region defined by a single variable z: $z_0 \le z \le z_1$, where z_0, z_1 are constants. Then the dividing boundary of the two regions is given by the hypersurface $z = z_1$. The values of the physical fields on the hypersurface will provide the initial data to be filled in the field equations: One then finds that the nature of the initial data generates a unique solution to the field equations.

The basic problem is the fitting of the similarity solution of the first region with that of a known solution in the second region with the possibility of retaining self-similarity. For example, all static solutions retain self-similarity.[10]

However, only the simplest kinds of matter (e.g., perfect fluid, gas, electromagnetic fields, and dust) are allowed to retain self-similarity between two regions. Even a mixture of these are generally disallowed since the boundary of a region where one component of the mixture dominates another would define an intrinsic scale, spoiling self-similarity. In view of this negative phenomenon, one must seek other methods for maintaining the continuity of the fluid matter when the self-similarity breaks. For this purpose, we now discuss another known technique.

B. Conformal symmetry

If there exists a map $M \to M$ such that the metric g transforms under the rule

$$g \to g' = e^{2\phi} g, \quad \phi = \phi(x^a), \tag{2.4}$$

then M is said to have conformal symmetry.[18] Two subcases are self-similarity ($\phi =$ nonzero constant) and isometry ($\phi = 0$). Equation (2.4) implies the existence of a one-parameter group of conformal motions generated by a conformal Killing vector field ξ such that

$$L_\xi g_{ab} = \xi_{a;b} + \xi_{b;a} = 2\phi g_{ab}, \tag{2.5}$$

where L is the Lie-derivative operator.

Example: Consider the Einstein static fluid space-time

$$ds^2 = -dt^2 + (1 - r^2)^{-1} dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2),$$

with the fluid four-velocity vector $u^a = \delta^a_0$. This space-time admits a conformal Killing vector[19]

$$\xi^a = (1 - r^2)^{1/2} \cos t\delta^a_0 - r(1 - r^2)^{1/2} \sin t\delta^a_1.$$

For details on properties of conformal motions see Yano.[18]

Since the celebrated work of Weyl[20] conformal symmetry property has been an essential geometric prescription for a good part of physics. For example, all equations of massless particles,[21] such as the graviton, are conformally invariant. Also, conformal invariance is the root of the twistor program.[22] However, the role of proper conformal symmetry in general relativity is limited to the following recent results.

1. Garfinkle and Tian[23]

Let (M,g) be a solution of the vacuum Einstein equations with the nonzero cosmological constant Λ. Let ξ be a proper conformal vector field of M. Then (M,g) is locally isometric to de Sitter (anti-de Sitter) space-time if Λ is $> 0 (< 0)$.

2. Garfinkle[24]

A space-time that is asymptotically Minkowskian, vacuum, and has positive Bondy energy does not admit any conformal vector field which is not a Killing vector field.

3. Eardley et al.[25]

Eardley *et al.*[25] examined restrictions on space-times via the existence of a conformal vector field for various forms of the stress-energy tensor, with various assumptions about global structure. The theorems support strong indications that proper conformal symmetries have little role to play in general relativity and that homothetic (self-similar) symmetries are useful only for model space-times that are neither spatially compact nor asymptotically flat.

On the other hand, for the study of fluids, conformal symmetry is useful. To illustrate this point, consider a simple example of a liquid that changes to gas when heated through its boiling point. When the pressure is raised, the transition becomes less and less abrupt until at a critical pressure it is continuous. However, at this critical point (the self-similarity breaks) the density fluctuations occur at all length scales. It is a remarkable phenomenon of universality that most physical systems do respond in a natural way to the local conformal symmetry at those critical points. Thus conformal symmetry measures the response of the fluid subject to large density fluctuations and describes the leading finite-size correction to scaling at critical points. For example, it is known that conformal symmetry fits static spherical symmetric distribution of matter[26,27] and in particular, viscous

fluid,[28] to the exterior Schwarzschild metric. We observe that this fitting is not possible for self-similar symmetry.

However, in general, the conformal symmetry breaks down as a result of the action of L_ξ on the Christoffel symbols defined by

$$\left\langle \begin{matrix} c \\ ab \end{matrix} \right\rangle = \frac{1}{2} g^{cd} [g_{bd,a} + g_{ad,b} - g_{ab,d}], \tag{2.6}$$

$$L_x \left\langle \begin{matrix} c \\ ab \end{matrix} \right\rangle = \frac{1}{2} g^{cd} [(L_x g_{bd})_{;a} + (L_x g_{ad})_{;b} - (L_x g_{ab})_{;d}]. \tag{2.7}$$

Indeed, setting $X = \xi$, a conformal vector field, and substituting (2.5), we conclude that every conformal symmetry satisfying (2.5) must also satisfy

$$L_\xi \left\langle \begin{matrix} c \\ ab \end{matrix} \right\rangle = \delta^a_b \phi_{;c} + \delta^a_c \phi_{;b} - g_{bc} g^{ad} \phi_{;d}, \tag{2.8}$$

where δ^a_b is the Kronecker tensor. This raises the following question: Does (2.8) pull back conformal symmetry? In general, the answer is negative. In fact, it can be easily seen [$X = \xi$ in (2.7) $\Rightarrow (L_\xi g_{ab} - 2\phi g_{ab})_{;c} = 0 \Leftrightarrow$(2.9)] that (2.8) is equivalent to

$$L_\xi g_{ab} = 2\phi g_{ab} + K_{ab}, \quad K_{\{ab\}} = 0 = K_{ab;c}, \tag{2.9}$$

where K_{ab} is a symmetric, covariant constant (and therefore a Killing tensor) associated with ξ. For basic information on Killing tensors and their use in physics, see the Appendix.

Based on the above, in Sec. III we discuss a higher symmetry which takes over when the conformal symmetry breaks.

III. CONFORMAL COLLINEATIONS: KINEMATIC RESULTS

A. Conformal collineations

A space-time (M,g) admits a symmetry called "conformal collineation,"[5-8] generated by an affine conformal vector (ACV) field ξ if (2.8), or equivalently, (2.9) holds. Subcases are conformal symmetry ($K_{ab} = Ag_{ab}, A = \text{const}$) and affine collineation ($\phi = \text{const}$), which further includes self-similarity and isometry. In this paper, K_{ab} will be called a conformal Killing tensor (CKT). A special ACV is characterized by (2.9), together with

$$\phi_{;ab} = 0. \tag{3.1}$$

A proper ACV ($K_{ab} \neq g_{ab}$, $\phi = \text{nonconst}$) cannot exist in the general case. For example, it is known[29] that nonflat spaces of constant curvature do not admit a proper ACV since they admit only one CKT $K_{ab} = g_{ab}$.

On the existence of this symmetry, Katzin *et al.*[30] have shown that a nonflat conformally flat space C_n ($n \geqslant 3$) admits a proper ACV such that

$$K_{ab} = BR_{ab}, \quad B = \text{nonzero scalar}. \tag{3.2}$$

This means (in view of $K_{ab;c} = 0$) that C_n, with a proper ACV, is Ricci recurrent.[31] Denoting this class of space by C^*_n, it was further proved[30] that C^*_n is reducible (irreducible) iff it is symmetric ($R = 0 \Rightarrow K^a_a \equiv 0$).

If a reducible M admits a proper ACV ξ, then it is necessarily a combination of a proper affine collineation vector and a proper conformal vector. Mason and Maartens[32] displayed the following example for the Einstein static fluid space-time:

$$ds^2 = -dt^2 + (1-r^2)^{-1} dr^2 + r^2(d\theta^2 + \sin^2\theta\, d\phi^2),$$

with the fluid four-velocity vector $u^a = \delta^a_0$. This space-time admits a proper conformal vector

$$\xi^a_1 = (1-r^2)^{1/2} \cos t \delta^a_0 - r(1-r^2)^{1/2} \sin t \delta^a_1$$

and a proper affine collineation vector $\xi^a_2 = t\delta^a_0$. Thus a combination $\xi^a = \xi^a_1 + \xi^a_2$ is a proper ACV such that

$$\xi^a = [t + (1-r^2)^{1/2} \cos t]\delta^a_0 - r(1-r^2)^{1/2} \sin t \delta^a_1,$$
$$\phi = -(1-r^2)^{1/2} \sin t, \quad K_{ab} = -2t_{,a}t_{,b}. \tag{3.3}$$

In general, for an ACV ξ, the following identities hold (the proof is similar to results on conformal symmetry; see Yano[18]):

$$L_\xi R^d_{abc} = \delta^d_a \phi_{c;b} + \delta^d_b \phi_{a;c} - \phi^d_{;a} g_{bc} + \phi^d_{;b} g_{ac}, \tag{3.4a}$$

$$L_\xi R_{ab} = (\Box\phi) g_{ab} - 2\phi_{;ab}, \tag{3.4b}$$

$$L_\xi R = 6\Box\phi - 2\phi R - R', \tag{3.4c}$$

$$L_\xi C^d_{abc} = \tfrac{1}{2}[g_{bc} K^d_m R^m_a - g_{ac} K^d_m R^m_b - K_{bc} R^d_a + K_{ac} R^d_b] + (R/6)[K_{bc}\delta^d_a - K_{ac}\delta^d_b] - (R'/6)[g_{bc}\delta^d_a - g_{ac}\delta^d_b], \tag{3.4d}$$

where C^d_{abc} is the Weyl curvature tensor of M and

$$\Box\phi \equiv g^{ab}\phi_{;ab}, \quad R' \equiv K^a_b R^b_a. \tag{3.5}$$

Observe that in general, for an ACV, $L_\xi C^d_{abc} \neq 0$. However, any affine collineation implies $L_\xi R^d_{abc} = 0$. Also, conformal symmetry implies $L_\xi C^d_{abc} = 0$.

Using the identities (3.4), it is easy to show that a proper Einstein space [$R_{ab} = (R/4)g_{ab}, R \neq 0$] admits a proper ACV, with

$$K_{ab} = \tfrac{2}{3}((k/4) - \phi)g_{ab} - (8/R)\phi_{;ab}, \quad K \equiv K^a_a. \tag{3.6}$$

It is notable that an affine collineation ($\phi = \text{nonzero const}$) in a proper Einstein space reduces to a full isometry (the proof is easy).

A complete analysis on the existence of ACV's still remains open for indefinite manifolds (for positive definite manifolds this problem was solved by Tshairo[5]).

In view of the above information, we postulate the following prescription for K_{ab}

$$K_{ab} = Ag_{ab} - BR_{ab} + C\phi_{;ab}, \tag{3.7}$$

where A, B, and C are suitable scalars such that $K_{ab;c} = 0$. In the following we show the motivation for this choice. The choice holds for a proper Einstein space with $A = \frac{2}{3}((K/4) - \phi)$, $B = 0$, and $C = -8/R$. Also, nonflat conformally flat spaces [see Eq. (3.2)] can admit this prescription for $A = 0$ and [$C = 0$ or $\phi_{;ab} = 0$]. Finally, this choice is applicable to a variety of physical problems in relativistic fluids (we discuss this in Sec. IV.)

B. Kinematic results

Equation (2.9) may be rewritten as

$$\xi_{(b;a)} = 2\phi g_{ab} + K_{ab}, \quad K_{[ab]} = 0 = K_{ab;c}. \tag{3.8}$$

A decomposition of $\xi_{a;b}$ into its symmetric and skew-symmetric parts and the use of (3.8) provides

$$\xi_{a;b} = \phi g_{ab} + F_{ab} + \tfrac{1}{2}K_{ab}, \tag{3.9}$$

where F is such that

$$F_{ab} = \xi_{[a;b]}. \tag{3.10}$$

A bivector satisfying (3.10) will be called a conformal collineation bivector. This terminology follows that of the homothetic bivector of McIntosh.[33]

Suppose ξ is a non-null ACV. Then, it is useful to discuss the kinematic quantities (shear, expansion, and vorticity) of ξ. A thorough discussion on kinematics may be found in any standard text on cosmology, e.g., Ellis,[34] whose notations will be followed (both for timelike and spacelike ξ). We form the projection tensor h, with the components

$$h_{ab} = g_{ab} - \epsilon\alpha^2\xi_a\xi_b \tag{3.11}$$

such that $h_{ab}\xi^b = 0$ and $\xi\cdot\xi = \epsilon\alpha^2(\epsilon = +1$ or -1 for spacelike or timelike ξ and $\alpha > 0$ is a real scalar).

The following results on conformal symmetry are known.[35,36]

Theorem 1: A space-time admits a timelike (spacelike) conformal motion with the symmetry vector $\xi \| \mu(\xi \| n)$ iff it is shear-free, where $u\cdot n = 0$ and $n\cdot n = 1$.

For an ACV, we have the following corresponding result.

Theorem 2: A space-time admits a non-null ACV field $\xi(\xi\cdot\xi = \epsilon\alpha^2, \epsilon = \pm 1$, and $\alpha > 0)$ iff

$$\sigma_{cd} = (2\alpha)^{-1}\left[h^a_c h^b_d - \tfrac{1}{3}h^{ab}h_{cd}\right]K_{ab}, \tag{3.12a}$$

$$\theta = \alpha^{-1}\left[3\phi + \tfrac{1}{2}h^{ab}K_{ab}\right], \tag{3.12b}$$

where σ_{ab} and θ are the shear tensor and the expansion of ξ.

Proof: Consider the tensor V_{ab} (which represents the relative velocities of neighboring particles) expressed by

$$V_{ab} = h_{ac}h_{bd}\xi^{c;d} = \theta_{ab} + \omega_{ab},$$

where θ_{ab} and ω_{ab} are the expansion and vorticity tensors. Using $\xi^{c;d} = \phi g^{cd} + F^{cd} + \tfrac{1}{2}K^{cd}$, we obtain

$$\theta_{ab} = \theta_{(ab)} = \sigma_{ab} + \tfrac{1}{3}\theta h_{ab}$$

$$= \tfrac{1}{2}h_{ac}h_{bd}K^{cd} + \phi h_{ab}. \tag{3.13}$$

Contracting (3.13) with g^{ab} and using $g^{ab}\sigma_{ab} = 0$, we obtain (3.12b). Substituting (3.12b) in (3.13) and lowering/raising some indices we obtain (3.12a). Conversely, assuming Eqs. (3.12) hold for some CKT K_{ab}, one can show that there exists a non-null ACV ξ^a. ■

Remark 1: It is evident from Theorems 1 and 2 that at the breaking point of the conformal symmetry ($K_{ab} \neq g_{ab}$), shear that causes distortion in the fluid appears. This change in the fluid pattern is governed by the deviation of the CKT K_{ab} from the metric tensor g_{ab}. For example, the fluid will be shear-free (Theorem 1) when $K_{ab} = Ag_{ab}$ (no deviation). Our result (Theorem 2) is consistent with a remark of Collins[37]: "Shear-free perfect-fluid solutions might also be of interest as asymptotic states of space-times in which the flow is not shear-free." Indeed, $\sigma_{ab} \to 0$ iff $K_{ab} \to Ag_{ab}$.

Conclusion 1: The conformal collineation symmetry therefore plays the role of smooth transition between the shear-free and distorted regions.

For a physical feeling of this natural phenomenon, consider a simple case of distortion ($\sigma_{ab} \neq 0$) with no change in volume. Then, if the fluid is at one instant (with conformal symmetry) a spherical ball, a short while later (when the conformal collineation symmetry takes over) it will change to an ellipsoid of the same volume.

Now we are ready to discuss some applications of Theorem 2 based on a suitable prescription for K_{ab}.

IV. APPLICATIONS

We consider the following general form of the energy momentum tensor of a fluid[34]:

$$T_{ab} = (\mu + pc^{-2})u_a u_b + pg_{ab} + \pi_{ab} + q_a u_b + q_b u_a, \tag{4.1}$$

where μ, p, q^a, and π^{ab} are the density, thermodynamic pressure, energy flux vector, and anisotropic pressure tensor, respectively ($q_a u^a = 0$, $\pi_{ab}u^b = 0$, $\pi^a_a = 0$). Let the field equations be (here we set $8\pi G = 1$)

$$R_{ab} - \tfrac{1}{2}Rg_{ab} = c^{-4}T_{ab}. \tag{4.2}$$

Suppose there exists a conformal collineation symmetry defined by Eq. (2.9): This symmetry will in general have unreasonable physical properties unless the CKT K_{ab} has some physically meaningful prescription. Based on postulate (3.11), we set

$$K_{ab} = Ag_{ab} - BR_{ab}, \quad C = 0, \quad B \neq 0. \tag{4.3}$$

The general case (when $C \neq 0$) will be discussed in another paper. Since we are investigating fluid-filled spaces for which $R_{ab} \neq g_{ab}$, the choice of (4.3) (which is not unique) is suitable for a proper ($K_{ab} \neq g_{ab}$) ACV. Moreover, $C = 0$ will not restrict our discussion to a special ACV for which $\phi_{;ab} = 0$.

A. Viscous fluids

The following relations will hold:

$$R_{ab}u^a u^b = \tfrac{1}{2}(\mu + 3pc^{-2}), \quad R_{ab}u^a h^b_c = -q_c, \tag{4.4}$$

$$R_{ab}h^a_c h^b_d = (\mu - pc^{-2})h_{cd} + \pi_{cd},$$

where $h_{ab} = g_{ab} + c^{-2}u_a u_b$ is the projection tensor. Therefore, it follows from (3.11) that ξ is timelike such that

$$\xi^a = \alpha c^{-1}u^a. \tag{4.5}$$

Using (4.3) and (4.4) in (3.12a), we obtain

$$\pi_{ab} = -(2\alpha/B)\sigma_{ab}. \tag{4.6}$$

It is known that a particular case of anisotropic fluid is

viscous fluids,[34,38–42] characterized by (4.1) and (4.6), where the kinematical viscosity coefficient η is

$$\eta = 2\alpha/B > 0, \quad \alpha = (\epsilon\xi\cdot\xi)^{1/2} > 0.$$

Thus we have established the relativistic equivalence to the Navier–Stokes theory of Newtonian fluid mechanics.[39]

Example: Consider FRW models

$$ds^2 = -c^2\,dt^2 + S^2(t)[(1-kr^2)^{-1}\,dr^2 + r^2(d\theta^2 + \sin^2\theta\,d\phi^2)],$$

where t, $S(t)$, and $k(0 \text{ or } \pm 1)$ are the cosmic time, scalar function, and curvature constant, respectively. Introducing the dimensionless function $\epsilon(t) = p/\mu c^2$, one can find various FRW models by prescribing $\epsilon(t)$. For example, $\epsilon = \text{const}$ recovers the standard FRW models and $\epsilon \to \frac{1}{3}$ as $t \to 0$ and $\epsilon \to 0$ as $t \to \infty$ are the radiationlike models (representing the epoch near the initial singularity for small t) and dustlike models [the ultimate (including present) stage of the universe for large t]. The interest in such models has its roots in the discovery of the cosmic microwave background.[43,44]

Related to the present paper, Coley and Tupper[42] have recently proved that FRW models with $k = 0$ (flat geometry) can be exact solutions for a viscous fluid (with or without an electromagnetic field). Consider Coley and Tupper's radiationlike viscous fluid solution with $S(t) = t/2$ and the tilting velocity vector

$$\bar{u} = (2c/t)[\cosh\Phi\,\partial_t + \sinh\Phi\,\partial_r], \tag{4.7}$$

where $\Phi(t,r)$ is the hyperbolic tilt angle. The shear tensor is

$$\sigma_{ab} = \sqrt{3}\sigma[\tfrac{1}{3}h_{ab} - n_a n_b],$$

where $\sigma = \sigma(t,r)$ is the magnitude of σ_{ab}, $n = (2/t)_r$ is the unit radial vector, and

$$\pi_{ab} = -\eta\sigma_{ab}, \quad q^a \neq 0.$$

Relating the above example with the ACV data of this paper, we observe that since FRW models are conformally flat, it follows from Eq. (3.2) and the constant scalar curvature R that B is constant. Therefore, using (4.5) and (4.7), we obtain

$$\xi^a = (\alpha t/2c)\bar{u}^a, \quad \eta = (\alpha/B), \quad B = \text{nonzero const.} \tag{4.8}$$

Discussion: The tilt of the matter velocity field relative to the radiation field causes turbulence which drives the viscosity to high magnitude. Attempts have been made to obtain the upper limit on the viscosity (for example, see Goicoechea and Sauz[45] for FRW models with $k = 0$ and a homogeneous velocity field). By (4.8), it follows that the viscosity coefficient η is constantly related to the magnitude α of the ACV ξ.

Conclusion 2: By adjusting the magnitude of its ACV ξ, conformal collineation symmetry may be used to measure the response of viscous fluid subject to large viscosity fluctuations in order to describe the leading finite-size correction to scaling.

In general, viscous fluids are of great interest in explaining the relativistic dissipative processes in situations near the thermodynamic equilibrium[38] and have been used to describe neutron stars in certain density ranges.[46]

B. Material curves

A material curve in a fluid is a curve that moves with the fluid as the fluid evolves (this is also known as the "frozen in" curve to the fluid). For instance, if an observer comoving with the fluid can be chosen at any one point, then a comoving observer can be employed at any other point along a spacelike congruence in the fluid iff the curves of the congruence are material curves in the fluid.

The interest in material curves has its roots in the theory of spacelike congruences first formulated by Greenberg[47]: This theory plays an important role in relativistic fluid dynamics and relativistic electrodynamics of continuous media.

We show that non-null ACV's lead to physically significant material curves in fluid space-times. Our results agree with known results on conformal motions. For this, we need the following. The effect of an ACV on any non-null unit vector X^a is[6]

$$\begin{aligned} L_\xi X^a &= -[\phi + (\epsilon/2)K_{bc}X^bX^c]X^a + Y^a, \\ L_\xi X_a &= [\phi - (\epsilon/2)K_{bc}X^bX^c]X_a + K_{ab}X^b + Y_a, \end{aligned} \tag{4.9}$$

where Y^a is some vector orthogonal to X^a. In general, $Y^a \neq 0$: See Ref. 6 for an explicit example of an ACV with $Y^a \neq 0$. In particular, for a fluid four-velocity vector u^a, we obtain

$$\begin{aligned} L_\xi u^a &= -(\phi - \tfrac{1}{2}K_{bc}u^bu^c)u^a + v^a, \\ L_\xi u_a &= (\phi + \tfrac{1}{2}K_{bc}u^bu^c)u_a + K_{ab}u^b + v_a, \end{aligned} \tag{4.10a}$$

where $v\cdot u = 0$. By (4.10), it follows that the integral curves of an ACV ξ^a are material curves iff $v^a = 0$. This means that

$$L_\xi u^a = -(\phi - \tfrac{1}{2}K_{bc}u^bu^c)u^a. \tag{4.10b}$$

For further analysis, we decompose $\xi^a = \beta u^a + Y^a$, where $\beta = -u_a\xi^a$ and $Y^a u_a = 0$. The following kinematic result is known[34]:

$$u_{a;b} = \sigma_{ab} + \tfrac{1}{3}\theta h_{ab} + \omega_{ab} - \dot{u}_a u_b,$$

where we set the velocity of light $c = 1$. Using this result, we obtain

$$L_\xi u_a = \dot{\beta}u_a + \beta[\dot{u}_a + (\log\beta^{-1})_{;b}] + Y^b(\dot{u}_a u_b + 2\omega_{ab}).$$

Equating the above result with (4.10a) and then contracting with u^a provides $\phi = \dot{\beta} + \dot{u}_a\xi^a + \frac{1}{2}K_{ab}u^au^b$. Finally, eliminating ϕ, we obtain

$$v_a = 2\omega_{ab}\xi^b + \beta[\dot{u}_a + (\log\beta^{-1})_{;b}h^b{}_a] - (K_{bc}u_bu_c)u_a - K_{ab}u^b. \tag{4.11}$$

For an ACV ξ with K_{ab} satisfying (4.3), the following holds [here we use Eq. (4.4) for $c = 1$]:

$$(K_{bc}u^bu^c)u_a + K_{ab}u^b = 0. \tag{4.12}$$

Therefore, Eq. (4.11) reduces to

$$v_a = 2\omega_{ab}\xi^b + \beta\left[\dot{u}_a + (\log\beta^{-1})_{;b}h^b{}_a\right]. \quad (4.13)$$

Remark 2: It is evident from (4.13) that the relation between v^a and the kinematic quantities is independent of K_{ab}. This means that the properties of material curves (when $v^a = 0$) forbid the influence of K_{ab} if K_{ab} satisfies (4.3). Based on this, in particular, the following relevant results on spacelike congruences and non-null conformal symmetry will also hold for conformal collineation symmetry.

Theorem 3: If a unit tangent vector n^a to any spacelike congruence is everywhere orthogonal to the four-velocity vector field u^a of a self-gravitating fluid, then comoving observers can be employed all along a curve of the congruence iff the curves are material curves in the fluid (Tsamparlis and Mason[48]).

Theorem 4: If a rotational fluid ($\omega \neq 0$) space-time admits a timelike conformal vector $\xi \| u$, then the vortex lines are material curves in the fluid (Ehlers *et al.*[49]).

Theorem 5: Suppose a fluid space-time admits a spacelike conformal vector $\xi \| n$, where $n\cdot n = 1$ and $u\cdot n = 0$.

(i) If the fluid is irrotational ($\omega = 0$), then the integral curves of n must be the material curves in the fluid.

(ii) If the vorticity of the fluid is nonzero ($\omega \neq 0$), then the integral curves of n are material curves in the fluid iff they are vortex lines (Mason and Tsamparlis[36]).

Conclusion 3: If its CKT K_{ab} is prescribed by Eq. (4.3), conformal collineation symmetry may be useful to retain the properties of material curves under a change in fluid pattern when the conformal symmetry breaks.

V. CONCLUDING REMARKS

We have seen that metric symmetries (discussed within the scope of the present paper) are useful in the study of fluid space-times. This discussion indicates several topics for further investigation. The results may be used to derive general properties of non-null conformal collineations. This may be extended to null conformal collineations, for which the relevant kinematic quantities can be defined in a manner similar to that outlined in the present work. Finally, it is hoped that the present investigation, together with earlier works,[6–8] will help stimulate further research on this topic.

ACKNOWLEDGMENTS

This research was supported by the Natural Science and Engineering Council of Canada and the University of Windsor Research Board under a special project of its Fluid Dynamics Research Institute.

APPENDIX: KILLING TENSORS

A vector K is called a Killing vector if

$$K_{(a;c)} = \tfrac{1}{2}(K_{a;c} + K_{c;a}) = 0. \quad (A1)$$

As a generalization of Killing vectors, a Killing tensor of order m is a symmetric tensor $K_{a_1,\ldots,a_m}$ which satisfies

$$K(a_1,\ldots,a_m;c) = 0. \quad (A2)$$

Within the scope of this paper, we are interested in Killing tensors of order 2, which in accordance with (6.2) obey

$$K_{[ab]} = 0 = K_{(ab;c)}. \quad (A3)$$

Trivial examples of a Killing tensor are the metric tensor g_{ab}; all products $K_{ab} = U_{(a}V_{b)}$ of the Killing vectors U_a, V_b (not necessarily different); and linear combinations of these with constant coefficients. Killing tensors that do not admit this type of representation are referred to as nontrivial.

From the point of view of physics, interest in the Killing tensors originated in their connection with the separability of various PDE's e.g., the Hamilton–Jacobi equation. For further details regarding this connection, we refer the reader to Refs. 50–55.

In reference to the existence of Killing tensors, we state the following *open problems*: (i) find all Killing tensors of a given space-time and/or (ii) classify space-times with respect to the nontrivial Killing tensors they admit. Neither problem has yet been solved, although some specific results are known. Here we only state one fundamental result (Hauser and Malhiot[56]). A four-dimensional space-time admits at most 50 linearly independent Killing tensors of order 2. The maximum number 50 is attained iff the space-time is of constant curvature, for which all 50 Killing tensors are trivial. For the general study of Killing tensors we refer the reader to Refs. 57–60.

[1] A. H. Taub, Phys. Rev. **103**, 454 (1956).
[2] A. H. Taub, Ann. Rev. Fluid Mech. **10**, 301 (1978).
[3] A. Lichnerowicz, *Relativistic Hydrodynamics and Magnetohydrodynamics* (Benjamin, New York, 1967).
[4] S. W. Hawking and G. F. R. Ellis, *The Large Scale Structure of Space-time* (Cambridge U. P., Cambridge, 1973).
[5] Y. Tashiro, Math. J. Okayama Univ. **10**, 75 (1960).
[6] K. L. Duggal and R. Sharma, J. Math. Phys. **27**, 2511 (1986).
[7] D. P. Mason and R. Maartens, J. Math. Phys. **28**, 2182 (1987).
[8] K. L. Duggal, J. Math. Phys. **28**, 2700 (1987).
[9] D. Eardley, Commun. Math. Phys. **37**, 287 (1974).
[10] M. E. Cahill and A. H. Taub, Commun. Math. Phys. **21**, 1 (1971).
[11] R. N. Henrikson and P. S. Wesson, Astrophys. Space Sci. **53**, 429 (1978).
[12] G. V. Bienell and R. N. Henrikson, Astrophys. J. **219**, 1043 (1978).
[13] G. V. Bienell and R. N. Henrikson, Astrophys. J. **225**, 237 (1978).
[14] D. E. Wilson, Phys. Fluids **29**, 632 (1986).
[15] S. M. Caridi, Phys. Fluids **29**, 672 (1986).
[16] R. B. Lazarus, Phys. Fluids **25**, 1146 (1982).
[17] J. R. Wainwright, in *Axisymmetric Systems*, edited by M. A. H. MacCallum (Cambridge U. P., Cambridge, 1985).
[18] K. Yano, *The Theory of Lie Derivatives and its Applications* (North-Holland, Amsterdam, 1955).
[19] R. Maartens and S. D. Maharaj, Class. Quant. Grav. **3**, 1005 (1986).
[20] H. Weyl, Cottingen Nucher. **99** (1921).
[21] R. Penrose, Proc. R. Soc. London Ser. A **284**, 159 (1965).
[22] R. Penrose, *Twistor Theory, Its Aims and Achievements*, Quantum Gravity, Oxford Symposium (Clarendon, Oxford, 1975).
[23] D. Garfinkle and Q. Tian, Class. Quant. Grav. **4**, 137 (1987).
[24] D. Garfinkle, J. Math. Phys. **28**, 28 (1987).
[25] D. Eardley, J. Isenberg, J. Marsden, and V. Monreif, Commun. Math. Phys. **106**, 137 (1986).
[26] L. Herrera and J. Jimenez, Phys. Rev. D **28**, 2987 (1983).
[27] L. Herrera, J. Jimenez, L. Leal, and J. Ponce de Leon, J. Math. Phys. **25**, 3274 (1984).
[28] J. Carot and L. I. Mas, J. Math. Phys. **27**, 2336 (1986).
[29] J. Levine and G. H. Katzin, Tensor (NS) **19**, 317 (1968).
[30] G. H. Katzin, J. Levine, and W. R. Davis, Tensor (NS) **21**, 51 (1970).
[31] G. S. Hall, Phys. Lett. A **56**, 17 (1976).
[32] D. P. Mason and R. Maartens, J. Math. Phys. **28**, 2511 (1986).
[33] C. B. G. McIntosh, Gen. Relativ. Gravit. **7**, 199 (1976).

[34]G. F. R. Ellis, in *XLIII Enrico Fermi Summer School Proceedings*, edited by R. K. Sachs (Academic, New York, 1971).
[35]D. R. Oliver and W. R. Davis, Gen. Relativ. Gravit. **8**, 905 (1977).
[36]D. P. Mason and M. Tsamparlis, J. Math. Phys. **26**, 2881 (1985).
[37]B. Collins, Can. J. Phys. **9**, 191 (1986).
[38]C. Eckart, Phys. Rev. **58**, 919 (1940).
[39]L. Landau and E. M. Lifshitz, *Fluid Mechanics* (Addison-Wesley, Reading, MA, 1958).
[40]B. O. J. Tupper, J. Math. Phys. **22**, 2666 (1981).
[41]S. R. Maiti and S. N. Das, J. Math. Phys. **26**, 2840 (1985).
[42]A. A. Coley and B. O. J. Tupper, Astrophysics. J. **288**, 418 (1985).
[43]A. A. Penzias and R. W. Wilson, Astrophys. J. **142**, 419 (1965).
[44]P. G. Roll and D. G. Wilkinson, Phys. Rev. Lett. **16**, 405 (1966).
[45]L. J. Goicoechea and J. L. Sanz, Astrophys. J. **286**, 392 (1984).
[46]V. Canuto, Solvay Conference on Astrophysics and Gravitation, Brussels, 1973.
[47]P. J. Greenberg, J. Math. Anal. Appl. **30**, 128 (1970).
[48]M. Tsamparlis and D. P. Mason, J. Math. Phys. **24**, 1577 (1983).
[49]J. Ehlers, P. Geren, and R. K. Sachs, J. Math. Phys. **9**, 1344 (1986).
[50]P. Havas, J. Math. Phys. **16**, 1461, 2476 (1975).
[51]N. M. J. Woodhouse, Commun. Math. Phys. **44**, 9 (1975).
[52]W. Dietz, J. Phys. A **9**, 519 (1976).
[53]C. D. Collinson and J. Fugere, J. Phys. A **10**, 745 (1977).
[54]C. B. Boyer, E. G. Kalnins, and W. Miller, Commun. Math. Phys. **59**, 285 (1978).
[55]S. Benenti and M. Francaviglia, Gen. Relativ. Gravit. **10**, 79 (1979).
[56]I. Hausen and R. J. Malhiot, J. Math. Phys. **16**, 150 (1975); J. Math. Phys. **16**, 1625 (1975).
[57]L. P. Eisenhart, Trans. Am. Math. Soc. **25**, 297 (1923).
[58]E. M. Patterson, Q. J. Math. Oxford **2** (2), 151 (1951).
[59]M. Walker and R. Penrose, Commun. Math. Phys. **18**, 265 (1970).
[60]R. Geroch, J. Math. Phys. **11**, 1955 (1970).

Debye potentials for Rarita–Schwinger fields in curved space-times

G. F. Torres del Castillo
Departamento de Física Matemática, Instituto de Ciencias de la Universidad Autónoma de Puebla, 72000 Puebla, Mexico and Departamento de Física, Centro de Investigación y de Estudios Avanzados del IPN, Apartado Postal 14–740, 07000 Mexico, D.F., Mexico

(Received 9 September 1988; accepted for publication 4 January 1989)

Expressions for the complete solution of the Rarita–Schwinger equation in terms of complex scalar potentials are obtained by means of Wald's method of adjoint operators. The background space-time is required to be an algebraically special solution of the Einstein vacuum field equations with cosmological constant or a solution of the Einstein–Maxwell equations such that one principal null direction of the electromagnetic field is geodetic and shear-free.

I. INTRODUCTION

The zero-rest-mass field equations on a curved space-time constitute a highly involved system of differential equations because of the coupling between the various components of the field and the background geometry. When the space-time is not (conformally) flat, the usual zero-rest-mass field equations for spins greater than 1 have nontrivial integrability conditions that restrict their solutions. In the case of a spin-2 field, these integrability conditions are identically satisfied if the field is identified with a gravitational perturbation. For spin-$\frac{3}{2}$ fields, consistent equations are obtained by linearizing the supergravity field equations; however, the equations obtained in this way do vary depending on the presence of a cosmological constant or another field.

In the case of the algebraically special vacuum space-times, for the zero-rest-mass fields of spin $\frac{1}{2}$, 1, and 2, there exists a decoupled equation for a certain component of the field,[1] and the complete solution is expressible in terms of a single scalar potential that has to satisfy a wavelike equation.[2,3] Decoupled equations have also been derived in the case of spin $\frac{3}{2}$ when the background space-time satisfies the Einstein vacuum field equations without cosmological constant[4,5] and when it satisfies the Einstein–Maxwell equations with one of the principal null directions of the electromagnetic field being geodetic and shear-free.[6,7] Expressions for all the components of the field in terms of a complex potential have been obtained only for the case of vacuum without cosmological constant.[5]

The solution of the decoupled equations derived from the zero-rest-mass field equations gives a component of the field only, which constitutes limited but useful information.[1] Nevertheless it turns out that in these cases, whenever there exists an appropriate decoupled equation, it is possible to express the complete solution in terms of scalar potentials.[3,8–10] The procedure to obtain these expressions and the equations for the potentials starting from the decoupled equations is essentially simple and is based on the concept of the adjoint of a linear differential operator.

In the present paper, the equations for zero-rest-mass fields of spin $\frac{3}{2}$ given by the linearized supergravity theory are reduced to a single wavelike equation by means of the method of adjoint operators, assuming that the background space-time is a solution of the Einstein vacuum field equations with cosmological constant, or a solution of the Einstein–Maxwell equations such that one principal null direction of the electromagnetic field is geodetic and shear-free. In both cases we recover the expressions found by direct integration in Ref. 5, applicable to the case of a vacuum without cosmological constant.

In Sec. II, the method of adjoint operators, which allows the reduction of systems of homogeneous linear partial differential equations to equations for potentials that determine the complete solution of the original system, is presented following Ref. 3. In Sec. III, from the linearized field equations of simple supergravity with cosmological constant[11] we obtain a decoupled equation for the spin-$\frac{3}{2}$ field, assuming that the background space-time is algebraically special, and then the expressions for all the components of the field in terms of a scalar (Debye) potential are obtained. In Sec. IV an analogous derivation is given, starting from the linearized field equations of the $N = 2$ extended supergravity[12] in the case of a background space-time that satisfies the Einstein–Maxwell equations with one principal null direction of the electromagnetic field being geodetic and shear-free. The spinor formalism and the Newman–Penrose notation are used throughout, following Ref. 13.

II. DECOUPLED EQUATIONS AND POTENTIALS

Let f be a tensor or spinor field governed by a system of homogeneous linear partial differential equations that can be written in the form

$$\mathscr{E}(f) = 0, \tag{1}$$

where $\mathscr{E}$ is a linear differential operator that maps tensor or spinor fields like f into tensor or spinor fields, possibly of a type different from that of f. Examples of (1) are the Dirac equation, the source-free Maxwell equations, the Rarita–Schwinger equation, and the linearized Einstein equations. By combining appropriately the equations in (1) and their derivatives, one may be able to obtain a decoupled equation of the form

$$\mathscr{O}(\chi) = 0, \tag{2}$$

where $\mathscr{O}$ is a linear differential operator that maps scalar fields into scalar fields, and χ is a function made out (linearly) of the components of f or their derivatives; the scalar field

 0022-2488/89/061323-06$02.50

χ can then be expressed as $\chi = \mathscr{T}(f)$, where $\mathscr{T}$ is another linear operator. The fact that Eq. (2) follows from Eq. (1) is equivalent to the existence of a linear operator $\mathscr{S}$ such that

$$\mathscr{S}\mathscr{E} = \mathscr{O}\mathscr{T} \tag{3}$$

identically, so that when both sides of (3) are applied to a solution f of Eq. (1), one gets the decoupled equation (2).

If g is a field of the same type of $\mathscr{E}(f)$, in such a way that the full contraction of g and $\mathscr{E}(f)$, denoted by $g\cdot\mathscr{E}(f)$, yields a scalar field, then the adjoint of $\mathscr{E}$, $\mathscr{E}^\dagger$ is defined as that linear operator such that[3]

$$g\cdot\mathscr{E}(f) - f\cdot\mathscr{E}^\dagger(g) = \nabla_\mu s^\mu \tag{4}$$

for every pair of fields f and g of the required types, where s^μ is some vector field. It can be readily verified that $(\mathscr{A}^\dagger)^\dagger = \mathscr{A}$, $(\mathscr{A}+\mathscr{B})^\dagger = \mathscr{A}^\dagger + \mathscr{B}^\dagger$, $(\mathscr{A}\mathscr{B})^\dagger = \mathscr{B}^\dagger\mathscr{A}^\dagger$, and if h is a function, $h^\dagger = h$. Thus from Eq. (3) one gets the identity

$$\mathscr{E}^\dagger\mathscr{S}^\dagger = \mathscr{T}^\dagger\mathscr{O}^\dagger, \tag{5}$$

and therefore if ψ is a function that satisfies the equation

$$\mathscr{O}^\dagger(\psi) = 0, \tag{6}$$

then the field $\mathscr{S}^\dagger(\psi)$ satisfies the equation

$$\mathscr{E}^\dagger(\mathscr{S}^\dagger(\psi)) = 0. \tag{7}$$

In particular, if $\mathscr{E}$ is self-adjoint or antiself-adjoint (i.e., $\mathscr{E}^\dagger = \pm\mathscr{E}$), then by virtue of (7), $f = \mathscr{S}^\dagger(\psi)$ is a solution of Eq. (1), provided that the potential function ψ obeys Eq. (6).

If, instead of being a single field, f is an array of fields and, similarly, χ has several components, then all the previous conclusions remain valid, with ψ being now an array of scalar potentials, provided that the contractions appearing in the definition (4) include sums over the components of the array in order for the left-hand side of Eq. (4) to be a scalar.[8–10]

III. SPIN-$\frac{3}{2}$ PERTURBATIONS OF VACUUM BACKGROUNDS WITH COSMOLOGICAL CONSTANT

The supergravity field equations reduce to the equations of general relativity when the spin-$\frac{3}{2}$ fields vanish (see, e.g., Ref. 14). If the supergravity field equations are linearized with respect to the spin-$\frac{3}{2}$ fields about a solution with vanishing spin-$\frac{3}{2}$ fields, apart from the equations of general relativity with a torsion-free connection, a consistent set of equations for the spin-$\frac{3}{2}$ fields is obtained. In this way one finds the following equations for a spin-$\frac{3}{2}$ massless field in a background space-time that satisfies the Einstein vacuum field equations with cosmological constant[11,14]:

$$\nabla_{AB'}\psi^A{}_{CD'} = \nabla_{CD'}\psi^A{}_{AB'} + ig(2\bar\psi_{B'D'C} - \bar\psi_{D'B'C}), \tag{8}$$

where the real constant g must be related to the scalar curvature through

$$g^2 = -\Lambda, \tag{9}$$

and $\bar\psi_{B'D'C} \equiv \overline{\psi_{BDC'}}$, which can be written in the equivalent form

$$H_{ABC} = H_{(ABC)}, \quad H_{AB'C'} = 0, \tag{10}$$

where

$$\begin{aligned} H^A{}_{BC} &\equiv \nabla_{(B|R'|}\psi^A{}_{C)}{}^{R'} + ig\epsilon^A{}_{(B}\bar\psi_{|R'|}{}^{R'}{}_{C)}, \\ H^A{}_{B'C'} &\equiv \nabla_{R(B'}\psi^{AR}{}_{C')} - ig\bar\psi_{(B'C')}{}^A \end{aligned} \tag{11}$$

(the round brackets denote symmetrization on the indices enclosed; the indices between bars are excluded from the symmetrization). These equations are invariant under the supersymmetry transformations

$$\psi_{ACD'} \to \psi_{ACD'} + \nabla_{CD'}\alpha_A + ig\epsilon_{AC}\bar\alpha_{D'}, \tag{12}$$

where α_A is an arbitrary spinor field and $\bar\alpha_{A'} = \overline{\alpha_A}$, provided that the traceless part of the Ricci tensor vanishes,

$$\Phi_{ABC'D'} = 0, \tag{13}$$

and condition (9) is fulfilled.

In order to obtain identities of the form (3) it is convenient to introduce the spinor field

$$J_{D'B'C} \equiv \nabla_{AB'}\psi^A{}_{CD'} - \nabla_{CD'}\psi^A{}_{AB'} - ig(2\bar\psi_{B'D'C} - \bar\psi_{D'B'C}), \tag{14}$$

where $\psi_{ABC'}$ is not restricted by any condition; then $\psi_{ABC'}$ satisfies Eq. (8) if and only if $J_{D'B'C} = 0$. From the definitions (11) and (14) one finds that

$$H^A{}_{AC} = \tfrac{1}{2}J^{R'}{}_{R'C}, \quad H_{AB'C'} = J_{(B'C')A}, \tag{15}$$

and using the Ricci identities together with Eqs. (9) and (13),

$$\nabla^{AR'}H_{ABC} = \Psi_{ABCD}\psi^{ADR'} + \nabla_{(B}{}^{S'}J_{|S'|}{}^{R'}{}_{C)} - ig\bar J_{(BC)}{}^{R'}, \tag{16}$$

where $\bar J_{BCR'} = \overline{J_{B'C'R}}$.

In order to express Eq. (8) as a system of linear differential equations, we need to consider $\psi_{ABC'}$ and $\bar\psi_{A'B'C}$ as the unknowns; then Eq. (8) together with its complex conjugate can be written as

$$\mathscr{E}\begin{bmatrix}\psi_{ABC'} \\ \bar\psi_{A'B'C}\end{bmatrix} = 0, \tag{17}$$

where

$$\mathscr{E}\begin{bmatrix}\psi_{ABC'} \\ \bar\psi_{A'B'C}\end{bmatrix} \equiv \begin{bmatrix}\bar J_{ABC'} \\ J_{A'B'C}\end{bmatrix}, \tag{18}$$

with $J_{A'B'C}$ defined by (14). This means that $\mathscr{E}$ is a 2×2 matrix of linear operators, which turns out to be antiself-adjoint. In fact, using the definition (14) one can easily see that for an arbitrary spinor field $\phi_{ABC'}$,

$$\begin{aligned} &[\phi^{ABC'} \quad \bar\phi^{A'B'C}]\mathscr{E}\begin{bmatrix}\psi_{ABC'} \\ \bar\psi_{A'B'C}\end{bmatrix} \\ &\quad + [\psi^{ABC'} \quad \bar\psi^{A'B'C}]\mathscr{E}\begin{bmatrix}\phi_{ABC'} \\ \bar\phi_{A'B'C}\end{bmatrix} \\ &\quad = \nabla_{AB'}(\bar\phi^{D'B'C}\psi^A{}_{CD'} - \bar\phi^{B'D'A}\psi^C{}_{CD'} \\ &\qquad + \phi^{CAD'}\bar\psi^{B'}{}_{D'C} - \phi^{ACB'}\bar\psi^{D'}{}_{D'C}). \end{aligned} \tag{19}$$

Taking $\phi_{ABC'} = \psi_{ABC'}$, this last identity shows that if $\psi_{ABC'}$ satisfies Eq. (8), then there exists a divergenceless real vec-

tor field made out of the spin-$\frac{3}{2}$ field [which is equal to $2(\bar{\psi}^{D'B'C}\psi^{A}{}_{CD'} - \bar{\psi}^{B'D'A}\psi^{C}{}_{CD'})$].

If $J_{A'B'C}$ is equal to zero [i.e., if $\psi_{ABC'}$ satisfies Eq. (8)], then the spinor field H_{ABC} obeys an equation identical to that found in Ref. 5 in the case of a Ricci-flat space-time [see Eq. (16)] which leads to a decoupled equation for each multiple principal spinor of the conformal curvature. Indeed if the conformal curvature has a multiple principal spinor, then by aligning appropriately the spin frame, we have $\Psi_0 = 0 = \Psi_1$ and, as a consequence of (13), $\kappa = 0 = \sigma$. Hence from Eq. (16) making use of the first equation in (15), it follows that

$$\begin{aligned}
&(\bar{\delta} - 3\alpha + \pi)H_{000} - (D - \epsilon - 3\rho)H_{100} \\
&\quad - \rho(J_{1'0'0} - J_{0'1'0}) \\
&\quad = \Psi_2\psi_{000'} + (\delta - \beta - 2\bar{\alpha} + \bar{\pi})J_{0'0'0} \\
&\qquad - (D - \epsilon - \bar{\rho})J_{1'0'0} + \bar{\rho}J_{0'1'0} - ig\bar{J}_{000'}, \\
&(\Delta - 3\gamma + \mu)H_{000} - (\delta - \beta - 3\tau)H_{100} \\
&\quad - \tau(J_{1'0'0} - J_{0'1'0}) \\
&\quad = \Psi_2\psi_{001'} + (\delta - \beta + \bar{\pi})J_{0'1'0} \\
&\qquad - (D - \epsilon + 2\bar{\epsilon} - \bar{\rho})J_{1'1'0} - \bar{\lambda}J_{0'0'0} \\
&\qquad + \bar{\pi}J_{1'0'0} - ig\bar{J}_{001'}. \qquad (20)
\end{aligned}$$

Then by applying $(\delta - 2\beta - \bar{\alpha} - 3\tau + \pi)$ to the first equation and $(D - 2\epsilon + \bar{\epsilon} - 3\rho - \rho)$ to the second one and subtracting, the terms with H_{100} cancel by virtue of the identity[1]

$$\begin{aligned}
&[D + (p-1)\epsilon + \bar{\epsilon} + q\rho - \bar{\rho}](\delta + p\beta + q\tau) \\
&\quad = [\delta + (p-1)\beta - \bar{\alpha} + q\tau + \bar{\pi}](D + p\epsilon + q\rho),
\end{aligned} \qquad (21)$$

where p and q are two arbitrary constants, which follows from $\kappa = \sigma = \Psi_1 = 0$. Making use of the Bianchi identities $(D - 3\rho)\Psi_2 = 0$ and $(\delta - 3\tau)\Psi_2 = 0$, and of the equality $H_{000} = (D - 2\epsilon + \bar{\epsilon} - \bar{\rho})\psi_{001'} - (\delta - 2\beta - \bar{\alpha} + \bar{\pi})\psi_{000'}$ which follows from Eq. (11), one finds that

$$\begin{aligned}
&[(D - 2\epsilon + \bar{\epsilon} - 3\rho - \bar{\rho})(\Delta - 3\gamma + \mu) \\
&\qquad - (\delta - 2\beta - \bar{\alpha} - 3\tau + \bar{\pi})(\bar{\delta} - 3\alpha + \pi) - \Psi_2]H_{000}, \\
&\quad = K,
\end{aligned} \qquad (22)$$

where we have introduced

$$\begin{aligned}
K \equiv\ & (D - 2\epsilon + \bar{\epsilon} - 3\rho - \bar{\rho})[(\delta - \beta - \tau + \bar{\pi})J_{0'1'0} + (\tau + \bar{\pi})J_{1'0'0} - (D - \epsilon + 2\bar{\epsilon} - \bar{\rho})J_{1'1'0} - \bar{\lambda}J_{0'0'0} - ig\bar{J}_{001'}] \\
& - (\delta - 2\beta - \bar{\alpha} - 3\tau + \bar{\pi})[(\delta - \beta - 2\bar{\alpha} + \bar{\pi})J_{0'0'0} + (\bar{\rho} - \rho)J_{0'1'0} - (D - \epsilon - \rho - \bar{\rho})J_{1'0'0} - ig\bar{J}_{000'}].
\end{aligned} \qquad (23)$$

Thus by defining the linear operators $\mathscr{S}$ and $\mathscr{T}$ by

$$\mathscr{S}\begin{bmatrix} \bar{J}_{ABC'} \\ J_{A'B'C} \end{bmatrix} \equiv \begin{bmatrix} K \\ \bar{K} \end{bmatrix}, \quad \mathscr{T}\begin{bmatrix} \psi_{ABC'} \\ \bar{\psi}_{A'B'C} \end{bmatrix} \equiv \begin{bmatrix} H_{000} \\ \overline{H_{000}} \end{bmatrix}, \qquad (24)$$

from Eqs. (18) and (22) we see that

$$\mathscr{S}\mathscr{C}\begin{bmatrix} \psi_{ABC'} \\ \psi_{A'B'C} \end{bmatrix} = \mathscr{O}\mathscr{T}\begin{bmatrix} \psi_{ABC'} \\ \bar{\psi}_{A'B'C} \end{bmatrix}, \qquad (25)$$

where $\mathscr{O}$ is a diagonal matrix formed by $(D - 2\epsilon + \bar{\epsilon} - 3\rho - \bar{\rho})$ $(\Delta - 3\gamma + \mu)$ $-(\delta - 2\beta - \bar{\alpha} - 3\tau + \bar{\pi}) \times (\bar{\delta} - 3\alpha + \pi) - \Psi_2$ and its complex conjugate. Therefore using the fact that

$$\begin{aligned}
D^{\dagger} &= -D - \epsilon - \bar{\epsilon} + \rho + \bar{\rho}, \\
\Delta^{\dagger} &= -\Delta + \gamma + \bar{\gamma} - \mu - \bar{\mu}, \\
\delta^{\dagger} &= -\delta - \beta + \bar{\alpha} + \tau - \bar{\pi}, \\
\bar{\delta}^{\dagger} &= -\bar{\delta} + \alpha - \bar{\beta} - \pi + \bar{\tau},
\end{aligned} \qquad (26)$$

we find that $\mathscr{O}^{\dagger}$ is also a diagonal matrix with $(\Delta + 2\gamma - \bar{\gamma} + \bar{\mu})$ $(D + 3\epsilon + 2p) - (\bar{\delta} + 2\alpha + \bar{\beta} - \bar{\tau})$ $(\delta + 3\beta + 2\tau) - \Psi_2$ and its complex conjugate along the diagonal, and from Eqs. (23), (24), and (26) one can readily obtain

$$\begin{aligned}
&[\bar{J}_{ABC'}J_{A'B'C}]S^{\dagger}\begin{bmatrix} \psi \\ \bar{\psi} \end{bmatrix} \\
&\quad = J^{1'1'1}\{(\delta + 2\beta + \bar{\alpha} - \tau)(\delta + 3\beta + 2\tau)\psi - \bar{\lambda}(D + 3\epsilon + 2\rho)\psi - ig(\bar{\delta} + 3\bar{\beta} + 2\bar{\tau})\bar{\psi}\} \\
&\qquad + J^{1'0'1}\{(\delta + 2\beta - \bar{\alpha})(D + 3\epsilon + 2\rho)\psi + (\bar{\rho} - \rho)(\delta + 3\beta + 2\tau)\psi\} \\
&\qquad + J^{0'1'1}\{(D + 2\epsilon + \bar{\epsilon})(\delta + 3\beta + 2\tau)\psi - (\tau + \bar{\pi})(D + 3\epsilon + 2\rho)\psi\} \\
&\qquad + J^{0'0'1}\{(D + 2\epsilon - \bar{\epsilon} - \rho)(D + 3\epsilon + 2\rho)\psi\} + J^{1'1'0}\{-ig(D + 3\bar{\epsilon} + 2\bar{\rho})\bar{\psi}\} + \text{complex conjugate},
\end{aligned} \qquad (27)$$

from which one can identify the components of $\mathscr{S}^\dagger[\psi\bar{\psi}]$ as the coefficients of $\overline{J}_{ABC'}$ and $J_{A'B'C}$.

Hence from the general discussion presented in Sec. II, we conclude that if ψ is a solution of

$$[(\Delta + 2\gamma - \bar{\gamma} + \bar{\mu})(D + 3\epsilon + 2\rho) - (\bar{\delta} + 2\alpha + \bar{\beta} - \bar{\tau})(\delta + 3\beta + 2\tau) - \Psi_2]\psi = 0, \quad (28)$$

then $\mathscr{S}^\dagger$ acting on the transpose of $[\psi\ \bar{\psi}]$ yields a solution of Eq. (17). The components of this solution, given in (27), can also be written in the form

$$\begin{aligned}
\bar{\psi}_{1'1'1} &= (\delta + 2\beta + \bar{\alpha} + \tau)(\delta + 3\beta)\psi \\
&\quad - \bar{\lambda}(D + 3\epsilon)\psi - ig(\bar{\delta} + 3\bar{\beta} + 2\bar{\tau})\bar{\psi}, \\
\bar{\psi}_{1'0'1} &= (D + 2\epsilon + \bar{\epsilon} + \rho)(\delta + 3\beta)\psi - \bar{\pi}(D + 3\epsilon)\psi, \\
\bar{\psi}_{0'1'1} &= (\delta + 2\beta - \bar{\alpha} + \tau)(D + 3\epsilon)\psi + \bar{\rho}(\delta + 3\beta)\psi, \\
\bar{\psi}_{0'0'1} &= (D + 2\epsilon - \bar{\epsilon} + \rho)(D + 3\epsilon)\psi, \\
\bar{\psi}_{1'1'0} &= -ig(D + 3\bar{\epsilon} + 2\bar{\rho})\bar{\psi}, \\
\bar{\psi}_{1'0'0} &= \bar{\psi}_{0'1'0} = \bar{\psi}_{0'0'0} = 0,
\end{aligned} \quad (29)$$

where we have made use of the Ricci identities

$$\begin{aligned}
&(\delta - \beta + \bar{\alpha} - \tau)\tau = \bar{\lambda}\rho, \quad (\delta - \beta - \bar{\alpha} - \tau)\rho = -\bar{\rho}\tau, \\
&(D - \epsilon + \bar{\epsilon} - \rho)\tau = \bar{\pi}\rho, \quad (D - \epsilon - \bar{\epsilon} - \rho)\rho = 0,
\end{aligned} \quad (30)$$

and of the relation (21).

The field given by Eqs. (27) and (29) is, locally, the most general solution of Eq. (8), modulo the supersymmetry transformations (12). The remaining gauge freedom corresponds to a certain arbitrariness involved in the definition of $\mathscr{S}$: we can add to $\mathscr{S}$ any operator that composed with $\mathscr{E}$ gives zero identically, without altering Eq. (25).[8] By using the Ricci identities with Eqs. (9), (13), and (14) one can verify that

$$\nabla^{CB'} J_{A'B'C} = ig\overline{J}_C{}^C{}_{A'}. \quad (31)$$

Hence the operator $\mathscr{G}$ defined by

$$\mathscr{G}\begin{bmatrix} \overline{J}_{ABC'} \\ J_{A'B'C} \end{bmatrix} \equiv \begin{bmatrix} \overline{\beta}^{A'} \nabla^{CB'} J_{A'B'C} - ig\epsilon^{AB}\overline{\beta}^{C'}\overline{J}_{ABC'} \\ \beta^A \nabla^{BC'} \overline{J}_{ABC'} + ig\epsilon^{A'B'}\beta^C J_{A'B'C} \end{bmatrix}, \quad (32)$$

with β_A arbitrary, is such that $\mathscr{G}\mathscr{E} = 0$, identically. Therefore $\mathscr{S}^\dagger + \mathscr{G}^\dagger$ also yields a solution of Eq. (17); the contribution from $\mathscr{G}^\dagger$ corresponds precisely to the transformation (12) with $\alpha_A = \bar{\psi}\beta_A$. Thus as in the case of the electromagnetic field, where the gauge invariance is related to the charge conservation, the invariance of Eq. (8) under the supersymmetry transformations (12) follows from the "continuity equation" (31).

In the case of a type-D or conformally flat solution of the Einstein vacuum field equations with cosmological constant, Eq. (28) is solvable by separation of variables[4,15,16]; furthermore if $\kappa = \sigma = \nu = \lambda = 0$, from Eqs. (8), (16), and (30) we find that

$$\begin{aligned}
&(D + \epsilon - 2\bar{\epsilon} - 5\rho)(D + 2\epsilon - \bar{\epsilon} - 3\rho)(D + 3\epsilon - \rho)H_{111} \\
&\quad = (\bar{\delta} - \alpha - 2\bar{\beta} + 5\pi)(\bar{\delta} - 2\alpha - \bar{\beta} + 3\pi) \\
&\qquad \times (\bar{\delta} - 3\alpha + \pi)H_{000} - ig\Psi_2\,\overline{H_{000}},
\end{aligned}$$

$$\begin{aligned}
&(\Delta - \gamma + 2\bar{\gamma} + 5\mu)(\Delta - 2\gamma + \bar{\gamma} + 3\mu)(\Delta - 3\gamma + \mu)H_{000} \\
&\quad = (\delta + \beta + 2\bar{\alpha} - 5\tau)(\delta + 2\beta + \bar{\alpha} - 3\tau) \\
&\qquad \times (\delta + 3\beta - \tau)H_{111} + ig\Psi_2\,\overline{H_{111}}
\end{aligned}$$

(cf. Refs. 5 and 16).

By simply taking g, or Λ, equal to zero in the foregoing results, they coincide with those found in Ref. 5 by a different procedure. In fact, the decoupled equation satisfied by H_{000} [Eq. (22)] and the equation for the potential ψ [Eq. (28)] do not contain the cosmological constant explicitly [cf. also Eq. (46) below].

IV. SPIN-$\frac{3}{2}$ PERTURBATIONS OF SOLUTIONS OF THE EINSTEIN–MAXWELL EQUATIONS

In the case of a solution of the Einstein–Maxwell equations, without cosmological constant, a consistent set of equations for spin-$\frac{3}{2}$ fields is obtained from the $N = 2$ extended supergravity.[12,14] These equations involve a doublet of spin-$\frac{3}{2}$ fields and are given by [6,7]

$$\nabla_{AB'}\psi^{jA}{}_{CD'} = \nabla_{CD'}\psi^{jA}{}_{AB'} - i\sqrt{2}\epsilon^{jk}\varphi^A{}_C\bar{\psi}^k{}_{D'B'A}, \quad (33)$$

where $j,k = 1,2$, ϵ^{jk} is the usual Levi-Civita symbol, and φ_{AB} is the electromagnetic spinor, or equivalently in the form

$$H^j{}_{ABC} = H^j{}_{(ABC)}, \quad H^j{}_{AB'C'} = 0, \quad (34)$$

where

$$\begin{aligned}
H^{jA}{}_{BC} &\equiv \nabla_{(B|R'|}\psi^{jA}{}_{C)}{}^{R'} - i\sqrt{2}\epsilon^{jk}\varphi^A{}_{(B}\bar{\psi}^k{}_{|R'|}{}^{R'}{}_{C)}, \\
H^{jA}{}_{B'C'} &\equiv \nabla_{R(B'}\psi^{jAR}{}_{C')} - i\sqrt{2}\epsilon^{jk}\varphi^A{}_R\bar{\psi}^k{}_{(B'C')}{}^R.
\end{aligned} \quad (35)$$

These equations are consistent if the background gravitational and electromagnetic fields satisfy the coupled Einstein–Maxwell equations,

$$\begin{aligned}
&\Phi_{ABC'D'} = 2\varphi_{AB}\bar{\varphi}_{C'D'}, \\
&\Lambda = 0, \\
&\nabla^A{}_{C'}\varphi_{AB} = 0,
\end{aligned} \quad (36)$$

and then they are invariant under the supersymmetry transformations

$$\psi^j{}_{ACD'} \to \psi^j{}_{ACD'} + \nabla_{CD'}\alpha^j{}_A - i\sqrt{2}\epsilon^{jk}\varphi_{AC}\bar{\alpha}^k{}_{D'}. \quad (37)$$

By defining the fields

$$J^j{}_{D'B'C} \equiv \nabla_{AB'}\psi^{jA}{}_{CD'} - \nabla_{CD'}\psi^{jA}{}_{AB'} + i\sqrt{2}\epsilon^{jk}\varphi^A{}_C\bar{\psi}^k{}_{D'B'A}, \quad (38)$$

one finds that

$$H^{jA}{}_{AC} = \tfrac{1}{2}J^{jR'}{}_{R'C}, H^j{}_{AB'C'} = J^j{}_{(B'C')A}, \quad (39)$$

and, using the Ricci identities and Eqs. (36) and (38),

$$\begin{aligned}
\nabla^{AR'}H^j{}_{ABC} &= \Psi_{ABCD}\psi^{jADR'} + i\sqrt{2}\epsilon^{jk}\bar{\psi}^k{}_{S'}{}^{R'A}\nabla_B{}^{S'}\varphi_{AC} \\
&\quad + \nabla_{(B}{}^{S'}J^j{}_{|S'|}{}^{R'}{}_{C)} + i\sqrt{2}\epsilon^{jk}\varphi^A{}_{(B}\overline{J}^k{}_{|A|C)}{}^{R'}.
\end{aligned} \quad (40)$$

Equations (33) can be written in the form

$$\mathscr{E}\begin{bmatrix} \psi^j{}_{ABC'} \\ \bar{\psi}^j{}_{A'B'C} \end{bmatrix} = 0, \quad (41)$$

where

$$\mathscr{E}\begin{bmatrix}\psi^{j}{}_{ABC'}\\ \bar{\psi}^{j}{}_{A'B'C}\end{bmatrix}\equiv\begin{bmatrix}\bar{J}^{j}{}_{ABC'}\\ J^{j}{}_{A'B'C}\end{bmatrix}. \tag{42}$$

This operator is also antiself-adjoint, since

$$\begin{aligned}&[\phi^{jABC'}\quad \bar{\phi}^{jA'B'C}]\mathscr{E}\begin{bmatrix}\psi^{j}{}_{ABC'}\\ \bar{\psi}^{j}{}_{A'B'C}\end{bmatrix}\\ &\quad+[\psi^{jABC'}\quad \bar{\psi}^{jA'B'C}]\mathscr{E}\begin{bmatrix}\phi^{j}{}_{ABC'}\\ \bar{\phi}^{j}{}_{A'B'C}\end{bmatrix}\\ &=\nabla_{AB'}(\bar{\phi}^{jD'B'C}\psi^{jA}{}_{CD'}-\bar{\phi}^{jB'D'A}\psi^{jC}{}_{CD'}\\ &\quad+\phi^{jCAD'}\bar{\psi}^{jB'}{}_{D'C}-\phi^{jACB'}\bar{\psi}^{jD'}{}_{D'C}).\end{aligned} \tag{43}$$

Therefore there exists a divergenceless real vector field, associated with each solution of Eq. (33), which is the sum of two similar terms whose divergence vanish separately when the electromagnetic field vanishes [cf. Eq. (19)].

As shown in Ref. 7, if one of the principal null directions of the background electromagnetic field is geodetic and shear-free, then one can derive a decoupled equation from Eq. (33). By choosing the spin frame in such a way that $\varphi_0=0$ and $\kappa=\sigma=0$, then from Eqs. (36) it follows that $\Psi_0=\Psi_1=0$, and from Eqs. (39) and (40) and the Maxwell equations

$$(D-2\rho)\varphi_1=0,\quad (\delta-2\tau)\varphi_1=0, \tag{44}$$

one finds

$$\begin{aligned}&(\bar{\delta}-3\alpha+\pi)H^{j}{}_{000}-(D-\epsilon-3\rho)H^{j}{}_{100}-\rho(J^{j}{}_{1'0'0}-J^{j}{}_{0'1'0})=\Psi_2\psi^{j}{}_{000'},\\ &\quad+i2\sqrt{2}\epsilon^{jk}\varphi_1(\rho\bar{\psi}^{k}{}_{1'0'0}-\tau\bar{\psi}^{k}{}_{0'0'0})+(\delta-\beta-2\bar{\alpha}+\bar{\pi})J^{j}{}_{0'0'0}\\ &\quad-(D-\epsilon-\bar{\rho})J^{j}{}_{1'0'0}+\bar{\rho}J^{j}{}_{0'1'0}+i\sqrt{2}\epsilon^{jk}\varphi_1\bar{J}^{k}{}_{000'},\\ &(\Delta-3\gamma+\mu)H^{j}{}_{000}-(\delta-\beta-3\tau)H^{j}{}_{100}-\tau(J^{j}{}_{1'0'0}-J^{j}{}_{0'1'0})\\ &\quad=\Psi_2\psi^{j}{}_{001'}+i2\sqrt{2}\epsilon^{jk}\varphi_1(\rho\bar{\psi}^{k}{}_{1'1'0}-\tau\bar{\psi}^{k}{}_{0'1'0})\\ &\quad+(\delta-\beta+\bar{\pi})J^{j}{}_{0'1'0}-(D-\epsilon+2\bar{\epsilon}-\bar{\rho})J^{j}{}_{1'1'0}-\bar{\lambda}J^{j}{}_{0'0'0}+\bar{\pi}J^{j}{}_{1'0'0}+i\sqrt{2}\epsilon^{jk}\varphi_1\bar{J}^{k}{}_{001'}.\end{aligned}$$

Then, using Eq. (21), the complex conjugates of

$$\begin{aligned}&(\bar{\delta}-2\alpha-\bar{\beta}+\pi)\psi^{j}{}_{000'}-(D-\bar{\epsilon}-\rho)\psi^{j}{}_{010'}+\rho\psi^{j}{}_{100'}=i\sqrt{2}\epsilon^{jk}\varphi_1\bar{\psi}^{k}{}_{0'0'0}-J^{j}{}_{0'0'0},\\ &(D+2\epsilon-\bar{\epsilon}-\rho)\psi^{j}{}_{110'}-(\bar{\delta}-\bar{\beta}+\pi)\psi^{j}{}_{100'}-\pi\psi^{j}{}_{010'}+\lambda\psi^{j}{}_{000'}\\ &\quad=i\sqrt{2}\epsilon^{jk}(\varphi_1\bar{\psi}^{k}{}_{0'0'1}-\varphi_2\bar{\psi}^{k}{}_{0'0'0})+J^{j}{}_{0'0'1},\end{aligned}$$

which follow from Eq. (38), and the Bianchi identities

$$(D-3\rho)\Psi_2=2\rho\Phi_{11},\quad (\delta-3\tau)\Psi_2=-2\tau\Phi_{11}+2\rho\Phi_{12},$$

we get the decoupled equations

$$\begin{aligned}&[(D-2\epsilon+\bar{\epsilon}-3\rho-\bar{\rho})(\Delta-3\gamma+\mu)\\ &\quad-(\delta-2\beta-\bar{\alpha}-3\tau+\bar{\pi})(\bar{\delta}-3\alpha+\pi)-\Psi_2]H^{j}{}_{000}=K^{j},\end{aligned}$$

with

$$\begin{aligned}K^{j}\equiv&(D-2\epsilon+\bar{\epsilon}-3\rho-\bar{\rho})[(\delta-\beta-\tau+\bar{\pi})J^{j}{}_{0'1'0}+(\tau+\bar{\pi})J^{j}{}_{1'0'0}\\ &-(D-\epsilon+2\bar{\epsilon}-\bar{\rho})J^{j}{}_{1'1'0}-\bar{\lambda}J^{j}{}_{0'0'0}]-(\delta-2\beta-\bar{\alpha}-3\tau+\bar{\pi})[(\delta-\beta-2\bar{\alpha}+\bar{\pi})J^{j}{}_{0'0'0}+(\bar{\rho}-\rho)J^{j}{}_{0'1'0'}\\ &-(D-\epsilon-\rho-\bar{\rho})J^{j}{}_{1'0'0}]+i\sqrt{2}\epsilon^{jk}\varphi_1[(D-2\epsilon+\bar{\epsilon}+\rho-\bar{\rho})\bar{J}^{k}{}_{001'}-(\delta-2\beta-\bar{\alpha}+\tau+\bar{\pi})\bar{J}^{k}{}_{000'}].\end{aligned}$$

Therefore by defining operators $\mathscr{S}$, $\mathscr{T}$, and $\mathscr{O}$ in an analogous manner to that followed in the preceding section and making use of (26), we get

$$\begin{aligned}&[\bar{J}^{j}{}_{ABC'}J^{j}{}_{A'B'C}]\mathscr{S}^{\dagger}\begin{bmatrix}\psi^{j}\\ \bar{\psi}^{j}\end{bmatrix}\\ &\quad=J^{j1'1'1}\{(\delta+2\beta+\bar{\alpha}-\tau)(\delta+3\beta+2\tau)\psi^{j}-\bar{\lambda}(D+3\epsilon+2\rho)\psi^{j}-i\sqrt{2}\epsilon^{jk}(\bar{\delta}+3\bar{\beta}-2\bar{\tau})\,\overline{\varphi_1\bar{\psi}^{k}}\}\\ &\quad+J^{j1'0'1}\{(\delta+2\beta-\bar{\alpha})(D+3\epsilon+2\rho)\psi^{j}+(\bar{\rho}-\rho)(\delta+3\beta+2\tau)\psi^{j}\}\\ &\quad+J^{j0'1'1}\{(D+2\epsilon+\bar{\epsilon})(\delta+3\beta+2\tau)\psi^{j}-(\tau+\bar{\pi})(D+3\epsilon+2\rho)\psi^{j}\}\\ &\quad+J^{j0'0'1}\{(D+2\epsilon-\bar{\epsilon}-\rho)(D+3\epsilon+2\rho)\psi^{j}\}\\ &\quad+J^{j1'1'0}\{-i\sqrt{2}\epsilon^{jk}(D+3\bar{\epsilon}-2\bar{\rho})\,\overline{\varphi_1\psi^{k}}\}+\text{complex conjugate.}\end{aligned} \tag{45}$$

The operator $\mathscr{S}^{\dagger}$ acting on the transpose of $[\psi^j \ \bar{\psi}^j]$ yields a solution of (41), provided that the complex potentials ψ^j satisfy

$$[(\Delta + 2\gamma - \bar{\gamma} + \bar{\mu})(D + 3\epsilon + 2\rho) - (\bar{\delta} + 2\alpha + \bar{\beta} - \bar{\tau}) \times (\delta + 3\beta + 2\tau) - \Psi_2]\psi^j = 0. \quad (46)$$

By using Eqs. (30) and (44), the solution given implicitly in (45) can also be written in the form

$$\begin{aligned}
\bar{\psi}^j{}_{1'1'1} &= (\delta + 2\beta + \bar{\alpha} + \tau)(\delta + 3\beta)\psi^j \\
&\quad - \bar{\lambda}(D + 3\epsilon)\psi^j - i\sqrt{2}\epsilon^{jk}\,\overline{\varphi_1}(\bar{\delta} + 3\bar{\beta})\bar{\psi}^k, \\
\bar{\psi}^j{}_{1'0'1} &= (D + 2\epsilon + \bar{\epsilon} + \rho)(\delta + 3\beta)\psi^j - \bar{\pi}(D + 3\epsilon)\psi^j, \\
\bar{\psi}^j{}_{0'1'1} &= (\delta + 2\beta - \bar{\alpha} + \tau)(D + 3\epsilon)\psi^j + \bar{\rho}(\delta + 3\beta)\psi^j, \\
\bar{\psi}^j{}_{0'0'1} &= (D + 2\epsilon - \bar{\epsilon} + \rho)(D + 3\epsilon)\psi^j, \\
\bar{\psi}^j{}_{1'1'0} &= -i\sqrt{2}\epsilon^{jk}\,\overline{\varphi_1}(D + 3\bar{\epsilon})\bar{\psi}^k, \\
\bar{\psi}^j{}_{1'0'0} &= \bar{\psi}^j{}_{0'1'0} = \bar{\psi}^j{}_{0'0'0} = 0\,. \qquad (47)
\end{aligned}$$

In the present case, the invariance of Eq. (33) under the transformations (37) can be derived from the identity

$$\nabla^{CB'} J^j{}_{A'B'C} = i\sqrt{2}\epsilon^{jk}\varphi^{DC}\bar{J}^k{}_{DCA'}\,,$$

which follows from Eqs. (36) and (38) and the Ricci identities [cf. Eq. (31)].

When the conformal curvature is of type D and the principal null directions of the electromagnetic field coincide with those of the conformal curvature (as in the case of the Kerr–Newman solution), Eq. (46) can be solved by separation of variables.[6,7,15]

Again, one recovers the results for a vacuum without cosmological constant by simply setting the electromagnetic spinor equal to zero in the results of this section.

V. CONCLUDING REMARKS

The foregoing results show that, in spite of the various versions of the spin-$\frac{3}{2}$ massless field equations given by the linearized supergravity theory, in the cases treated here the complete solution is determined by the solutions of the second-order linear partial differential equation

$$\begin{aligned}
&[(\Delta + (2s+1)\gamma - \bar{\gamma} + \bar{\mu})(D + 2s\epsilon + (2s-1)\rho) \\
&\quad - (\bar{\delta} + (2s-1)\alpha + \bar{\beta} - \bar{\tau})(\delta + 2s\beta + (2s-1)\tau) \\
&\quad - (s-1)(2s-1)\Psi_2]\psi = 0\,, \qquad (48)
\end{aligned}$$

with $s = \frac{3}{2}$ [see Eqs. (28) and (46)]. On the other hand, the solutions of (48) with $s = \frac{1}{2}$, 1, or 2 generate the complete solution of the Weyl neutrino equation, the source-free Maxwell equations, or the linearized Einstein field equations, respectively, in an algebraically special vacuum space-time.[2,3] However, in the case of a background space-time with electromagnetic field and cosmological constant, under the assumptions made in Sec. IV, the solution of the spin-$\frac{3}{2}$ massless field equations given by the linearized $N = 2$ extended supergravity with cosmological constant[17] is determined by two complex potentials that obey a coupled system of two second-order linear partial differential equations.

ACKNOWLEDGMENT

The author acknowledges support from the Sistema Nacional de Investigadores.

[1] S. A. Teukolsky, Astrophys. J. **185**, 635 (1973).
[2] L. S. Kegeles and J. M. Cohen, Phys. Rev. D **19**, 1641 (1979).
[3] R. M. Wald, Phys. Rev. Lett. **41**, 203 (1978).
[4] R. Güven, Phys. Rev. D **22**, 2327 (1980).
[5] G. F. Torres del Castillo, J. Math. Phys. **30**, 446 (1989).
[6] P. C. Aichelburg and R. Güven, Phys. Rev. D **24**, 2066 (1981).
[7] G. F. Torres del Castillo, "Spin-$\frac{3}{2}$ perturbations of algebraically special solutions of the Einstein–Maxwell equations," to appear in J. Math. Phys.
[8] R. M. Wald, Proc. R. Soc. London Ser. A **369**, 67 (1979).
[9] G. F. Torres del Castillo, "Generation of perturbations by means of decoupled equations and their adjoints," to appear in J. Math. Phys.
[10] G. F. Torres del Castillo, Class. Quantum Gravit. **5**, 649 (1988).
[11] P. K. Townsend, Phys. Rev. D **15**, 2802 (1977).
[12] S. Ferrara and P. van Nieuwenhuizen, Phys. Rev. Lett. **37**, 1669 (1976).
[13] R. Penrose and W. Rindler, *Spinors and space-time* (Cambridge U.P., Cambridge, 1984), Vol. 1.
[14] P. van Nieuwenhuizen, in *Superspace and Supergravity*, edited by S. W. Hawking and M. Roček (Cambridge U.P., Cambridge, 1981).
[15] N. Kamran, J. Math. Phys. **26**, 1740 (1985).
[16] G. F. Torres del Castillo, J. Math. Phys. **29**, 2078 (1988).
[17] D. Z. Freedman and A. Das, Nucl. Phys. B **120**, 221 (1977).

On the thermodynamic solution of the Boltzmann equation and nonlinear irreversible thermodynamics

M. Chen
Vanier College, 821 Ste-Croix Blvd., St. Laurent, Quebec H4L 3X9, Canada and Department of Decision Sciences, Concordia University, Montreal, Quebec H3G 1M8, Canada

(Received 10 March 1988; accepted for publication 2 February 1989)

Based on the results of a previous paper on the thermodynamic solution of the Boltzmann equation, some important questions in nonlinear irreversible thermodynamics are reexamined; specifically, the Gibbs relations, the Onsager relations, and the relationship between thermodynamic stability in terms of the entropy balance equation and the dynamical stability of the hydrodynamic equations.

I. INTRODUCTION

Thermodynamics is a macroscopic theory. It is well known that some of the fundamental results in equilibrium thermodynamics can be derived from equilibrium statistical mechanics. The same assertion can also be made for the linear irreversible thermodynamics.[1] However, the problem becomes very difficult in the nonlinear region. Recently, in a series of papers, Eu[2] has proposed an interesting theory of nonlinear irreversible thermodynamics based on the Boltzmann equation.[3] Following Eu's work, in this paper we investigate some of the open questions, such as the Gibbs relation,[1] the Onsager relations,[4] and the relation between thermodynamic stability and dynamic stability, in terms of the thermodynamic solution[5] of the Boltzmann equation which we have obtained recently.

Although the methodology discussed in the following is similar to Glansdorff and Prigogine,[6] and to Eu,[2] there is, however, a major distinction in the structure of the entropy density S. In traditional irreversible thermodynamics, S is assumed to be a differentiable function of the conserved extensive variables, while in Eu's theory, it is assumed to be a differentiable function of the conserved extensive variables as well as the fluxes of stress tensor, heat flow, and mass flow. On the other hand, the entropy density in our approach is a differentiable function of the conserved extensive variables as well as the stationary state of the fluxes, which in turn, are nonlinear functionals of the extensive variables and their spatial gradients. Consequently, the ensuing result that follows from the entropy density can be considered as a bridge between the traditional theory of irreversible thermodynamics and Eu's theory on the extended irreversible thermodynamics.

Now, consider a system of gases with r components contained in an arbitrary region Ω with volume V, where no chemical reactions take place. Let $f_i(t,\mathbf{u}_i,\mathbf{r})$ be the one-particle distribution function of species i with molecular velocity $\mathbf{u}_i$ and at position $\mathbf{r}$. The Boltzmann equation for the system can be written as

$$\frac{\partial f_i}{\partial t} + \mathbf{u}_i \cdot \nabla f_i = \sum_j C(f_i, f_j) , \tag{1}$$

with $C(f_i,f_j)$ representing the collision integral.

Let S be the entropy density, $\mathbf{J}_s$ the entropy current, and σ the entropy production, defined, respectively, by

$$\rho S = -k \sum_i \int_{\Omega_i} d\mathbf{u}_i \, f_i (\log f_i - 1) , \tag{2}$$

$$J_s = -k \sum_i \int_{\Omega_i} d\mathbf{u}_i (\mathbf{u}_i - \mathbf{v}) f_i (\log f_i - 1) , \tag{3}$$

$$\sigma = -k \sum_{i,j} \int_{\Omega_i} d\mathbf{u}_i \, C(f_i, f_j) \log f_i , \tag{4}$$

where k is the Boltzmann constant and Ω_i is an arbitrary region of $\mathbf{u}_i$.

From Eqs. (1)–(4), we can easily obtain the entropy balance equation

$$\rho \frac{d}{dt} S = -\nabla \cdot \mathbf{J}_s + \sigma , \tag{5}$$

where

$$\frac{d}{dt} = \frac{\partial}{\partial t} + \mathbf{v} \cdot \nabla .$$

On the other hand, suppose there exists some f_i, such that ρS, $\mathbf{J}_s$, and σ satisfy the entropy balance equation, and

$$H(t) = \sum_i \int_\Omega d\mathbf{r} \int_{\Omega_i} d\mathbf{u}_i \, f_i \log f_i$$

satisfies the Boltzmann H equation

$$\frac{d}{dt} H = -\sum_i \int_{\Omega_i} d\mathbf{u}_i \int_{\partial\Omega} (\mathbf{u}_i f_i \log f_i) \cdot d\mathbf{A} + \sum_{i,j} \int_\Omega d\mathbf{r} \int_{\Omega_i} d\mathbf{u}_i \, C(f_i f_j) \log f_i ,$$

where the boundary condition on $\partial\Omega$ is described by Cercignani,[7] or by Darozès and Guiraud.[8] Then f_i can be shown to satisfy the Boltzmann equation.[5] Thus the entropy balance equation can be considered as a nonlinear transformation of the Boltzmann equation. Instead of solving the Boltzmann equation directly, we look for some f_i indirectly through Eq. (5), which satisfies the Boltzmann H equation with appropriate boundary conditions. In order to achieve this goal, we assume the functional hypothesis that f_i depends on time t and position $\mathbf{r}$ in terms of the thermohydrodynamic variables $w = \{\rho_i, \mathbf{v}, E, \overleftrightarrow{\Pi}_i, \mathbf{Q}_i', \mathbf{J}_i, \alpha\}$, where E is the energy density, $\mathbf{v}$ is the hydrodynamic velocity, ρ_i, $\overleftrightarrow{\Pi}_i$, $\mathbf{Q}_i'$, and $\mathbf{J}_i$, respectively, are the particle density, the stress tensor, the heat flux, and the mass flux of species i, and α denotes the spatial derivatives of ρ_i, $\mathbf{v}$, E, $\overleftrightarrow{\Pi}_i$, $\mathbf{Q}_i'$, $\mathbf{J}_i$, which in turn are

 0022-2488/89/061329-09$02.50

governed by the following evolution equations:

$$\frac{d}{dt}\rho = -\rho\nabla\cdot\mathbf{v}, \quad \rho\frac{d}{dt}c_i = -\nabla\cdot\mathbf{J}_i, \tag{6}$$

$$\rho\frac{d}{dt}\mathbf{v} = -\nabla\cdot\overleftrightarrow{P}, \tag{7}$$

$$\rho\frac{d}{dt}E = -\nabla\cdot\mathbf{Q} - \overleftrightarrow{P}:\nabla\mathbf{v}, \tag{8}$$

$$\frac{d}{dt}\overleftrightarrow{\Pi}_i = \overleftrightarrow{Z}_i^{(p)} + \sum_j \overleftrightarrow{\Lambda}_{ij}^{(p)}, \tag{9}$$

$$\frac{d}{dt}\mathbf{Q}_i' = \mathbf{Z}_i^{(h)} + \sum_j \mathbf{\Lambda}_{ij}^{(h)}, \tag{10}$$

$$\frac{d}{dt}\mathbf{J}_i = \mathbf{Z}_i^{(d)} + \sum_j \mathbf{\Lambda}_{ij}^{(d)}, \tag{11}$$

where the notations given in Eqs. (6)–(11) and in the subsequent discussions are summarized in the Appendix.

For simplicity, we assume that $\nabla\cdot\mathbf{v} = 0$ in the following discussions. To the linear order of $[\nabla\mathbf{v}]$,[2] ∇T, and $\nabla(\mu_i/T)$, the entropy balance equation yields the following results[5]:

$$f_i = f_i^{(0)}\exp\left\{1 + \frac{m_i}{2P_i kT}[(\mathbf{u}_i-\mathbf{v})(\mathbf{u}_i-\mathbf{v})]^{(2)}:\overleftrightarrow{\Pi}_i + \frac{2m_i}{5k^2P_iT}\left[\frac{1}{2}(\mathbf{u}_i-\mathbf{v})^2 - \frac{5}{2}kT\right](\mathbf{u}_i-\mathbf{v})\cdot\mathbf{Q}_i' + \frac{m_i}{\rho_i kT}(\mathbf{u}_i-\mathbf{v})\cdot\mathbf{J}_i\right\}, \tag{12}$$

$$T\frac{dS}{dt} = \frac{dE}{dt} + p\frac{dv}{dt} - \sum_i \mu_i\frac{dc_i}{dt} + \sum_i\left\{\hat{X}_i^{(P_0)}:\frac{d}{dt}\overleftrightarrow{\Pi}_i + \hat{X}_i^{(h_0)}\cdot\frac{d}{dt}\mathbf{Q}_i' + \hat{X}_i^{(d_0)}\cdot\frac{d}{dt}\mathbf{J}_i\right\}, \tag{13}$$

$$\sigma = \sum_{ij}\{\overleftrightarrow{X}_i^{(P_0)}:\overleftrightarrow{\Lambda}_{ij}^{(P_0)} + \overleftrightarrow{X}_i^{(h_0)}\cdot\mathbf{\Lambda}_{ij}^{(h_0)} + \mathbf{X}_i^{(d_0)}\cdot\mathbf{\Lambda}_{ij}^{(d_0)}\}$$
$$= \sum_i\left\{\mathbf{X}_i^{(P_0)}:\frac{d}{dt}\overleftrightarrow{\Pi}_i + \mathbf{X}_i^{(h_0)}\cdot\frac{d}{dt}\mathbf{Q}_i' + \mathbf{X}_i^{(d_0)}\cdot\frac{d}{dt}\mathbf{J}_i\right\} - \frac{1}{T}\overleftrightarrow{\Pi}:\nabla\mathbf{v} - \frac{1}{T}\mathbf{Q}'\cdot\nabla\log T - \frac{1}{T}\sum_j \mathbf{J}_i\cdot\nabla\mu_i, \tag{14}$$

where $f_i^{(0)}$ is the one-particle distribution function of species i at equilibrium,

$$\hat{X}_1^{(P_0)} = \frac{T}{\rho}\overleftrightarrow{X}_i^{(P_0)} = -\frac{1}{2P_i\rho}\overleftrightarrow{\Pi}_i,$$

$$\hat{X}_i^{(h_0)} = \frac{T}{\rho}\mathbf{X}_i^{(h_0)} = -\frac{2m_i}{5kP_iT\rho}\mathbf{Q}_i',$$

$$\hat{X}_i^{(d_0)} = \frac{T}{\rho}\mathbf{X}_i^{(d_0)} = -\frac{1}{\rho_i\rho}\mathbf{J}_i,$$

and $\overleftrightarrow{\Lambda}_{ij}^{(P_0)}$, $\mathbf{\Lambda}_{ij}^{(h_0)}$, $\mathbf{\Lambda}_{ij}^{(d_0)}$ are obtained from $\overleftrightarrow{\Lambda}_{ij}^{(P)}$, $\mathbf{\Lambda}_{ij}^{(h)}$, and $\mathbf{\Lambda}_{ij}^{(d)}$, respectively, with f_i given by (12).

Since $\hat{X}_i^{(P_0)}$, $\hat{X}_i^{(h_0)}$, and $\hat{X}_i^{(d_0)}$ are independent of ρ, they are intensive variables, and thus, they can be defined as the conjugate variables of $\overleftrightarrow{\Pi}_i$, $\mathbf{Q}_i'$, and $\mathbf{J}_i$, respectively. By Eqs. (2) and (12), the entropy density becomes

$$TS = E + pv - \sum_i \mu_i c_i + \sum_i\{\hat{X}_i^{(P_0)}:\overleftrightarrow{\Pi}_i + \hat{X}_i^{(h_0)}\cdot\mathbf{Q}_i' + \hat{X}_i^{(d_0)}\cdot\mathbf{J}_i\}. \tag{15}$$

Hence S is a first degree homogeneous function of the extensive variables E, v, c_i, $\overleftrightarrow{\Pi}_i$, $\mathbf{Q}_i'$, $\mathbf{J}_i$, and Eq. (13) is a generalized Gibbs formula. Furthermore, by (12) and (14), σ can be shown to be semipositive definite.

It is interesting to note that f_i can be linearized to yield the same expression as given by Grad's thirteen moment method.[9] Moreover, if

$$\frac{d}{dt}\overleftrightarrow{\Pi}_i = \frac{d}{dt}\mathbf{Q}_i' = \frac{d}{dt}\mathbf{J}_i = 0,$$

we then recover the first-order Chapman–Enskog[10] solution of the Boltzmann equation.

In the following sections, we discuss the stationary solutions of the evolution equations (9)–(11) for $\overleftrightarrow{\Pi}_i$, $\mathbf{Q}_i'$, $\mathbf{J}_i$, and their implications to nonlinear irreversible thermodynamics.

II. STATIONARY SOLUTIONS OF $\overleftrightarrow{\Pi}_i$, $\mathbf{Q}_i'$, and $\mathbf{J}_i$

To the linear order of $[\nabla\mathbf{v}]^2$, ∇T, and $\nabla(\mu_i/T)$ in the entropy balance equation, we have obtained a unique f_i given by (12), which not only satisfies the generalized Gibbs formula (13), but also gives rise to a semipositive definite entropy production σ. We define such f_i as the thermodynamic solution of the Boltzmann equation.[5] In view of Eqs. (6)–(11) and (12), ρ_i, $\mathbf{v}$, E, $\overleftrightarrow{\Pi}_i$, $\mathbf{Q}_i'$, and $\mathbf{J}_i$ are governed by the evolution equations

$$\rho\frac{d}{dt}c_i = -\nabla\cdot\mathbf{J}_i, \tag{6'}$$

$$\rho\frac{d}{dt}\mathbf{v} = -\nabla\cdot\overleftrightarrow{P}, \tag{7'}$$

$$\rho\frac{d}{dt}E = -\nabla\cdot\mathbf{Q} - \overleftrightarrow{P}:\nabla\mathbf{v}, \tag{8'}$$

$$\frac{d}{dt}\overleftrightarrow{\Pi}_i = \overleftrightarrow{Z}_i^{(P_0)} + \sum_j \overleftrightarrow{\Lambda}_{ij}^{(P_0)}, \tag{9'}$$

$$\frac{d}{dt}\mathbf{Q}_i' = \mathbf{Z}_i^{(h_0)} + \sum_j \mathbf{\Lambda}_{ij}^{(h_0)}, \tag{10'}$$

$$\frac{d}{dt}\mathbf{J}_i = \mathbf{Z}_i^{(d_0)} + \sum_j \mathbf{\Lambda}_{ij}^{(d_0)}, \tag{11'}$$

where $\overleftrightarrow{Z}_i^{(P_0)} = -2P_i[\nabla\mathbf{v}]^2$, $\mathbf{Z}_i^{(h_0)} = -(5kP_i/2m_i)\nabla T$, and $Z_i^{(d_0)} = \nabla(\mu_i/T)$.

In order to simplify the notations, we introduce the following column vectors H, X, Φ, Z, Λ with $3r-1$ components given by

$$H=[H_1^{(1)},...,H_r^{(1)},H_1^{(2)},...,H_r^{(2)},H_1^{(3)},...,H_{r-1}^{(3)}],$$
$$X=[X_1^{(1)},...,X_r^{(1)},X_1^{(2)},...,X_r^{(2)},X_1^{(3)},...,X_{r-1}^{(3)}],$$
$$\Phi=[\overleftrightarrow{\Pi}_1,...,\overleftrightarrow{\Pi}_r,\mathbf{Q}_1',...,\mathbf{Q}_r',\mathbf{J}_1,...,\mathbf{J}_{r-1}],$$
$$Z=[Z_1^{(1)},...,Z_r^{(1)},Z_1^{(2)},...,Z_r^{(2)},Z_1^{(3)},...,Z_{r-1}^{(3)}],$$
$$\Lambda=[\Lambda_1^{(1)},...,\Lambda_r^{(1)},\Lambda_1^{(2)},...,\Lambda_r^{(2)},\Lambda_1^{(3)},...,\Lambda_{r-1}^{(3)}],$$

where

$$H_i^{(1)}=[m_i(\mathbf{u}_i-\mathbf{v})(\mathbf{u}_i-\mathbf{v})]^{(2)},$$
$$H_i^{(2)}=[\tfrac{1}{2}m_i(\mathbf{u}_i-\mathbf{v})^2-\tfrac{5}{2}kT](\mathbf{u}_i-\mathbf{v}),$$
$$H_i^{(3)}=m_i(\mathbf{u}_i-\mathbf{v}),$$
$$X_i^{(1)}=\mathbf{X}_i^{(P_0)},\quad X_i^{(2)}=\mathbf{X}_i^{(h_0)},\quad X_i^{(3)}=\mathbf{X}_i^{(d_0)},$$
$$Z_i^{(1)}=\overleftrightarrow{Z}_i^{(P_0)},\quad Z_i^{(2)}=\mathbf{Z}_i^{(h_0)},\quad Z_i^{(3)}=\mathbf{Z}_i^{(d_0)},$$

$$\Lambda_i^{(\alpha)}=\sum_j\Lambda_{ij}^{(\alpha)}$$
$$=\sum_j\int d\Gamma_{ij}f_i^{(0)}f_j^{(0)}H_i^{(\alpha)}\{e^{-y_{ij}}-e^{-X_{ij}}\},$$
$$\alpha=1,2,3,$$

and

$$x_{ij}=\frac{1}{kT}\sum_{\beta=1}^{3}\{X_i^{(\beta)}H_i^{(\beta)}+X_j^{(\beta)}H_j^{(\beta)}\},$$
$$y_{ij}=x_{ij}^*=\text{post collision value of } x_{ij},$$
$$d\Gamma_{ij}=\text{integral measure in the Boltzmann collision integral.}$$

The evolution equations for $\overleftrightarrow{\Pi}_i$, $\mathbf{Q}_i'$, and $\mathbf{J}_i$ can then be written as

$$\frac{d}{dt}\Phi_i^{(\alpha)}=\Lambda_i^{(\alpha)}(\rho_j,\mathbf{v},T,\Phi)+Z_i^{(\alpha)}(\rho_j,\mathbf{v},T,\nabla\mathbf{v},\nabla T,\nabla(\mu_j/T)),$$
$$i=1,2,...,3r-1,\quad \alpha=1,2,3, \tag{16}$$

which represents a system of first-order quasilinear partial differential equations with the same principal part

$$\frac{d}{dt}\Phi_i^{(\alpha)}=\frac{\partial}{\partial t}\Phi_i^{(\alpha)}+(\mathbf{v}\cdot\nabla)\Phi_i^{(\alpha)}.$$

The stationary solution $\Phi_i^{(\alpha)}(\text{s.t.})$ of Eq. (16) is defined as the solution of the homogeneous equation

$$\frac{d}{dt}\Phi_i^{(\alpha)}=\frac{\partial}{\partial t}\Phi_i^{(\alpha)}+(\mathbf{v}\cdot\nabla)\Phi_i^{(\alpha)}=0. \tag{17}$$

Thus $\Phi_i^{(\alpha)}(\text{s.t.})$ can be obtained from the system of algebraic equations

$$Z_i^{(\alpha)}=-\Lambda_i^{(\alpha)}(\rho_j,\mathbf{v},T,\Phi(\text{s.t.})). \tag{18}$$

By inverting $\Phi_i^{(\alpha)}(\text{s.t.})$ in terms of ρ_j, $\mathbf{v}$, T, and $Z_i^{(\alpha)}$, we obtain the nonlinear constitutive relations between $\Phi_i^{(\alpha)}\times(\text{s.t.})$ and $Z_1^{(\alpha)}$. On the other hand, the characteristic equations of (17) are given by

$$\frac{dx}{v_x}=\frac{dy}{v_y}=\frac{dz}{v_z}=dt, \tag{19}$$

and $\Phi_i^{(\alpha)}(\text{s.t.})=\text{const}$ on the characteristics. In Eu's theory, the state of a thermodynamic system is described by the extensive variables $(\rho_i,\mathbf{v},E,\overleftrightarrow{\Pi}_i,\mathbf{Q}_i,\mathbf{J}_i)$. Thus the entire set of evolution equations (6′)–(11′) must be considered; however, in view of Eq. (19), $\Phi_i^{(\alpha)}(\text{s.t.})$ is constant on the characteristics. This suggests that we can separate the evolution equations (6′)–(8′) into two parts, Eqs. (6′)–(8′) for the conserved extensive variables, and Eqs. (9′)–(11′) for the fluxes. From a physical point of view, due to the chaotic molecular collisions, $\Phi_i^{(\alpha)}$ changes rapidly in space and time as compared to ρ_i, $\mathbf{v}$, T. Thus $\Phi_i^{(\alpha)}$ reaches the stationary state defined by $\Phi_i^{(\alpha)}(\text{s.t.})$ within a short period of time. During this period of time, ρ_i, $\mathbf{v}$, T can be considered as if they are constant. Once $\Phi_i^{(\alpha)}$ is in the stationary state, $\Phi_i^{(\alpha)}(\text{s.t.})$, then it depends on t and $\mathbf{r}$ in terms of ρ_i, $\mathbf{v}$, T which satisfy Eqs. (6′)–(8′). In other words, we consider two time scales τ_1 and τ_2 with $\tau_1\ll\tau_2$. In the first time scale τ_1, $\Phi_i^{(\alpha)}$ evolves according to Eq. (16) where ρ_i, $\mathbf{v}$, T are almost constant. Once $\Phi_i^{(\alpha)}$ has reached the stationary state, the second time scale τ_2 starts immediately. During this period of time, the evolutions of ρ_i, $\mathbf{v}$, T are described in Eqs. (6′)–(8′). Notice that τ_1 and τ_2 are macroscopic times. They are infinite as compared to the molecular time.

The separation of the two time scales τ_1 and τ_2 has been suggested by Mori *et al.*[11] and Fox[12] in the past. On the other hand, it has also been questioned by Grad[13] and Keizer[14] for the structureless gas molecules. The explanation given above is only a plausible argument from the physical point of view. However, from a mathematical point of view, the separation of the time scale can be construed as a mathematical technique employed to decouple the evolution equations (6′) and (11′).

The physical consideration given above can be formulated as follows. As proved by Courant and Hilbert,[15] the system of first-order quasilinear partial differential equations (16) with the same principal part is equivalent to the system of ordinary differential equations

$$\frac{dx}{v_x}=\frac{dy}{v_y}=\frac{dz}{v_z}=dt,$$
$$\frac{d}{dt}\Phi_i^{(\alpha)}=\Lambda_i^{(\alpha)}(\rho_j,\mathbf{v},T,\Phi)+Z_i^{(\alpha)}\left(\rho_j,\mathbf{v},T,\nabla\mathbf{v},\nabla T,\nabla\frac{\mu_j}{T}\right). \tag{20}$$

Consider the time interval τ_1. Let $\Phi_i^{(\alpha)}=\psi_i^{(\alpha)}+\Phi_i^{(\alpha)}(\text{s.t.})$, where $\psi_i^{(\alpha)}\in\mathcal{L}_2(\Omega)$ such that $\psi_i^{(\alpha)}=0$ on the boundary $\partial\Omega$ of the region $\Omega C R^3$. Denote the Hilbert space X by

$$X\equiv\{\psi=(\psi_1^{(1)},...,\psi_r^{(1)},\psi_1^{(2)},...,\psi_r^{(2)},\psi_1^{(3)},...,\psi_{r-1}^{(3)})\,|\,\psi_i^{(\alpha)}\in\mathcal{L}_2(\Omega),\ \psi_i^{(\alpha)}=0\text{ on }\partial\Omega\}$$

with norm $\|\cdot\|$ given by

$$\|\psi\|^2=\sum_{i,\alpha}\int_\Omega\psi_i^{(\alpha)}\,d\Omega$$

and the inner product $\langle\ ,\ \rangle$ induced by the norm $\|\cdot\|$.

As $\Lambda_i^{(\alpha)}$ is an analytical functional of $\psi_i^{(\alpha)}$, (20) can be rewritten in the form

$$\frac{d}{dt}\psi_i^{(\alpha)} = \sum_{j,\beta} (\tilde{R})_{ij}^{(\alpha\beta)}\psi_j^{(\beta)} + g_i^{(\alpha)}(\psi),$$
$$\psi_i^{(\alpha)}(0) = (\psi_0)_i^{(\alpha)}, \tag{21}$$

where $g_i^{(\alpha)}(\psi)$ is Lipshitz continuous and the matrix elements $\partial g_i^{(\alpha)}/\partial\psi_j^{(\beta)}$ are continuous functions of $\psi_i^{(\alpha)}$ and $\|g(\psi)\|/\|\psi\| \to 0$ as $\|\psi\| \to 0$, while $(\tilde{R})_{ij}^{(\alpha\beta)}$ are functions of ρ_i, $\mathbf{v}$, T, which are considered as constant in the time interval τ_1.

First, we consider the homogeneous equation of (21)

$$\frac{d}{dt}\psi_i^{(\alpha)} = \sum_{j,\beta} (\tilde{R})_{ij}^{(\alpha\beta)}\psi_j^{(\beta)},$$
$$\psi_i^{(\alpha)}(0) = (\psi_0)_i^{(\alpha)}. \tag{22}$$

By examining the property of $(\tilde{R})_{ij}^{(\alpha\beta)}$ we notice that, except for the negative sign, $(\tilde{R})_{ij}^{(\alpha\beta)}$ is related to the coefficients in the linear constitutive relations between the fluxes and the thermodynamic forces. Thus $\tilde{R}$ is a negative definite matrix $\langle\psi, \tilde{R}\psi\rangle < 0$ *and* $\langle\psi, \tilde{R}\psi\rangle = 0$ if and only if $\psi = 0$. It is obvious that $\psi = 0$ is an equilibrium solution of (22). Define a continuous function $V: X \to R$ by $V(\psi) = \|\psi\|^2$. Then $(d/dt)V = 2\langle\psi, \tilde{R}\psi\rangle < 0$, which implies that V is a continuous Liapunov function for the linear dynamical system generated by $\tilde{R}$ in X; hence $\psi = 0$ is asymptotically stable.

Next, we consider the nonlinear equation (21). As g is Lipshitz continuous in X, (21) has a unique solution satisfying the initial condition ψ_0. Again, let $V: X \to R$ be given by $V(\psi) = \|\psi\|^2$. Then

$$\frac{d}{dt}V(\psi) = 2\langle\psi, \tilde{R}\psi + g(\psi)\psi\rangle$$
$$= 2\langle\psi, \tilde{R}\psi\rangle + 2\langle\psi, g(\psi)\psi\rangle.$$

Since $\|g(\psi)\|/\|\psi\| \to 0$ as $\|\psi\| \to 0$, by the Lipshitz continuity of g, there exists a neighborhood N of $\psi = 0$ such that $2\langle\psi, \tilde{R}\psi\rangle + \langle\psi, g(\psi)\psi\rangle \leqslant -\alpha\|\psi\|^2$, where α is a positive real number. Thus $(d/dt)V < -\alpha\|\psi\|^2$ for all $\psi \in N \setminus \{0\}$ and, consequently, V is a Liapunov function for the dynamical system generated by Eq. (21) in X. Therefore, the equilibrium solution $\psi = 0$ of (21) is asymptotically and exponentially stable.

Now that $\Phi \to \Phi(\text{s.t.})$ exponentially, the entropy production σ in (14) reduces to

$$\sigma_{\text{s.t.}} = \sum_{i,j} \{\overleftrightarrow{X}_i^{(p_0)} : \overleftrightarrow{\Lambda}_{ij}^{(p_0)} + \mathbf{X}_i^{(h_0)} \cdot \mathbf{\Lambda}_{ij}^{(h_0)}$$
$$+ \mathbf{X}_i^{(d_0)} \cdot \mathbf{\Lambda}_{ij}^{(d_0)}\}\text{s.t.}$$
$$= \overleftrightarrow{\Pi}(\text{s.t.}) : (-T^{-1}\nabla\mathbf{v}) + \mathbf{Q}'(\text{s.t.}) \tag{23}$$
$$\cdot(-T^{-1}\nabla \log T) + \sum_i \mathbf{J}_i(\text{s.t.})$$
$$(-T^{-1}\nabla\mu_i) \geqslant 0,$$

and the entropy density S in (15) reduces to

$$TS = E + p\nu - \sum_i \mu_i c_i + \sum_{i,\alpha} \{\hat{X}_i^{(\alpha)}\Phi_i^{(\alpha)}\}\text{s.t.}, \tag{24}$$

while the generalized Gibbs formula (13) becomes

$$T\frac{dS}{dt} = \frac{dE}{dt} + p\frac{d\nu}{dt} - \sum_i \mu_i \frac{dC_i}{dt} + \sum_{i,\alpha} \{\hat{X}_i^{(p_0)} : \frac{d}{dt}\overleftrightarrow{\Pi}_i$$
$$+ \hat{X}_i^{(h_0)} \cdot \frac{d}{dt}\mathbf{Q}_i' + \hat{X}_i^{(d_0)} \cdot \frac{d}{dt}\mathbf{J}_i\}\text{s.t.}, \tag{25}$$

where $\{\}$s.t. is evaluated at $\Phi_i^{(\alpha)} = \Phi_i^{(\alpha)}(\text{s.t.})$.

We emphasize that $\Phi_i^{(\alpha)}(\text{s.t.})$ is now a function of ρ_i, $\mathbf{v}$, T, and the thermodynamic forces $Z_i^{(\alpha)}$.

III. NONLINEAR IRREVERSIBLE THERMODYNAMICS

Once $\Phi_i^{(\alpha)}$ has reached the stationary state defined by $\Phi_i^{(\alpha)}(\text{s.t.})$, the second time scale τ_2 starts immediately. During this interval of time, $\Phi_i^{(\alpha)}(\text{s.t.})$ depends on t and $\mathbf{r}$ in terms of ρ_i, $\mathbf{v}$, T, and $Z_i^{(\alpha)}$, which in turn are governed by Eqs. (6′)–(8′). In the meantime, $\sigma_{\text{s.t.}}$ is given by (23), and the entropy density is a first degree homogeneous function of E, ν, c_i, $\Phi_i^{(\alpha)}(\text{s.t.})$ given in Eq. (24). We now consider the consequences of $\Phi_i^{(\alpha)}(\text{s.t.})$ in nonlinear irreversible thermodynamics.

A. Onsager relations

By definition

$$Z_i^{(\alpha)} = -\Lambda_i^{(\alpha)} = -\sum_j \int d\Gamma_{ij} f_i^{(0)} f_j^{(0)} H_i^{(\alpha)}\{e^{-v_{ij}} - e^{-x_{ij}}\},$$
$$\alpha = 1,2,3. \tag{26}$$

If $\Lambda_i^{(\alpha)}$ is linearized, we then obtain

$$Z_i^{(1)} = (A)_{ii}^{(1,1)} X_i^{(1)}(\text{s.t.})$$
$$+ \sum_{j \neq i} (A)_{ij}^{(1,1)} X_j^{(1)}(\text{s.t.}), \tag{27}$$

$$Z_i^{(\beta)} = \sum_\gamma (A)_{ii}^{(\beta,\gamma)} X_i^{(\gamma)}(\text{s.t.})$$
$$+ \sum_{j \neq i} \sum_\gamma (A)_{ij}^{(\beta,\gamma)} X_j^{(\gamma)}(\text{s.t.}), \quad \beta, \gamma = 2,3, \tag{28}$$

where

$$(A)_{ij}^{(1,1)} = (A)_{ji}^{(1,1)}, \quad (A)_{ii}^{(\beta,\gamma)} = (A)_{ii}^{(\gamma,\beta)},$$
$$(A)_{ij}^{(\beta,\gamma)} = (A)_{ji}^{(\gamma,\beta)}. \tag{29}$$

Expression (29) represents the Onsager relations in linear irreversible thermodynamics. In general, by (26) we can derive the following symmetry relations:

$$\frac{\partial\Lambda_j^{(1)}}{\partial X_i^{(1)}(\text{s.t.})} = \frac{\partial\Lambda_i^{(1)}}{\partial X_j^{(1)}(\text{s.t.})}, \tag{30}$$

$$\frac{\partial\Lambda_j^{(\beta)}}{\partial X_i^{(\gamma)}(\text{s.t.})} = \frac{\partial\Lambda_i^{(\gamma)}}{\partial X_j^{(\beta)}(\text{s.t.})}, \quad \beta, \gamma = 2,3. \tag{31}$$

Thus, in any order of approximation of $\Lambda_i^{(\alpha)}$, the symmetry relations given above must be retained. These relations are useful in the study of transport coefficients.

B. Nonlinear thermodynamic relations

In view of the nonlinear constitutive relations between $\Phi_i^{(\alpha)}(\text{s.t.})$ and $Z_i^{(\alpha)}$, the entropy density S is also a nonlinear function of $Z_i^{(\alpha)}$. This viewpoint is different from that of Eu[2] or Glansdorff and Prigogine,[6] where S is assumed to be

independent of the thermodynamic forces $Z_i^{(\alpha)}$. Particularly, Glansdorff and Prigogine assumed that S was a differentiable function of E, v, and c_i. However, this is mainly due to the constraint imposed on the Chapman–Enskog method in solving the Boltzmann equation.

By Eq. (24), S is a differentiable function of E, v, c_i, and $\Phi_i^{(\alpha)}(\text{s.t.})$, where $\hat{X}_i^{(\alpha)}(\text{s.t.})$ is the conjugate variable of $\Phi_i^{(\alpha)}(\text{s.t.})$. Thus

$$T\,dS = dE + p\,dv - \sum_i \mu_i\,dc_i + \sum_{i,\alpha} \{\hat{X}_i^{(\alpha)} d\Phi_i^{(\alpha)}\}\text{s.t.} \tag{32}$$

Note that the exact one-form $T\,dS$ given by (32) is different from the expression given by Eu, or the expression considered by Glansdorff and Prigogine. Particularly, Glansdorff and Prigogine considered the one-form

$$T\,dS = dE + p\,dv - \sum_i \mu_i\,dc_i\,. \tag{33}$$

By the Legendre transformations, we now define the following nonequilibrium thermodynamic functions:

Enthalpy function,

$$\overline{H} = E + pv = \overline{H}(S, p, c_i, \Phi_i^{(\alpha)}(\text{s.t.}))\,,$$

Helmholtz function,

$$F = E - TS = F(T, v, c_i, \Phi_i^{(\alpha)}(\text{s.t.}))\,,$$

and Gibbs function,

$$G = \overline{H} - TS = G(T, p, c_i, \Phi_i^{(\alpha)}(\text{s.t.}))\,.$$

By the differential form (32), we can obtain the exact differential forms of $\overline{H}$, F, and G as follows:

$$d\overline{H} = T\,dS + v\,dp + \sum_i \mu_i\,dc_i - \sum_{i,\alpha} \{\hat{X}_i^{(\alpha)}\,d\Phi_i^{(\alpha)}\}_{\text{s.t.}}\,, \tag{34}$$

$$dF = -S\,dT - p\,dv + \sum_i \mu_i\,dc_i - \sum_{i,\alpha} \{\hat{X}_i^{(\alpha)}\,d\Phi_i^{(\alpha)}\}_{\text{s.t.}}\,, \tag{35}$$

$$dG = -S\,dT + v\,dp + \sum_i \mu_i\,dc_i - \sum_{i,\alpha} \{\hat{X}_i^{(\alpha)}\,d\Phi_i^{(\alpha)}\}_{\text{s.t.}}\,. \tag{36}$$

Finally, if we define

$$\Xi = TS - pv - E + \sum_i \mu_i c_i = \Xi(T, p, \mu_i, \Phi_i^{(\alpha)}(\text{s.t.}))\,,$$

then, by Eqs. (24) and (32) we have

$$d\Xi = S\,dT - v\,dp + \sum_i c_i\,d\mu_i + \sum_{i,\alpha} \{\hat{X}_i^{(\alpha)}\,d\Phi_i^{(\alpha)}\}_{\text{s.t.}} = \sum_{i,\alpha} \{\hat{X}_i^{(\alpha)}\,d\Phi_i^{(\alpha)} + \Phi_i^{(\alpha)}\,d\hat{X}_i^{(\alpha)}\}_{\text{s.t.}}\,,$$

which implies

$$\sum_{i,\alpha} \{\Phi_i^{(\alpha)}\,d\hat{X}_i^{(\alpha)}\}_{\text{s.t.}} = S\,dT - v\,dp + \sum_i c_i\,d\mu_i\,. \tag{37}$$

It is interesting to note that differential form (37) is a nonequilibrium generalization of the Gibbs–Duhem relation[16] in equilibrium thermodynamics. Moreover, by (37), $\hat{X}_i^{(\alpha)}(\text{s.t.})$ can be considered as a function of T, p, μ_i, and $\Phi_i^{(\alpha)}(\text{s.t.})$. The partial derivatives of $\hat{X}_i^{(\alpha)}(\text{s.t.})$ with respect to these variables will be useful in the discussion of thermodynamic stability.

From the differential forms (32), (34)–(37), we can obtain a set of nonequilibrium Maxwell relations. These relations are similar to the expressions given in Eu's paper if $X_i^{(\alpha)}$ and $\phi_i^{(\alpha)}$ in Eu's paper are replaced by $\hat{X}_i^{(\alpha)}(\text{s.t.})$ and $\Phi_i^{(\alpha)}(\text{s.t.})$, respectively.

C. The Gibbs–Duhem theory of thermodynamic stability

The nonequilibrium differential forms given in (32), (34)–(37) are generalizations of the corresponding differential forms in equilibrium thermodynamics. We now generalize the thermodynamic stability theory of Gibbs and Duhem[17] to nonlinear irreversible thermodynamics by means of these differential forms.

According to the second law of thermodynamics, the total entropy increases toward the maximum value at thermodynamic equilibrium. Therefore, the thermodynamic equilibrium is stable if and only if the total entropy decreases from its equilibrium value for all possible variations of the thermodynamic variables describing the thermodynamic state of the system. The Gibbs–Duhem criterion of thermodynamic stability can be formulated as

$$\delta^2 \mathscr{S} = \int_\Omega d\mathbf{r}\,\rho\,\delta^2 S < 0\,,$$

$$\delta^2 \mathscr{S} = 0 \text{ at thermodynamic equilibrium}\,.$$

Since S is a differentiable function of E, v, c_i, $\Phi_i^{(\alpha)}(\text{s.t.})$, the thermodynamic state of the system can be specified by these variables. Thus

$$T\,\delta S = \delta E + p\,dv - \sum_i \mu_i\,\delta c_i + \sum_{i,\alpha} \hat{X}_i^{(\alpha)}(\text{s.t.})\delta\Phi_i^{(\alpha)}(\text{s.t.}) \tag{38}$$

$$T\,\delta^2 S = -\delta T\,\delta S + \delta p\,\delta v - \sum_i \delta\mu_i\,\delta c_i + \sum_{i,\alpha} \delta\hat{X}_i^{(\alpha)}(\text{s.t.})\delta\Phi_i^{(\alpha)}(\text{s.t.})\,. \tag{39}$$

By the nonequilibrium Maxwell relations derived from the differential forms (32), (34)–(37), Eq. (39) can be simplified as

$$T\,\delta^2 S = -\delta T(\delta S)_\phi + \delta p(\delta v)_\phi - \sum_i \delta\mu_i(\delta c_i)_\phi + \sum_{i,\alpha} \left[\frac{\partial \hat{X}_i^{(\alpha)}(\text{s.t.})}{\partial \phi_i^{(\alpha)}(\text{s.t.})}\right] \delta\Phi_i^{(\alpha)}(\text{s.t.})\,, \tag{40}$$

where

$$(\delta S)_\phi = \delta S - \sum_{i,\alpha}\left[\frac{\partial S}{\partial \phi_i^{(\alpha)}(\text{s.t.})}\right]\delta\Phi_i^{(\alpha)}(\text{s.t.}),$$

$$(\delta \nu)_\phi = \delta \nu - \sum_{i,\alpha}\left[\frac{\partial \nu}{\partial \phi_i^{(\alpha)}(\text{s.t.})}\right]\delta\Phi_i^{(\alpha)}(\text{s.t.}),$$

$$(\delta c_i)_\phi = \delta c_i - \sum_{i,\alpha}\left[\frac{\partial c_i}{\partial \phi_j^{(\alpha)}(\text{s.t.})}\right]\delta\Phi_j^{(\alpha)}(\text{s.t.}),$$

and the variables which do not appear in the differentiation in [ ] are held constant as is customarily done in equilibrium thermodynamics.

We notice that the variations given in (38)–(40) are completely arbitrary. Thus our formulation is static in nature. Now, define the specific heat at constant volume C_v and the thermocompressibility constant κ by

$$\left[\frac{\partial S}{\partial T}\right]_{\nu,\, c_i,\, \phi_i^{(\alpha)}(\text{s.t.})} \equiv C_v T^{-1},$$

$$\left[\frac{\partial p}{\partial \nu}\right]_{T,\, c_i,\, \phi_i^{(\alpha)}(\text{s.t.})} \equiv -\nu^{-1}\kappa^{-1}.$$

Equation (40) finally becomes

$$\begin{aligned} T\delta^2 S = &-T^{-1}C_v(\delta T)^2 - \nu^{-1}\kappa^{-1}(\delta\nu)_\phi^2 \\ &- \sum_{i,j}\left[\frac{\partial \mu_i}{\partial c_j}\right](\delta c_i)_\phi(\delta c_j)_\phi \\ &+ \sum_{i,\alpha}\left[\frac{\partial \hat{X}_i^{(\alpha)}(\text{s.t.})}{\partial \Phi_i^{(\alpha)}(\text{s.t.})}\right] \\ &\times(\delta\Phi_i^{(\alpha)}(\text{s.t.}))(\delta\phi_i^{(\alpha)}(\text{s.t.})). \end{aligned} \tag{41}$$

Since T, C_v, κ are all positive real numbers, while

$$\left[\frac{\partial \hat{X}_i^{(\alpha)}(\text{s.t.})}{\partial \phi_i^{(\alpha)}(\text{s.t.})}\right] = \begin{cases} -1/2p_i\rho, & \text{for } \alpha = 1, \\ -2m_i/5kp_i\rho, & \text{for } \alpha = 2, \\ -1/p_i\rho, & \text{for } \alpha = 3, \end{cases}$$

thus $T\delta^2 S < 0$ and consequently $\delta^2 \mathcal{S} < 0$, if

$$\sum_{i,j}\left[\frac{\partial \mu_i}{\partial c_j}\right](\delta c_i)_\phi(\delta c_j)_\phi > 0, \tag{42}$$

while $T\delta^2 S = 0$ if $\delta T = (\delta\nu)_\phi = (\delta c_i)_\phi = \delta\Phi_i^{(\alpha)}(\text{s.t.}) = 0$ at thermodynamic equilibrium.

First, we note that condition (42) is a generalization of the stability condition with respect to diffusion in equilibrium thermodynamics.[19] Second the formulation discussed above is based on the differential forms (32) and (34)–(37). Except for some minor changes, we have recovered the result of Glansdorff and Prigogine, or the result of Eu.

In view of the results discussed in this section, it is evident that the stationary state of the fluxes $\Phi_i^{(\alpha)}(\text{s.t.})$ plays an important role in nonlinear irreversible thermodynamics. As a matter of fact, the dependence of thermodynamic potentials on the fluxes was first considered by Keizer[18] in a molecular kinetic theory of dissipations and fluctuations of transport equations. His proposal has recently been supported by some experimental evidence.[19] Insofar as linear irreversible thermodynamics is concerned, our result is similar to Keizer's. However, it would be difficult to compare the results in the nonlinear regime because the two methodologies and assumptions are different.

IV. THE RELATIONSHIP BETWEEN THERMODYNAMIC STABILITY AND DYNAMICAL STABILITY

In this section we consider the problem of thermodynamic stability from the dynamical point of view in terms of the evolution equations (6′)–(8′). Since $\Phi_i^{(\alpha)}(\text{s.t.})$ is a function of ρ_i, $\mathbf{v}$, E, and $Z_i^{(\alpha)}$, by (38) we can obtain the evolution equation of $\rho\,\delta^2 S$ as

$$\begin{aligned} \rho\frac{d}{dt}(\delta^2 S) = &(\delta T^{-1})\rho\frac{d}{dt}(\delta E) + \delta(pT^{-1})\rho\frac{d}{dt}(\delta\nu) \\ &- \sum_i \delta(\mu_i T^{-1})\rho\frac{d}{dt}(\delta c_i) \\ &+ \sum_{i,\alpha}[\delta X_i^{(\alpha)}(\text{s.t.})]\frac{d}{dt}\delta\Phi_i^{(\alpha)}(\text{s.t.}), \end{aligned} \tag{43}$$

where the fluctuations δE, $\delta\nu$, δc_i, and $\delta\Phi_i^{(\alpha)}(\text{s.t.})$ are variations from their equilibrium values and obey the variational evolution equations

$$\rho\frac{d}{dt}\delta c_i = -\nabla\cdot\delta\mathbf{J}_i(\text{s.t.}), \tag{44}$$

$$\rho\frac{d}{dt}\delta\mathbf{v} = -\nabla\cdot\delta\overleftrightarrow{P}(\text{s.t.}), \tag{45}$$

$$\rho\frac{d}{dt}\delta E = -\nabla\cdot\delta\mathbf{Q}(\text{s.t.}) - \delta(\overleftrightarrow{P}(\text{s.t.}):\nabla\mathbf{v}), \tag{46}$$

$$\delta Z_i^{(\alpha)} = -\delta\Lambda_i^{(\alpha)}(\rho_j, \mathbf{v}, T, \Phi(\text{s.t.})). \tag{47}$$

It is important to notice that in (44)–(46), ρ is the density weight of the independent variables δc_i, $\delta\mathbf{v}$, and δE. Thus ρ is unvariated. By (43)–(46), the evolution equation for $\rho\,\delta^2 S$ finally becomes

$$\begin{aligned} \frac{d}{dt}(\rho\,\delta^2 S) = &-\nabla\cdot[\delta Q(\text{s.t.})(\delta T^{-1})] + \sum_i \nabla\cdot[\delta(\mu_i T^{-1})\delta\mathbf{J}_i(\text{s.t.})] - \delta\mathbf{Q}(\text{s.t.})\cdot\delta(T^{-1}\nabla\log T) - \delta\overleftrightarrow{\Pi}(\text{s.t.}):\delta(T^{-1}\nabla\mathbf{v}) \\ &+ \sum_i \delta\mathbf{J}_i(\text{s.t.})\cdot\nabla(\mu_i T^{-1}) + \sum_{i,\alpha}[\delta X_i^{(\alpha)}(\text{s.t.})]\frac{d}{dt}\delta\Phi_i^{(\alpha)}(\text{s.t.}), \end{aligned} \tag{48}$$

which can be considered as a second-order entropy balance equation. Let $\delta\mathbf{Q}(\text{s.t.}) = \delta\mathbf{J}_i(\text{s.t.}) = 0$ on the boundary $\partial\Omega$ of the region Ω. Then

$$\frac{d}{dt}\delta^2\mathscr{S} = \int_\Omega \frac{d}{dt}(\rho\,\delta^2 S)d\Omega$$
$$= \int_\Omega d\Omega\left\{ -\delta\mathbf{Q}(\text{s.t.})\cdot\delta(T^{-1}\nabla\log T) - \delta\overleftrightarrow{\Pi}(\text{s.t.}):\ \delta(T^{-1}\nabla\mathbf{v}) + \sum_i \delta\mathbf{J}_i(\text{s.t.})\cdot\nabla(\mu_i T^{-1})\right\}$$
$$+ \int_\Omega d\Omega \sum_{i,\alpha}\left(\delta X_i^{(\alpha)}(\text{s.t.})\,\frac{d}{dt}\delta\Phi_i^{(\alpha)}(\text{s.t.})\right). \tag{49}$$

Following Glansdorff and Prigogine, the first term on the right-hand side of (49) is called the generalized excess entropy production. In view of (23), this term is semipositive definite and it vanishes at thermodynamic equilibrium. On the other hand, the second term is essentially determined by the nonlinear constitutive relations (47) and the variational evolution equations (44)–(46).

According to the second law of thermodynamics, the total entropy increases toward the maximum value at thermodynamic equilibrium. Hence $\delta^2\mathscr{S}\leqslant 0$. Suppose, by the variational evolution equations (44)–(47),

$$\sum_{i,\alpha}\int_\Omega d\Omega\left[\delta X_i^{(\alpha)}(\text{s.t.})\right]\frac{d}{dt}\delta\Phi_i^{(\alpha)}(\text{s.t.})\geqslant 0\,, \tag{50}$$

then $(d/dt)\delta^2\mathscr{S}\geqslant 0$, and the thermodynamic equilibrium is stable. Equation (50) is therefore a sufficient condition for thermodynamic stability.

On the other hand, if Eqs. (44)–(46) can be formulated as a dynamical system, and if condition (50) holds, then $-\delta^2\mathscr{S}\geqslant 0$ and $(d/dt)(-\delta^2\mathscr{S})\leqslant 0$. In other words, $-\delta^2\mathscr{S}$ is a Liapunov function for the dynamical system and, consequently, the thermodynamic equilibrium state is asymptotically stable.

In order to illustrate this argument, let us, for simplicity, consider the linearized hydrodynamic equations for an isotropic single component system. Equations (44)–(46) then reduce to

$$\partial_t\rho = -\gamma\,\partial_i v_i\,, \tag{51}$$
$$\partial_t v_i = -\beta\,\partial_i T - \alpha\,\partial_i\rho + \nu\,\partial_j\partial_j v_i + \delta\,\partial_i\partial_j v_j\,, \tag{52}$$
$$\partial_t T = -\mu\,\partial_i v_i + \partial_i\partial_i T\,, \tag{53}$$

where $i,j = 1,2,3$; $\gamma,\beta,\nu,\delta,\mu$, are positive constants, while α is a negative constant. Here, ρ, v_i, T are variations from their corresponding equilibrium values, and repeated index means summation, i.e., $\partial_j v_j = \nabla\cdot\mathbf{v}$.

Let Ω be an open, regular, connected, and bounded region. Denote $x = (\rho, v_1, v_2, v_3, T)$ and

$$X = \mathcal{L}_2(\Omega)\times(\mathcal{L}_2(\Omega))^3\times\mathcal{L}_2(\Omega)\,,$$

with norm

$$\|x\| = \left\{\int_\Omega(\gamma^{-1}\rho^2 + \alpha^{-1}v_i v_i + \mu^{-1}T^2)d\Omega\right\}^{1/2}$$

and inner product $\langle\,,\rangle$ induced by the norm $\|\cdot\|$.

Define a linear operator A: $D(A)\subset X\to X$ by

$$\partial_t x = Ax\,,$$

where the components of Ax are given by the right-hand side of (51)–(53), respectively. The boundary conditions are set up as follows:

$$x = 0,\quad \partial_i\rho = \partial_i T = \partial_i v_j = 0 \text{ on } \partial\Omega\,.$$

The linearized hydrodynamic equations (51)–(53) can be reformulated as an abstract evolution equation

$$\partial_t x = Ax\,,\qquad x(0) = x_0\in D(A)\,, \tag{54}$$

where the domain $D(A)$ is given by

$$D(A) = \{x\in X\,|\,\partial_i\rho\in\mathcal{L}_2(\Omega),\ \partial_i\partial_j v_k\in\mathcal{L}_2(\Omega),\ \partial_i\partial_j T\in\mathcal{L}_2(\Omega);$$
$$x = 0,\ \partial_i v_j = \partial_i\rho = \partial_i T = 0 \text{ on } \partial\Omega\}$$
$$\equiv \hat{W}^2(\Omega)\times(\hat{W}_2^2(\Omega))^3\times\hat{W}_2^2(\Omega)\,.$$

First, it can be shown easily that $D(A)$ is dense in X. Second, we notice that

$$\langle x, Ax\rangle = \int_\Omega d\Omega\{(\beta\alpha^{-1} - 1)T\,\partial_i v_i - \nu\alpha^{-1}(\partial_j v_i)(\partial_j v_i)$$
$$- \delta\alpha^{-1}(\partial_i v_i)(\partial_j v_j) - \xi\mu^{-1}(\partial_i T)(\partial_i T)\}\,. \tag{55}$$

Suppose there exists a uniform temperature T_0 such that

$$\int_\Omega d\Omega\, T\,\partial_i v_i \geqslant T_0\int_\Omega d\Omega\,\partial_i v_i\,.$$

Then, by the boundary condition, $\int_\Omega d\Omega\, T\,\partial_i v_i\geqslant 0$ and thus $\langle x, Ax\rangle\leqslant 0$. This implies that $-A$ is an accretive operator.[20]

Finally, by taking the Fourier transform of (51)–(53), we can show that, for sufficiently small positive λ, the range of $I-\lambda A$ is X, i.e., $R(I-\lambda A) = X$.

Since $D(A)$ is dense, $-A$ is accretive, and $R(I-\lambda A) = X$ for sufficiently small positive λ, A is the infinitesimal generator[21] of a linear dynamical system $\{U(t)\}_{t\geqslant 0}$ on X such that $\|U(t)x\|\leqslant\|x\|$ for all $x\in X$ and $t\geqslant 0$. This implies that all motions $x(t) = U(t)x_0$ are bounded.

Next, we define a function V: $X\to R$ by $V(x) = \|x\|^2$. Then $\dot{V}(x) = 2\langle x,\dot{x}\rangle = 2\langle x, Ax\rangle\leqslant 0$. Let W: $X\to\bar{R}$ (the extended real number system) be a lower semicontinuous function defined by

$$W(x) = -2\langle x, Ax\rangle\geqslant 0 \text{ for all}$$
$$x\in X\cap\{\mathcal{L}^2(\Omega)\times(\hat{W}_2^1(\Omega))^3\times\hat{W}_2^1(\Omega)\}\,,$$
$$W(x) = \infty \text{ if } x\in X \text{ but}$$
$$v_i(x)\notin\hat{W}_2^1(\Omega),\ T\notin\hat{W}_2^1(\Omega)\,.$$

Then V is a continuous Liapunov function on X. Consider the set $M_2 = \{x\in X\,|\,W(x) = 0\}$. As ρ, v_i, and T are independent variables, $W(x) = 0$ implies $\partial_i v_j = \partial_i T = 0$. By the boundary condition, we have $v_i = T\equiv 0$. Hence $M_2 = \{x\in X\,|\,x = (\rho, 0, 0)\}$. Since $M_2\not\subset D(A)$, the evolution equation $\dot{x}(t) = Ax(t)$ does not apply to some of the motions originating in M_2. However, by employing the theory

of distributions,[22] we can obtain a distributional evolution equation satisfied by all motions of $\{U(t)\}_{t>0}$, and from this equation it can be concluded that the largest positive invariant set M^+ in $M_2 \cap D(A)$ is $M^+ = \{0\}$. On the other hand, we notice that the equilibrium solution x_e of the evolution equation (54) is $x_e = 0$.

Let $\hat{I}$ be the injection operator defined by $\hat{I}$: $(D(A), \|\cdot\|_g) \to X$, where $\|\cdot\|_g$ denotes the graph norm. By the Sobolev embedding theorem, $\hat{W}_2^1(\Omega) x (\hat{W}_2^2(\Omega))^3 x (\hat{W}_2^2(\Omega)) \subset \angle_2(\Omega) x (\angle_2(\Omega))^3 x \angle_2(\Omega)$, $\hat{I}$ is a compact operator.[23] Hence for sufficiently small positive λ, $J_\lambda \equiv (I - \lambda A)^{-1}: X \to X$ is also a compact operator.[24] Since A is the infinitesimal generator of the linear dynamical system $\{U(t)\}_{t>0}$ and J_λ is compact, all positive orbits $\gamma(x)$ are precompact.[25] By the invariance principle,[26] $U(t)x \to M^+ = \{0\}$ as $t \to \infty$. Consequently, the thermodynamic equilibrium state $x_e = 0$ is globally asymptotically stable.

When the thermodynamic system is near the equilibrium state, the fluxes $\delta\phi^{(\alpha)}$(s.t.) are proportional to the thermodynamic forces $\delta Z^{(\alpha)}$, where $\delta Z^{(P)} = -2p_0[\nabla \mathbf{v}]^{(2)}$ and $\delta Z^{(h)} = -(5kp_0/2m)\nabla T$. Taking the partial differentiation ∂_α on Eqs. (52) and (53), we can obtain the evolution equations of $\delta Z^{(\alpha)}$. Since

$$\begin{aligned}
\delta \overleftrightarrow{X}^{(P)}(\text{s.t.}) &= -(1/2p_0T_0)\overleftrightarrow{\Pi}(\text{s.t.}) \\
&= (1/2p_0T_0)\delta\Phi^{(P)}(\text{s.t.}) , \\
\delta \overleftrightarrow{X}^{(h)}(\text{s.t.}) &= -(2m/5kp_0T_0^2)\mathbf{Q}'(\text{s.t.}) \\
&= -(2m/5kp_0T_0^2)\delta\Phi^{(h)}(\text{s.t.}) .
\end{aligned}$$

Thus

$$\delta X^{(\alpha)}(\text{s.t.}) \frac{d}{dt}\delta\Phi^{(\alpha)}(\text{s.t.}) = K^{(\alpha)} \frac{d}{dt}[\delta\Phi^{(\alpha)}(\text{s.t.})]^2 ,$$

where $K^{(\alpha)} < 0$, $\alpha = p, h$, and P_0, T_0 are the values of P, T, respectively, at equilibrium.

By the solutions of Eqs. (51)–(53), $x_e = 0$ is globally asymptotically stable, and all eigenvalues of the linear operator A have negative real parts. This implies that

$$\frac{d}{dt}[\delta\Phi^{(\alpha)}(\text{s.t.})]^2 \leqslant 0 .$$

Consequently, by Eq. (49),

$$\int_\Omega d\Omega \sum_\alpha \delta X^{(\alpha)}(\text{s.t.}) \frac{d}{dt}\delta\Phi^{(\alpha)}(\text{s.t.}) \geqslant 0 ,$$

and therefore, $-\delta^2\mathscr{S}$ can be considered as an alternative Liapunov function for the dynamical system generated by Eqs. (51)–(53). Of course, it will be much more interesting and challenging if $-\delta^2\mathscr{S}$ indeed can be proved to be a Liapunov function for the nonlinear variational evolution equations (44)–(46). At the moment, we believe that this is still an open problem.

ACKNOWLEDGMENTS

The author wishes to express his gratitude to Professor B. C. Eu, from whom the author has learned a great deal on the subject of nonlinear irreversible thermodynamics. The author also wishes to thank the referee for some valuable suggestions and for bringing the author's attention to Keizer's work. Finally, we wish to thank C. Robertson, assistant to the Academic Dean of Vanier College, for her constant encouragement.

This research was supported by a grant from FCAC of the Ministry of Education, the Province of Quebec.

APPENDIX

Vector: $\mathbf{A}_1$, scalar product of vectors: $\mathbf{A}_1 \cdot \mathbf{A}_2$, tensor: $\overleftrightarrow{B}_1$, tensor contraction: $\overleftrightarrow{B}_1 : \overleftrightarrow{B}_2$.

Local average of A: $\langle A \rangle = \int d\mathbf{u}_i \, A f(\mathbf{r}, \mathbf{u}_i, t)$.

Density: $\rho_i = m_i n_i = \langle m_i \rangle$, $\rho = \Sigma_i \rho_i$.

Concentration: $c_i = \rho_i / \rho$.

Hydrodynamic velocity: $\rho \mathbf{v} = \Sigma_i \langle m_i \mathbf{u}_i \rangle$,

Internal energy density: $\rho E = \Sigma_i \langle \frac{1}{2} m_i (\mathbf{u}_i - \mathbf{v}) \cdot (\mathbf{u}_i - \mathbf{v}) \rangle$.

Mass flux: $\mathbf{J}_i = \langle m_i (\mathbf{u}_i - \mathbf{v}) \rangle$.

Stress tensor:

$$\overleftrightarrow{P} = \sum_i \overleftrightarrow{P}_i = \sum \langle m_i (\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v}) \rangle ,$$

$$\begin{aligned}
\overleftrightarrow{\pi} &= \sum_i \overleftrightarrow{\pi}_i = \sum_i [\overleftrightarrow{P}_i]^{(2)} \\
&= \sum_i \left\{ \tfrac{1}{2}(\overleftrightarrow{P}_i + \overleftrightarrow{P}_i{}^t) - \tfrac{1}{3}(\overleftrightarrow{P}_i : \overleftrightarrow{I})\overleftrightarrow{I} \right\} \\
&= \sum_i \langle m_i [(\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v})]^{(2)} \rangle ,
\end{aligned}$$

$P_i = \frac{1}{3}(\overleftrightarrow{P}_i : \overleftrightarrow{I})$, $\overleftrightarrow{I}$: unit tensor.

Traceless symmetric part of second rank tensor $\overleftrightarrow{A} : [\overleftrightarrow{A}]^{(2)}$.

Heat flux:

$$\mathbf{Q} = \sum_i \mathbf{Q}_i = \sum_i \langle \tfrac{1}{2} m_i (\mathbf{u}_i - \mathbf{v}) \cdot (\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v}) \rangle ,$$

$$\mathbf{Q}_i' = \mathbf{Q}_i - \frac{5}{2}\left(\frac{kT}{m_i}\right)\mathbf{J}_i .$$

Third moment:

$$\overleftrightarrow{\psi}_i{}^{(p)} = \langle m_i (\mathbf{u}_i - \mathbf{v})[(\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v})]^{(2)} \rangle ,$$

$$\overleftrightarrow{\phi}_i{}^{(3)} = \langle m_i (\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v}) \rangle .$$

Fourth moment:

$$\overleftrightarrow{\psi}_i{}^{(h)} = \langle \tfrac{1}{2} m_i (\mathbf{u}_i - \mathbf{v}) \cdot (\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v}) \rangle .$$

Boltzmann collision integral:

$$C(f_i, f_j) = C_{ij} = \int d\mathbf{u}_j \, d\Phi \, db \, b g_{ij} (f_i' f_j' - f_i f_j) .$$

Collisional average:

$$\langle AC(f_i, f_j) \rangle = \langle AC_{ij} \rangle = \{A\}_{c_{ij}} ,$$

$$\overleftrightarrow{\Lambda}_{ij}{}^{(p)} = \langle m_i [(\mathbf{u}_i - \mathbf{v})(\mathbf{u}_i - \mathbf{v})]^{(2)} \rangle_{c_{ij}} ,$$

$$\Lambda_{ij}{}^{(h)} = \left\langle \left[\tfrac{1}{2} m_i (\mathbf{u}_i - \mathbf{v}) \cdot (\mathbf{u}_i - \mathbf{v}) - \tfrac{5}{2} kT \right] (\mathbf{u}_i - \mathbf{v}) \right\rangle_{c_{ij}} ,$$

$$\Lambda_{ij}{}^{(b)} = \left\langle \tfrac{1}{3} m_i (\mathbf{u}_i - \mathbf{v}) \cdot (\mathbf{u}_i - \mathbf{v}) \right\rangle_{c_{ij}} ,$$

$$\Lambda_{ij}{}^{(f)} = \langle m_i (\mathbf{u}_i - \mathbf{v}) \rangle_{c_{ij}} .$$

Other definitions:

$$\overleftrightarrow{Z}_i^{(p)} = -\nabla\cdot\overleftrightarrow{\psi}_i^{(p)} - \overleftrightarrow{\pi}_i\nabla\cdot\mathbf{v} - 2[\overleftrightarrow{\pi}_i\cdot\nabla\mathbf{v}]^{(2)} - 2p_i[\nabla\mathbf{v}]^{(2)} + \frac{2}{\rho}\sum_j [\mathbf{J}_i\nabla p_j]^{(2)} + \frac{2}{\rho}[\mathbf{J}_i\nabla\cdot\overleftrightarrow{\pi}_j]^{(2)},$$

$$\mathbf{Z}_i^{(h)} = \mathbf{Z}_i'^{(h)} - \frac{5}{2}\frac{kT}{m_i}\mathbf{Z}_i^{(f)} - \frac{5}{2}\frac{kT}{m_i}\mathbf{J}_i\frac{d}{dt}\log T,$$

$$\mathbf{Z}_i^{(h)} = -\nabla\cdot\overleftrightarrow{\psi}_i^{(h)} - \mathbf{Q}_i\nabla\cdot\mathbf{v} + \rho^{-1}\nabla\cdot\overleftrightarrow{P}:(\rho_i E_i\overleftrightarrow{I} + \overleftrightarrow{P}_i) - (\mathbf{Q}_i\cdot\nabla)\mathbf{v} - \overleftrightarrow{\phi}_i^{(3)}:\nabla\mathbf{v},$$

$$\mathbf{Z}_i^{(f)} = -\nabla\cdot(\overleftrightarrow{P}_i - c_i\overleftrightarrow{P}) - \overleftrightarrow{P}\cdot\nabla c_i - \mathbf{J}_i\nabla\cdot\mathbf{v} - \mathbf{J}_i,$$

$$Z_i^{(b)} = -\tfrac{5}{3}P_i\nabla\cdot\mathbf{Q}_i + (2/3\rho)\mathbf{J}_i\nabla:\overleftrightarrow{P} - \tfrac{2}{3}\overleftrightarrow{\pi}_i:\nabla\mathbf{v}.$$

[1] I. Prigogine, Physica **15**, 272 (1949); S. R. de Groot and P. Mazur, *Non-equilibrium Thermodynamics* (North-Holland, Amsterdam, 1962); I Prigogine, *Thermodynamics of Irreversible Process*, third Ed. (Interscience, New York, 1967).

[2] B. C. Eu, J. Chem. Phys. **73**, 2958 (1980); **74**, 2998 (1981); Ann. Phys. **140**, 341 (1982).

[3] L. Boltzmann, *Lectures on Gas Theory*, translated by S. G. Brush (University of California Press, Berkeley, 1964).

[4] L. Onsager, Phys. Rev. **37**, 405 (1931); **38**, 2265 (1931).

[5] M. Chen, J. Math. Phys. **27**, 603 (1986).

[6] P. Glansdorff and I. Prigogine, *Thermodynamic Theory of Structure, Stability and Fluctuations* (Wiley, New York, 1971).

[7] C. Cercignani, *Mathematical Methods in Kinetic Theory* (Plenum, New York, 1969).

[8] J. H. Ferziger and H. G. Kaper, *Mathematical Theory of Transport Processes in Gases* (North-Holland, Amsterdam, 1972), p. 77.

[9] H. Grad, Commun. Pure Appl. Math. **2**, 311 (1949).

[10] S. Chapman and T. G. Cowling, *Mathematical Theory of Non-Uniform Gases* (Cambridge U.P., London, 1961).

[11] K. T. Mashiyama and H. Mori, J. Stat. Phys. **18**, 383 (1978).

[12] J. Keizer, *Statistical thermodynamics of nonequilibrium processes* (Springer, New York, 1987), Chap. 9.

[13] H. Grad, *Handbuch der Physik*, edited by S. Flugge (Springer, Berlin, 1958), Vol. XII, pp. 204 and 205.

[14] J. Keizer, J. Stat. Phys. **31**, 485 (1983); **37**, 491 (1984).

[15] R. Courant and D. Hilbert, *Methods of Mathematical Physics* (Wiley, New York, 1955), p. 139.

[16] H. B. Callen, *Thermodynamics* (Wiley, New York, 1960).

[17] J. G. Gibbs, *The Scientific Papers of J. W. Gibbs* (Dover, New York, 1961), Vol. I.

[18] J. Keizer, J. Chem. Phys. **64**, 4466 (1976).

[19] J. Keizer and O. Chang, J. Chem. Phys. **87**, 4064 (1987).

[20] J. A. Walker, *Dynamical System and Evolution Equations* (Plenum, New York, 1980), p. 63.

[21] Reference 20, Chap. 3, Theorem 3.2, p. 92.

[22] K. Yosida, *Functional Analysis* (Springer, New York, 1977).

[23] Reference 20, p. 180.

[24] Reference 20, p. 181.

[25] Reference 20, Theorem 5.2, p. 179.

[26] Reference 20, Theorem 4.2, p. 168.

Multi-Hamiltonian structure of the Born–Infeld equation

Metin Arik
Department of Mathematics, Technical University of Istanbul, Istanbul, Turkey

Fahrünisa Neyzi
Department of Physics, Bosphorus University, Bebek, Istanbul, Turkey

Yavuz Nutku
Department of Physics, Technical University of Istanbul, Istanbul, Turkey

Peter J. Olver
School of Mathematics, University of Minnesota, Minneapolis, Minnesota 55455

John M. Verosky
Department of Mathematics, Heriot-Watt University, Riccarton, Currie, Edinburgh, EH14 4AS, Scotland

(Received 16 December 1988; accepted for publication 8 February 1989)

The multi-Hamiltonian structure, conservation laws, and higher order symmetries for the Born–Infeld equation are exhibited. A new transformation of the Born-Infeld equation to the equations of a Chaplygin gas is presented and explored. The Born–Infeld equation is distinguished among two-dimensional hyperbolic systems by its wealth of such multi-Hamiltonian structures.

I. INTRODUCTION

A nonlinear modification of Maxwell's electrodynamics was proposed by Born and Infeld in 1934.[1] The simplest example of this system of nonlinear field equations is the quasilinear second-order equation in 1 + 1 dimensions:

$$(1+\varphi_x^2)\varphi_{tt} - 2\varphi_t\varphi_x\varphi_{xt} - (1-\varphi_t^2)\varphi_{xx} = 0, \quad (1.1)$$

which is known as the Born–Infeld equation.[2] The Born–Infeld also governs minimal surfaces in 2 + 1-dimensional Minkowski space, which is a special case of the Nambu string.[3] The world sheet of the Nambu string is parametrized by harmonic coordinates, familiar from the theory of minimal surfaces, rather than the light cone gauge.[4] We will also consider the representation of Eq. (1.1) in null coordinates:

$$x' = x + t, \quad t' = x - t,$$

in terms of which the Born–Infeld equation can be rewritten as

$$\varphi_{x'}^2\varphi_{t't'} - 2(2+\varphi_{t'}\cdot\varphi_{x'})\varphi_{x't'} + \varphi_{t'}^2\varphi_{x'x'} = 0. \quad (1.2)$$

In this paper we shall discuss the Hamiltonian structure, symmetries, and conservation laws of the Born–Infeld equation. We shall find that it has a remarkably rich structure. The first step is to recast the Born–Infeld equation as a first-order quasilinear Hamiltonian system of hydrodynamic type.[5,6] Remarkably, this can be done in three inequivalent ways, one of which corresponds to a system of isentropic gas dynamics, with the adiabatic index $\gamma = -1$ corresponding to the pressure–density relation $P = -1/\rho$, which is known as a *Chaplygin gas*.[7] Each of these systems is separable; therefore, the extensive results on Hamiltonian structures, symmetries, and conservation laws of Sheftel'[8] and Olver and Nutku[9] can be used. Even among the separable two-dimensional systems, the Born–Infeld system has a much richer algebraic structure than most, in part due to the multiple Hamiltonian reformulations of the equation. We will see that the Born–Infeld equation admits (at least) six independent Hamiltonian structures, in contrast to two Hamiltonian structures for a general separable system and four Hamiltonian structures in the more general polytropic case. Moreover, the diagonalization techniques introduced by Verosky[10] are then applied to show that these systems admit first-order conserved densities depending on arbitrary functions—which is special to these particular systems.

We assume that the reader is familiar with the basics of Hamiltonian systems of evolution equations, symmetries, and conservation laws, as presented, for example, in Olver.[11] In the interests of brevity, we have omitted many of the more complicated computations.

II. HYPERBOLIC FORMS OF THE BORN–INFELD EQUATION

We begin by showing that the Born–Infeld equation can be rewritten in several ways as a first-order system of quasilinear hyperbolic evolution equations. All of these representations have the form[9]

$$u_t = -D_x\frac{\partial H}{\partial v}, \quad v_t = -D_x\frac{\partial H}{\partial u}, \quad (2.1)$$

where $\mathscr{H}[u,v] = \int H(u,v)dx$ is the Hamiltonian functional and D_x is the total x derivative. In vector form, if we let

$$\mathbf{u}(x,t) = \begin{pmatrix} u(x,t) \\ v(x,t) \end{pmatrix},$$

then Eqs. (2.1) are in elementary Hamiltonian form[11]:

$$\mathbf{u}_t = \mathscr{D}^* E_{\mathbf{u}}[H], \quad (2.2)$$

where $E_{\mathbf{u}}$ denotes the Euler operator, or variational derivative with respect to $\mathbf{u}$. The Hamiltonian operator in (2.2) is the constant coefficient skew-adjoint differential operator

$$\mathscr{D}^* = -\sigma_1\cdot D_x = \begin{pmatrix} 0 & -D_x \\ -D_x & 0 \end{pmatrix},$$

$$\text{where } \sigma_1 = \begin{pmatrix} 0 & 1 \\ 1 & 0 \end{pmatrix}. \quad (2.3)$$

The induced Poisson bracket on the space of densities is given by the standard formula

 0022-2488/89/061338-07$02.50

$$\{\mathscr{F},\mathscr{H}\} = \frac{1}{2}\int E_u[F]\cdot\mathscr{D}^* E_u[H]\,dx$$
$$= \frac{1}{2}\int \{E_v[F]D_x E_u[H] - E_u[F]D_x E_v[H]\}dx.$$

We begin by looking at Eq. (1.1) in the physical variables. Since (1.1) can be derived from a variational principal where the Lagrangian depends only on the gradient of φ, we know that it can be expressed as the integrability condition of a first-order system.[6] To effect this change, we introduce a new potential ψ given by

$$\psi_x = \varphi_x/\sqrt{1+\varphi_x^2-\varphi_t^2}, \quad \psi_t = \varphi_t/\sqrt{1+\varphi_x^2-\varphi_t^2}. \tag{2.4}$$

Inverting Eqs. (2.4) for the first derivatives of φ we find the same expressions, with the roles of φ and ψ interchanged. Equation (1.1) is then realized as the integrability condition for system (2.4): Moreover, its companion equation expressing the integrability conditions for φ is again (1.1), with ψ replacing φ. We shall now formulate these equations in terms of a pair of conservation laws. For this purpose, we introduce the variables

$$r = \varphi_x, \quad s = \psi_x.$$

Solving (2.4) for φ_t, ψ_t we deduce that the one-forms

$$\alpha = r\,dx + s\sqrt{(1+r^2)/(1+s^2)}\,dt = d\varphi,$$
$$\omega = s\,dx + r\sqrt{(1+s^2)/(1+r^2)}\,dt = d\psi,$$

are exact; the implication that they are closed gives rise to the following pair of quasilinear evolution equations:

$$r_t = \left[rs/\sqrt{(1+r^2)(1+s^2)}\right]r_x + \sqrt{(1+r^2)/(1+s^2)^3}\,s_x,$$
$$s_t = \sqrt{(1+s^2)/(1+r^2)^3}\,r_x + \left[rs/\sqrt{(1+r^2)(1+s^2)}\right]s_x. \tag{2.5}$$

We will call the quasilinear system (2.5) the *physical version* of the Born–Infeld equation. It is easy to see that (2.5) is in the standard Hamiltonian form (2.2), where

$$\tilde{H}^*(r,s) = \sqrt{(1+r^2)(1+s^2)} \tag{2.6}$$

is the Hamiltonian density. We note that there are alternative ways of reexpressing (1.1) as the integrability condition of a first-order system such as (2.2), but there is a unique choice of ψ which will result in a Hamiltonian system of equations. (An alternative first-order form of the Born–Infeld equation that is not Hamiltonian can be found in Whitham.[2])

A similar reasoning applies to the Born–Infeld equation, rewritten in the null coordinates (1.2). Dropping the primes on x, t, we similarly introduce a new potential χ by

$$\chi_x = -\varphi_x/\sqrt{1+\varphi_x\varphi_t}, \quad \chi_t = \varphi_t/\sqrt{1+\varphi_x\varphi_t}. \tag{2.7}$$

As in (2.4), the companion equation for χ is identical to (1.2). Define

$$z = \varphi_x, \quad w = \chi_x.$$

Note that the one-forms

$$\alpha = z\,dx - (1/z - z/w^2)dt = d\varphi,$$
$$\omega = w\,dx - (1/w - w/z^2)dt = d\chi$$

are exact, leading to an alternative system of quasilinear evolution equations:

$$z_t = (1/z^2 + 1/w^2)z_x - (2z/w^3)w_x,$$
$$w_t = -(2w/z^3)z_x + (1/z^2 + 1/w^2)w_x, \tag{2.8}$$

which will be called the *null coordinate version* of the Born–Infeld equation. Again, (2.8) are in Hamiltonian form (2.2), with the Hamiltonian density

$$\hat{H}^*(z,w) = z/w + w/z. \tag{2.9}$$

Although the two versions of the Born–Infeld equation can be obtained by a transformation between physical and null coordinates, it is rather remarkable that there is also a transformation of the dependent variables which maps one to the other, as shown in the following theorem.

Theorem 1: Given r, s with $rs > 1$, define the transformation

$$z = (1+r^2)^{1/4}(1+s^2)^{1/4}[(rs+1)^{1/2} + (rs-1)^{1/2}],$$
$$w = (1+r^2)^{1/4}(1+s^2)^{1/4}[(rs+1)^{1/2} - (rs-1)^{1/2}]. \tag{2.10}$$

If (r,s) satisfy the physical version of the Born–Infeld equation (2.5), then (z,w) satisfy the null coordinate version (2.8).

The proof is a straightforward, but lengthy calculation. In Sec. III we shall see how the transformation (2.10) can be systematically deduced by referring to the second Hamiltonian structure of (2.5).

We now turn to a remarkable transformation from the Born–Infeld system to a system of quasilinear equations arising in polytropic gas dynamics.

Theorem 2: Define the variables

$$u = -(1/z^2 + 1/w^2), \quad v = zw/2. \tag{2.11}$$

Then z, w satisfy the Born–Infeld system (2.8) if and only if u,v satisfy the gas dynamics system

$$u_t + uu_x + v^{-3}v_x = 0, \quad v_t + (uv)_x = 0. \tag{2.12}$$

The proof is again a straightforward calculation. The system (2.12) corresponds to the equations of isentropic, polytropic gas dynamics with the adiabatic index $\gamma = -1$, known as a Chaplygin gas.[7] The system (2.12) is distinguished from such quasilinear hyperbolic systems by the fact that shocks do not form[12,2]: This system is also in the elementary Hamiltonian form (2.2), with the Hamiltonian density

$$H^*(u,v) = u^2v/2 + 1/2v. \tag{2.13}$$

We remark that the reduction of a gas dynamics system to a single second-order hyperbolic equation, which includes the reduction of a Chaplygin gas to the Born–Infeld equation (1.2), can be found in Garabedian.[13] Note, also, that the physical version (2.5) can be transformed directly to the gas dynamics version (2.12) by composing the transformations (2.10) and (2.11):

$$u = rs/\sqrt{(1+r^2)(1+s^2)}, \quad v = \sqrt{(1+r^2)(1+s^2)}. \tag{2.14}$$

We thus have three distinct ways of reformulating the

Born–Infeld equation as a Hamiltonian system of quasilinear evolution equations of the type (2.2). To keep track of various functions and operators in the different coordinate systems, we will adopt the following conventions: In the physical (r,s) version, these quantities will have an overtilde, e.g., $\tilde{H}$; in the null (z,w) version, they will have a caret, e.g., $\hat{H}$; and the gas dynamics (u,v) coordinates will not have any distinguishing mark, e.g., H.

III. FIRST-ORDER HAMILTONIAN OPERATORS

We now investigate other first-order Hamiltonian structures for the Born–Infeld equation using the methods found in Refs. 6 and 9. First, we recall that the most general skew-adjoint first-order matrix differential operator has the form

$$\mathscr{D} = M\cdot D_x + D_x\cdot M + Q_x = \begin{pmatrix} 2mD_x + m_x & 2pD_x + p_x + q_x \\ 2pD_x + p_x - q_x & 2nD_x + n_x \end{pmatrix}, \tag{3.1}$$

where

$$M = \begin{pmatrix} m & p \\ p & n \end{pmatrix}$$

is a general symmetric matrix,

$$Q = \begin{pmatrix} 0 & q \\ -q & 0 \end{pmatrix}$$

is a general skew-symmetric matrix, and where the coefficients m, n, p, and q are allowed to depend on the dependent variables. The particular Hamiltonian operator (2.3) corresponds to the choice

$$\mathscr{D}^*\colon\ m^* = n^* = q^* = 0,\quad p^* = -\tfrac{1}{2}. \tag{3.2}$$

In order that the Poisson bracket associated with the operator (3.1) satisfies the Jacobi identity, the coefficients m, n, p, and q must satisfy additional first-order partial differential equations.[6]

Besides the standard Hamiltonian form (2.2), any polytropic gas dynamics system can be written in two additional, alternative Hamiltonian forms involving first-order Hamiltonian operators[6] and making it a tri-Hamiltonian system:

$$\mathbf{u}_t = \mathscr{D}_0 E_{\mathbf{u}}(H_2) = \mathscr{D}_1 E_{\mathbf{u}}(H_1) = \mathscr{D}_2 E_{\mathbf{u}}(H_0). \tag{3.3}$$

For the case of the adiabatic index $\gamma = -1$, the Hamiltonian operators in (3.3) have the form

$$\mathscr{D}_0 = \mathscr{D}^*\colon\ m_0 = 0,\quad n_0 = 0,\quad p_0 = -\tfrac{1}{2},\quad q_0 = 0, \tag{3.4}$$

$$\mathscr{D}_1\colon\ m_1 = 1/v^3,\quad n_1 = v,\quad p_1 = -u,\quad q_1 = 2u, \tag{3.5}$$

$$\mathscr{D}_2\colon\ m_2 = u/v^3,\quad n_2 = uv,\quad p_2 = -u^2/2 - 1/2v^2,\quad q_2 = u^2 \tag{3.6}$$

and are mutually compatible.[11] We note that (3.4)–(3.6) are genuinely distinct Hamiltonian operators, meaning that $\mathscr{D}_2$ is not related to $\mathscr{D}_0$ and $\mathscr{D}_1$ according to a well-known recurrence formula[14] which generates higher order Hamiltonian operators from any bi-Hamiltonian system. The corresponding Hamiltonian densities placing (2.12) in the tri-Hamiltonian form (3.3) are

$$H_0 = v,\quad H_1 = uv,\quad H_2 = u^2v/2 + 1/2v, \tag{3.7}$$

which appear in the well-known hierarchy of conserved densities for gas dynamics.[9] (See Sec. IV.)

Before proceeding to the tri-Hamiltonian structure of the null coordinate and physical versions of the Born–Infeld equation, it helps to recall how Hamiltonian operators transform under a change of variables.

Lemma 3:[14,15] Let $\mathbf{u} = \varphi(\mathbf{z})$ be a change of variables and let $\mathbf{J}$ denote the Jacobian matrix of φ. Let $\mathscr{D}$ denote a Hamiltonian operator in the $\mathbf{u}$ coordinates and $\hat{\mathscr{D}}$ the corresponding Hamiltonian operator in the $\mathbf{z}$ coordinates; then these two operators are related by the change of variables formula

$$\mathscr{D} = \mathbf{J}\cdot\hat{\mathscr{D}}\cdot\mathbf{J}^T. \tag{3.8}$$

Thus for Hamiltonian operators of the form (3.1), we find the corresponding coefficient matrices have the change of variables formula

$$M = \mathbf{J}\cdot\hat{M}\cdot\mathbf{J}^T,\quad Q_x = \mathbf{J}\cdot\hat{Q}_x\cdot\mathbf{J}^T + \mathbf{J}\cdot\hat{M}\cdot\mathbf{J}_x^T - \mathbf{J}_x\cdot\hat{M}\cdot\mathbf{J}^T. \tag{3.9}$$

Dubrovin and Novikov[5] have pointed out that the Poisson brackets defined by Hamiltonian operators for equations of hydrodynamic type give rise to Riemannian metrics with vanishing torsion and curvature. The metric corresponding to an operator of the form (3.1) is given by

$$ds^2 = (n\,du^2 - 2p\,du\,dv + m\,dv^2)/(mn - p^2). \tag{3.10}$$

Since the metric (3.10) is flat we know that a (possibly complex) change of variables $\mathbf{u} = \varphi(\mathbf{z})$ will bring it to the canonical form $d\hat{s}^2 = 2\,dz\,dw$, determining the maximal analytic extension of the metric and corresponding to the elementary Hamiltonian operator (2.3). Remarkably, the transformations (2.11) and (2.14) are precisely the ones needed to place the metrics determined by the Hamiltonian operators $\mathscr{D}_1$, $\mathscr{D}_2$ in canonical form.

Proposition 4: Under the transformations (2.11) and (2.14) the Hamiltonian operators and densities for the gas dynamics system (2.12) are mapped to the following Hamiltonian operators and densities for the null and physical versions of the Born–Infeld equation:

Null coordinate version—Hamiltonian operators:

$$\mathscr{D}_0\colon\ \hat{m}_0 = zw^{-3}(z^{-2} - w^{-2})^{-2},\quad \hat{n}_0 = wz^{-3}(z^{-2} - w^{-2})^{-2},$$

$$\hat{p}_0 = -\tfrac{1}{2}(z^{-2} - w^{-2})^{-1},\quad \hat{q}_0 = (z^{-2} - w^{-2})^{-1},$$

$$\hat{\mathscr{D}}_1 = -2\mathscr{D}^*\colon\ \hat{m}_1 = 0,\quad \hat{n}_1 = 0,\quad \hat{p}_1 = 1,\quad \hat{q}_1 = 0,$$

$$\hat{\mathscr{D}}_2\colon\ \hat{m}_2 = -\frac{z}{w^3},\quad \hat{n}_2 = -\frac{w}{z^3},\quad \hat{p}_2 = -\frac{1}{2z^2} + \frac{1}{2w^2},\quad \hat{q}_2 = \frac{1}{z^2} - \frac{1}{w^2}.$$

Hamiltonian densities:

$$\hat{H}_0(z,w) = zw/2,$$
$$\hat{H}_1(z,w) = -z/2w - w/2z,$$
$$\hat{H}_2(z,w) = w/4z^3 + 3/2zw + z/4w^3.$$

Physical version—Hamiltonian operators:

$$\tilde{\mathscr{D}}_0\colon\ \tilde{m}_0 = \frac{-2rs(1+r^2)^2}{(r^2-s^2)^2},\quad \tilde{n}_0 = \frac{-2rs(1+s^2)^2}{(r^2-s^2)^2},$$
$$\tilde{p}_0 = \frac{(r^2+s^2)(1+r^2)(1+s^2)}{(r^2-s^2)^2},$$
$$\tilde{q}_0 = \frac{(r^2+s^2+2r^2s^2)}{(r^2-s^2)},$$
$$\tilde{\mathscr{D}}_1\colon\ \tilde{m}_1 = \frac{(1+r^2)^{3/2}(r^2+s^2+2r^2s^2)}{(1+s^2)^{1/2}(r^2-s^2)^2},$$
$$\tilde{n}_1 = \frac{(1+s^2)^{3/2}(r^2+s^2+2r^2s^2)}{(1+r^2)^{1/2}(r^2-s^2)^2},$$
$$\tilde{p}_1 = \frac{(1+r^2)^{1/2}(1+s^2)^{1/2}}{(r^2-s^2)^2}rs(2+r^2+s^2),$$
$$\tilde{q}_1 = [2rs/(r^2-s^2)](1+r^2)^{1/2}(1+s^2)^{1/2},$$
$$\tilde{\mathscr{D}}_2 = \tilde{\mathscr{D}}^*\colon\ \tilde{m}_2 = 0,\quad \tilde{n}_2 = 0,\quad \tilde{p}_2 = -\tfrac{1}{2},\quad \tilde{q}_2 = 0.$$

Hamiltonian densities:

$$\tilde{H}_0(r,s) = \sqrt{(1+r^2)(1+s^2)},$$
$$\tilde{H}_1(r,s) = rs,$$
$$\tilde{H}_2(r,s) = (r^2s^2+1)/2\sqrt{(1+r^2)(1+s^2)}.$$

IV. RECURSION OPERATORS AND CONSERVED DENSITIES

According to Magri's theorem,[16] any compatible bi-Hamiltonian system has an associated recursion operator. The Hamiltonian operators $\mathscr{D}_0$, $\mathscr{D}_1$, and $\mathscr{D}_2$ are mutually compatible[6]; thus there are three recursion operators for the gas dynamics system,

$$\mathscr{R}_1 = \mathscr{D}_1\cdot\mathscr{D}_0^{-1},\quad \mathscr{R}_2 = \mathscr{D}_2\cdot\mathscr{D}_0^{-1},\quad \mathscr{R}_3 = \mathscr{D}_2\cdot\mathscr{D}_1^{-1}, \tag{4.1}$$

although there is a trivial relation between them:

$$\mathscr{R}_2 = \mathscr{R}_3\cdot\mathscr{R}_1.$$

Similar recursion operators can be constructed for the null coordinate and physical versions of the Born–Infeld equation. Now, a curious phenomenon occurs when we apply the recursion operator to the hierarchy where the Born–Infeld Hamiltonian lies. We find that the hierarchy of Hamiltonian flows $\mathscr{R}_1$ terminates after just two steps:

$$\mathscr{R}_1\colon\ H_0 \to H_1 \to H_2 \to 0$$

because the second Hamiltonian H_2 is a distinguished functional (Casimir) for the Hamiltonian structure determined by $\mathscr{D}_1$. Therefore, the hierarchy guaranteed by Magri's theorem[16] degenerates into just three independent Hamiltonians; we have a nontrivial example of a bi-Hamiltonian system which does not satisfy one of the technical hypotheses of Magri's theorem, which states that the hierarchy of Hamiltonians be independent functionals.[11,16] However, the second recursion operator $\mathscr{R}_2$ does generate further members of the gas dynamics hierarchy of conserved densities.[9] (We remark that in Ref. 9 we failed to show that this property of the hierarchy of flows generated by one of the recursion operators can occasionally degenerate. The equations following (4.3) of Ref. 9 should read as

$$\mathscr{R}_1(\mathbf{Q}_n) = (n\gamma - n - 1)\mathbf{Q}_{n+1},$$
$$\mathscr{R}_2(\mathbf{Q}_n) = (n/2)(n\gamma + \gamma - n - 3)\mathbf{Q}_{n+2},$$
$$\mathscr{R}_1(\tilde{\mathbf{Q}}_n) = (n\gamma - n + 1)\tilde{\mathbf{Q}}_{n+1},$$
$$\mathscr{R}_2(\tilde{\mathbf{Q}}_n) = [(n+1)/2](n\gamma - n + 1)\tilde{\mathbf{Q}}_{n+2},$$

leading to degeneracies if γ has one of the forms $1 \pm 1/n$, $1 \pm 2/n$ for some integer n.)

Another interesting anomaly occurs for the physical version of the Born–Infeld equation. Here, from the point of view of Ref. 6, the most natural recursion operator would be

$$\mathscr{R}^* = \tilde{\mathscr{D}}_1\cdot\tilde{\mathscr{D}}_2^{-1} = \tilde{\mathscr{D}}_1\cdot\mathscr{D}^{*-1}.$$

Again, this recursion operator does not produce a hierarchy of symmetries and conserved Hamiltonian densities. In fact, as the reader can check, the recursion operator repeats after two steps:

$$\mathscr{R}^*\colon\ H_0 \to H_1 \to H_0 \to H_1 \to H_0 \to \cdots,$$

resulting in an infinite loop; again the functionals produced by Magri's theorem[16] are not independent. [At first glance, this result does not seem reconciled with the gas dynamics version under the transformation (2.14). However, we note that since the recursion operator involves the inverse of the Hamiltonian operator $\mathscr{D}^*$, we can add in any element of its kernel at each step. Thus the explanation is that we have just chosen different elements of ker $\mathscr{D}_2$ to add in.]

The gas dynamics, null coordinate, and physical versions of the Born–Infeld equation are examples of *separable systems*,[8,9] meaning that the Hamiltonian density H in the representation (2.2) satisfies

$$H_{uu}/H_{vv} = \lambda(u)/\mu(v). \tag{4.2}$$

For the three versions, the separation coefficients are given by

gas dynamics [(2.12)]:

$$\lambda(u) = 1,\quad \mu(v) = v^{-4},$$

null version [(2.8)]:

$$\hat{\lambda}(z) = z^{-4},\quad \hat{\mu}(w) = w^{-4}, \tag{4.3}$$

physical version [(2.5)]:

$$\tilde{\lambda}(r) = (1+r^2)^{-2},\quad \tilde{\mu}(s) = (1+s^2)^{-2},$$

It is standard that the zeroth-order conserved densities for such a system can be found by solving a separable linear wave equation.[8,9]

Proposition 5: A function $F(u,v)$ is a conserved density of a separable Hamiltonian system (2.2) and (4.2) if and only if it is a solution to the linear wave equation

$$F_{uu}/\lambda(u) = F_{vv}/\mu(v). \tag{4.4}$$

Any Hamiltonian system (2.2) admits the conserved densities 1, u, v, and uv. In the separable case, there are four

fundamental hierarchies of solutions to the wave equation (4.4), each of the form

$$H_n(u,v) = \sum_{i=0}^{n} F_i(u)\cdot G_{n-i}(v), \tag{4.5}$$

where the functions F_i and G_i are generated by the recursion relations

$$\frac{\partial^2 F_i}{\partial u^2} = \lambda(u)F_{i-1}, \quad F_i(0) = F_i'(0) = 0,$$
$$\frac{\partial^2 G_i}{\partial v^2} = \mu(v)G_{i-1}, \quad G_i(0) = G_i'(0) = 0.$$

The hierarchies depend on the initial selection of $H_0 = F_0\cdot G_0$:

$$H_0^{(1)} = 1, \quad F_0^{(1)} = G_0^{(1)} = 1,$$
$$H_0^{(2)} = u, \quad F_0^{(2)} = u, \quad G_0^{(2)} = 1,$$
$$H_0^{(3)} = v, \quad F_0^{(3)} = 1, \quad G_0^{(3)} = v,$$
$$H_0^{(4)} = uv, \quad F_0^{(4)} = u, \quad G_0^{(4)} = v.$$

Our transformations do not respect this hierarchical structure of the conserved densities. For example, (2.11) maps the first and fourth null Born–Infeld hierarchies to combinations of all four gas dynamics hierarchies, so that up to a multiple,

$$\hat{H}_{2j}^{(1)} \to H_j^{(1)}, \quad \hat{H}_{2j}^{(4)} \to H_j^{(3)},$$
$$\hat{H}_{2j+1}^{(1)} \to H_j^{(2)}, \quad \hat{H}_{2j+1}^{(4)} \to H_j^{(4)}.$$

On the other hand, the second and third hierarchies are mapped to algebraic conserved densities for the gas dynamics version (2.12). For example, the conserved density $\hat{H}_0^{(2)} = z$ is mapped to the conserved density

$$\sqrt{v - uv^2} + \sqrt{-v - uv^2},$$

which does not show up in any of the standard gas dynamics hierarchies. The hierarchies in the physical r, s variables are no longer rational functions and we shall not write them explicitly: They do not correspond to any of the hierarchies in the other variables (with isolated exceptions) and provide yet other nonpolynomial conserved densities for gas dynamics system (2.12).

V. HIGHER ORDER HAMILTONIAN STRUCTURES

In Olver and Nutku[9] it was shown that any separable Hamiltonian system has a second Hamiltonian structure involving a complicated third-order matrix differential operator. The resulting recursion operator recovers results on symmetries and conservation laws due to Sheftel'.[8] For the Born–Infeld equations, each of the gas dynamics, null coordinate, and physical versions is separable, and so we are led to three distinct third-order Hamiltonian structures. This is probably quite special to these particular systems, but we have no proof of this fact. In particular, it would be interesting to see whether any of the other polytropic gas dynamics systems have additional Hamiltonian structures.

Theorem 6: Consider a separable Hamiltonian system (2.2), where the Hamiltonian density satisfies (4.2). Define the matrix variables

$$U_x = \begin{pmatrix} u_x & \mu(v)v_x \\ v_x & \lambda(u)u_x \end{pmatrix}, \quad V_x = \begin{pmatrix} \lambda(u)u_x & \mu(v)v_x \\ v_x & u_x \end{pmatrix}. \tag{5.1}$$

Then the system can be written in the bi-Hamiltonian form

$$\mathbf{u}_t = \mathscr{D}^* E_{\mathbf{u}}(H) = \mathscr{E} E_{\mathbf{u}}(H^*) \tag{5.2}$$

using the third-order matrix differential operator

$$\mathscr{E} = D_x\cdot V_x^{-1}\cdot D_x\cdot U_x^{-1}\cdot\sigma_1\cdot D_x$$
$$= D_x\cdot V_x^{-1}\cdot D_x\cdot\sigma_1\cdot V_x^{-T}\cdot D_x. \tag{5.3}$$

In particular, $\mathscr{E}$ is Hamiltonian and compatible with $\mathscr{D}^*$.

In the case of gas dynamics the matrix variables coincide:

$$U_x = V_x = \begin{pmatrix} u_x & v^{-4}v_x \\ v_x & u_x \end{pmatrix}$$

and the corresponding Hamiltonian operator (5.3) is

$$\mathscr{E}_0 = D_x\cdot U_x^{-1}\cdot D_x\cdot U_x^{-1}\cdot\sigma_1\cdot D_x, \tag{5.4}$$

which is compatible with $\mathscr{D}_0 = \mathscr{D}^*$. The second Hamiltonian in (5.2) turns out to be

$$H^* = H_2^{(3)} = u^4v/24 + u^2/2v + 1/24v^3,$$

which appears in the third hierarchy (4.5) of conserved densities. The corresponding recursion operator is the square of the simple recursion operator

$$\mathscr{R} = D_x\cdot U_x^{-1}, \tag{5.5}$$

so that

$$\mathscr{E}\cdot\mathscr{D}_0^{-1} = D_x\cdot U_x^{-1}\cdot D_x\cdot U_x^{-1} = \mathscr{R}^2.$$

Similarly, we have a third-order recursion operator in the null variables (z,w). We define the matrix variables

$$Z_x = \begin{pmatrix} z_x & w^{-4}w_x \\ w_x & z^{-4}z_x \end{pmatrix}, \quad W_x = \begin{pmatrix} z^{-4}z_x & w^{-4}w_x \\ w_x & z_x \end{pmatrix}$$

and the operator

$$\hat{\mathscr{E}}_1 = D_x\cdot W_x^{-1}\cdot D_x\cdot Z_x^{-1}\cdot\sigma_1\cdot D_x$$
$$= D_x\cdot W_x^{-1}\cdot D_x\cdot\sigma_1\cdot W_x^{-T}\cdot D_x$$

is Hamiltonian. Moreover, the Hamiltonian operators $\hat{\mathscr{E}}_1$ and $\hat{\mathscr{D}}_1 = -2\mathscr{D}^*$ are compatible; therefore, they form a Hamiltonian pair. The null Born–Infeld equation (2.8) can be written as a bi-Hamiltonian system

$$\mathbf{z}_t = \hat{\mathscr{D}}_1 E_{\mathbf{z}}[\hat{H}] = \hat{\mathscr{E}}_1 E_{\mathbf{z}}[\hat{H}^*], \tag{5.6}$$

where the Hamiltonian is a multiple of the Hamiltonian $\hat{H}_4^{(2)}$ in the fourth hierarchy (4.5):

$$\hat{H}^*(z,w) = 2\hat{H}_2^{(4)}(z,w) = w/12z^3 + 1/2zw + z/12w^3.$$

Note that the transformation (2.11) cannot map the above two higher order Hamiltonian operators to each other since the corresponding bi-Hamiltonian structures do not match, nor are the compatibility relations preserved. Indeed, a long calculation proves that the gas dynamics recursion operator arising from the bi-Hamiltonian Form (5.6) under the transformation (2.11) is the operator

$$\hat{\mathscr{R}}_1 = \hat{\mathscr{E}}_1\hat{\mathscr{D}}_1^{-1} \to -2\mathscr{R}_1\mathscr{R}^2,$$

where $\mathscr{R}$ is the gas dynamics recursion operator given by (5.5) and $\mathscr{R}_1$ is the recursion operator (4.1) arising from Nutku's[6] Hamiltonian structures for gas dynamics. Therefore the operator

$$\mathscr{E}_1 = -2\mathscr{R}_1\mathscr{E}_0\mathscr{R}_1$$

in another third-order Hamiltonian operator for Eqs. (2.12) which is compatible with the first-order Hamiltonian operator $\mathscr{D}_1$, but not with either $\mathscr{D}_0$ or $\mathscr{D}_2$.

Finally there is yet another third-order Hamiltonian operator arising from the physical version of the Born–Infeld equation. The operator takes the form

$$\begin{aligned}\tilde{\mathscr{E}}_2 &= D_x\cdot S_x^{-1}\cdot D_x\cdot R_x^{-1}\cdot\sigma_1\cdot D_x\\ &= D_x\cdot S_x^{-1}\cdot D_x\cdot\sigma_1\cdot S_x^{-T}\cdot D_x,\end{aligned}$$

where

$$R_x = \begin{pmatrix} r_x & (1+s^2)^{-2}s_x \\ s_x & (1+r^2)^{-2}r_x \end{pmatrix},$$

$$S_x = \begin{pmatrix} (1+r^2)^{-2}r_x & (1+s^2)^{-2}s_x \\ s_x & r_x \end{pmatrix}.$$

This Hamiltonian operator is compatible with $\tilde{\mathscr{D}}_2 = \mathscr{D}^*$ and so, when transformed back to the other coordinate systems, it provides yet another Hamiltonian structure for the Born–Infeld equation.

In summary, then, we have found that the Born–Infeld equation in any of its evolutionary forms (2.5), (2.8), or (2.12) possesses six distinct Hamiltonian structures: Three are first order, given by the operators $\mathscr{D}_0$, $\mathscr{D}_1$, and $\mathscr{D}_2$ and three are third order, given by the operators $\mathscr{E}_0$, $\mathscr{E}_1$, and $\mathscr{E}_2$. Moreover $\mathscr{D}_i$ is compatible with $\mathscr{E}_j$ if and only if $i=j$. Whether there are yet more Hamiltonian structures, not trivially related to these, remains an open question!

VI. DIAGONALIZATION AND HIGHER ORDER CONSERVATION LAWS

As shown by Olver and Nutku[9], for a generalized gas dynamics Hamiltonian system there is an additional hierarchy of higher order conservation laws generalizing Verosky's rational first-order conserved density[17]:

$$\hat{H}_1[u,v] = v_x/(u_x^2 - v^{\gamma-3}v_x^2)\,. \tag{6.1}$$

The case of a Chaplygin gas, $\gamma = -1$, is distinguished in that it admits an infinite collection of distinct first-order conserved densities (i.e., they do not differ by a divergence): The easiest way to see this is to apply a diagonalization technique, described by Verosky[10] and Tsarev.[18]

Definition 7: A first-order quasilinear system is said to be in *diagonal form* if it has the form

$$p_t = A(p,q)p_x, \quad q_t = B(p,q)q_x. \tag{6.2}$$

We remark that the existence of a diagonal form for a quasilinear first-order system is related to the existence of Riemann invariants.[18]

Proposition 8: For the Chaplygin system (2.12), the transformation

$$p = u + 1/v, \quad q = u - 1/v \tag{6.3}$$

place it in the diagonal form[19]

$$p_t = -qp_x, \quad q_t = -pq_x. \tag{6.4}$$

Theorem 9[10]: A two-dimensional diagonal quasilinear system (6.2) has a first-order conservation law $D_tT + D_xX = 0$, with conserved density and flux of the form

$$T = \frac{F(p,q)}{p_x} + \frac{G(p,q)}{q_x}, \quad X = \frac{AF}{p_x} + \frac{BG}{q_x} \tag{6.5}$$

if and only if F and G satisfy the system of differential equations

$$\begin{aligned}&(A-B)G_p = 2GB_p, \quad (B-A)F_q = 2FA_q,\\ &FA_p + GB_q = 0.\end{aligned} \tag{6.6}$$

For the special case $A = -q$, $B = -p$, corresponding to the Born–Infeld equation, the third equation in (6.6) is vacuous; thus there are the solutions

$$F(p,q) = \alpha(p)/(p-q)^2, \quad G(p,q) = \beta(q)/(p-q)^2 \tag{6.7}$$

depending on the arbitrary functions $\alpha(p), \beta(q)$. There are similar expressions for other gas dynamics systems with $\gamma \neq -1$, but then the third equation in (6.6) is not vacuous; this restricts the corresponding functions to satisfying $\alpha = -\beta$ and thus both coefficients must be constant! Thus the Born–Infeld case is very special.

In terms of the gas dynamics variables, the conserved densities have the form

$$\begin{aligned}T[u,v] &= v^4\alpha(u+v^{-1})/(v^2u_x - v_x)\\ &\quad + v^4\beta(u-v^{-1})/(v^2u_x + v_x)\,.\end{aligned}$$

Note that the case $\alpha = \frac{1}{2}, \beta = -\frac{1}{2}$ reproduces the conserved density (6.1) when $\gamma = -1$. Under the transformation (2.11), these turn into the following conserved densities for the null version of the Born–Infeld equation:

$$T[z,w] = \frac{z^4w^4\tilde{\alpha}(z^{-1}-w^{-1})}{w^2z_x - z^2w_x} + \frac{z^4w^4\tilde{\beta}(z^{-1}+w^{-1})}{w^2z_x + z^2w_x},$$

where

$$\tilde{\alpha}(s) = \alpha(-s^2)/8s, \quad \tilde{\beta}(s) = \beta(-s^2)/8s.$$

For the particular choices $\tilde{\alpha}(s) = 1$, $\tilde{\beta}(s) = \pm 1$, i.e., $\alpha(s) = 8\sqrt{-s}, \beta(s) = 8\sqrt{-s}$, we obtain the conserved densities

$$z^6w^6w_x/(w^4z_x^2 - z^4w_x^2), \quad z^6w^6z_x/(w^4z_x^2 - z^4w_x^2),$$

which are more like the first-order densities discovered in Verosky[17]; see, also, Olver and Nutku.[9] It is interesting that the transformation (2.11) does *not* map the Verosky-type densities to each other.

It can be shown that the third-order evolution equations corresponding to the above two densities are each bi-Hamiltonian systems; hence the recursion operators lead to two further hierarchies of higher order conserved densities.

ACKNOWLEDGMENTS

Two of (YN and PJO) would like to express our gratitude for the support and hospitality of the Institute for Mathematics and Its Applications (IMA) during the fall program of Nonlinear Waves–Solitons, 1988, during which this work was completed.

The research of author YN was supported in part by NATO Collaborative Research Grant No. RG 86/0055. The research of authors PJO and JMV was supported in part by NATO Collaborative Research Grant No. RG 86/0055 and NSF Grant No. DMS 86-02004.

[1]M. Born and L. Infeld, "Foundation of the new field theory," Proc. R. Soc. London A **144**, 425 (1934).
[2]G. B. Whitham, *Linear and Nonlinear Waves* (Wiley, New York, 1974), p. 617.
[3]Y. Nambu, "Quark model and the factorization of the Veneziano amplitude," in *Symmetries and Quark Models*, edited by R. Chand (Gordon and Breach, New York, 1970), pp. 269-277.
[4]P. Goddard, J. Goldstone, C. Rebbi, and C. B. Thorn, "Quantum dynamics of a massless relativistic string," Nucl. Phys. B **56**, 109 (1973).
[5]B. A. Dubrovin and S. A. Novikov, "Hamiltonian formalism of one-dimensional systems of hydrodynamic type and the Bogolyubov–Whitham averaging method," Sov. Math. Dokl. **27**, 665 (1983).
[6]Y. Nutku, "On a new class of completely integrable systems. II. Multi-Hamiltonian structure," J. Math. Phys. **28**, 2579 (1987).
[7]K. P. Stanyukovich, *Unsteady Motion of Continuous Media* (Pergamon, New York, 1960), p. 137.
[8]M. B. Sheftel', "Integration of Hamiltonian systems of hydrodynamic type with two dependent variables with the aid of the Lie–Bäcklund group," Func. Anal. Appl. **20**, 227 (1986).
[9]P. J. Olver and Y. Nutku, "Hamiltonian structures for systems of hyperbolic conservation laws," J. Math. Phys. **29**, 1610 (1988).
[10]J. M. Verosky, "Applications of the formal variational calculus to the equations of fluid dynamics," Ph.D. thesis, Tulane University, 1985.
[11]P. J. Olver, *Applications of Lie Groups to Differential Equations*, Graduate Texts in Mathematics (Springer, New York, 1986), Vol. 107.
[12]P. A. Thompson, *Compressible-fluid Dynamics* (McGraw-Hill, New York, 1972), p. 253.
[13]P. R. Garabedian, *Partial Differential Equations* (Wiley, New York, 1964), p. 519.
[14]A. S. Fokas and B. Fuchssteiner, "On the structure of symplectic operators and hereditary symmetries," Lett. Nuovo Cimento **28**, 299 (1980).
[15]P. J. Olver, "Darboux' theorem for Hamiltonian differential operators," J. Diff. Eq. **71**, 10 (1988).
[16]F. Magri, "A simple model of the integrable Hamiltonian equation," J. Math. Phys. **19**, 1156 (1978).
[17]J. M. Verosky, "First-order conserved densities for gas dynamics," J. Math. Phys. **27**, 3061 (1986).
[18]S. P. Tsarev, "On Poisson brackets and one-dimensional Hamiltonian systems of hydrodynamic type," Sov. Math. Dokl. **31**, 488 (1985).
[19]M. V. Pavlov, "Hamiltonian formalism of weakly nonlinear hydrodynamic systems," Theor. Math. Phys. **73**, 1242 (1986).

Noether's theorem and gauge transformations: Application to the bosonic string and CP_2^{n-1} model

C. Batlle,[a] J. Gomis, X. Gràcia, and J. M. Pons
Departament d'Estructura i Constituents de la Matèria, Universitat de Barcelona, Diagonal 647, E-08028 Barcelona, Spain

(Received 26 April 1988; accepted for publication 28 December 1988)

New results on the theory of constrained systems are applied to characterize the generators of Noether's symmetry transformations. As a byproduct, an algorithm to construct gauge transformations in Hamiltonian formalism is derived. This is illustrated with two relevant examples.

I. INTRODUCTION

It is superfluous to emphasize the relevance of gauge theories in modern physics. In spite of this, many aspects of the classical theory of constrained systems—those which have elbow room for gauge transformations—are not completely developed. The aim of this paper is to clarify the role of the Lagrangian Noether theorem in obtaining the generators of Hamiltonian gauge transformations. This is achieved by applying some results recently obtained concerning the relationship between the Hamiltonian and Lagrangian formalisms.[1-3] These new results apply to general constrained systems, with first- and second-class constraints, under the only regularity conditions of Ref. 2.

The paper is organized as follows. In Sec. II we set the notation and summarize some of the results of Refs. 1–3; they are used in Sec. III to characterize the Hamiltonian generators of a general symmetry Noether transformation. In Sec. IV the specific case of gauge transformations is considered. Section V is devoted to some relevant applications: the bosonic string and the CP_2^{n-1} model.

All structures are supposed to be C^∞. Indices of coordinates will be omitted.

II. PRELIMINARY RESULTS

Here we state some of the results needed in Sec. III. For more details see Refs. 2 and 3. Minor changes of notation have been done.

A configuration space Q and a Lagrangian L are given. We shall always work with natural coordinates such as (q,v) in $T(Q)$ and (q,p) in $T(Q)^*$.

Then the Euler–Lagrange equations for a curve $(q(t),p(t))$ in $T(Q)$ can be written as

$$\dot{q} = v, \tag{2.1}$$

$$W\dot{v} = \alpha, \tag{2.2}$$

where we have introduced the Hessian matrix

$$W := \frac{\partial^2 L}{\partial v\, \partial v} \tag{2.3}$$

and

$$\alpha := \frac{\partial L}{\partial q} - v\frac{\partial^2 L}{\partial q\, \partial v}. \tag{2.4}$$

The Legendre transformation $FL: T(Q) \to T(Q)^*$, with the local expression

$$FL(q,v) = \left(q, \frac{\partial L}{\partial v}\right), \tag{2.5}$$

has the image $M_0 \subset T(Q)^*$, which is assumed to be a submanifold (locally) defined by the m_0 primary Hamiltonian constraints ϕ_μ^0 $(1 \leqslant \mu \leqslant m_0)$.

The vertical vector fields

$$\Gamma_\mu := \gamma_\mu \frac{\partial}{\partial v} \tag{2.6}$$

constitute a frame for the sub-bundle $\mathrm{Ker}\, T(FL) \subset T(V)$, where

$$\gamma_\mu := FL^*\left(\frac{\partial \phi_\mu^0}{\partial p}\right) \tag{2.7}$$

are a basis for the null vectors of W.

An outstanding object in our development is the operator K^2, which is now understood[4] as a vector field along FL, that is to say, it is a mapping that makes the following diagram commutative:

$$\begin{array}{ccc} & & T(T(Q)^*) \\ & \overset{K}{\nearrow} & \downarrow \tau_{T(Q)^*} \\ T(Q) & \xrightarrow[FL]{} & T(Q)^* \end{array}$$

Its local expression is

$$K(q,v) = v\frac{\partial}{\partial q} + \frac{\partial L}{\partial q}\frac{\partial}{\partial p}.$$

In fact, we shall need K in the time-dependent case, so that we shall add $\partial/\partial t$ to it:

$$K(q,v,t) = v\frac{\partial}{\partial q} + \frac{\partial L}{\partial q}\frac{\partial}{\partial p} + \frac{\partial}{\partial t}. \tag{2.8}$$

Now K can be regarded as a differential operator as follows. If f is a function in $T(Q)^* \times \mathbf{R}$,

$$K\cdot f = v\, FL^*\left(\frac{\partial f}{\partial q}\right) + \frac{\partial L}{\partial q} FL^*\left(\frac{\partial f}{\partial p}\right) + FL^*\left(\frac{\partial f}{\partial t}\right) \tag{2.9}$$

is a function in $T(Q) \times \mathbf{R}$.

Now let H be a Hamiltonian function, that is, $FL^*(H) = E_L$, where

$$E_L = v\frac{\partial L}{\partial v} - L \tag{2.10}$$

is the Lagrangian energy.

[a] Present address: Department of Physics, Princeton University, Princeton, NJ 08540.

 0022-2488/89/061345-06$02.50

One can prove[2,5] that there exist m_0 functions λ^μ in $T(Q)$ such that

$$v = FL^*\{q,H\} + \lambda^\mu FL^*\{q,\phi^0_\mu\}, \tag{2.11}$$

$$\frac{\partial L}{\partial q} = FL^*\{p,\dot{H}\} + \lambda^\mu FL^*\{p,\phi^0_\mu\}. \tag{2.12}$$

These functions λ^μ are not FL projectable since $\Gamma_\mu \cdot \lambda^\nu = \delta^\nu_\mu$. Then, it is easy to obtain

$$K\cdot f = FL^*\{f,H\} + \lambda^\mu FL^*\{f,\phi^0_\mu\} + FL^*\left(\frac{\partial f}{\partial t}\right) \tag{2.13}$$

and

$$\Gamma_\mu \cdot (K\cdot f) = FL^*\{f,\phi^0_\mu\}. \tag{2.14}$$

A careful analysis[2] of Eqs. (2.11) and (2.12) lets us write the Hamilton–Dirac equations as

$$\dot{f} \underset{M_0}{=} \{f,H\} + \eta^\mu\{f,\phi^0_\mu\} + \frac{\partial f}{\partial t}, \tag{2.15}$$

where η^μ are arbitrary functions of time.

Now derivation of (2.11) with respect to v expresses the identity matrix as

$$I = MW + \gamma_\mu \otimes \frac{\partial \lambda^\mu}{\partial v}, \tag{2.16}$$

where

$$M = FL^*\left(\frac{\partial^2 H}{\partial p\,\partial p}\right) + \lambda^\mu FL^*\left(\frac{\partial^2 \phi^0_\mu}{\partial p\,\partial p}\right). \tag{2.17}$$

Application of (2.16) to (2.1) and (2.2) leads to the introduction of time-evolution fields in $T(Q)$:

$$D_u := D_0 + u^\mu \Gamma_\mu, \tag{2.18}$$

where u^μ are arbitrary functions of time and

$$D_0 := v\frac{\partial}{\partial q} + \alpha M\frac{\partial}{\partial v} + \frac{\partial}{\partial t}. \tag{2.19}$$

Then the Euler–Lagrange equations also read

$$\dot{g} \underset{S_1}{=} D_u \cdot g, \tag{2.20}$$

where $S_1 \subset T(Q)$ is the submanifold defined by the primary Lagrangian constraints

$$\chi^1_\mu = \alpha\gamma_\mu = K\cdot\phi^0_\mu. \tag{2.21}$$

Bearing all these relations in mind one can prove that

$$K\cdot f = D_u \cdot FL^*(f) + \chi^1_\mu (Y^\mu \cdot f), \tag{2.22}$$

where we have introduced m_0 vector fields along FL:

$$Y^\mu(q,v) = \frac{\partial \lambda^\mu}{\partial v}\frac{\partial}{\partial p}. \tag{2.23}$$

Finally, we want to point out that at the present time most of these objects and relations can be defined or written intrinsically: not only (2.5) and (2.10), which are well-known,[6] but also (2.1)–(2.2), (2.8), and (2.9)[4]; (2.6)[7]; (2.11)–(2.12), (2.13), (2.22), and (2.23)[5]; and (2.21).[1,2]

III. CHARACTERIZATION OF NOETHER TRANSFORMATIONS

In the following it will be useful to enlarge our space with a third set of independent coordinates, the accelerations a; that is to say, we shall work in the second tangent bundle $T^2(Q)$.

We shall consider the operator [which maps functions in $T(Q)\times\mathbf{R}$ to functions in $T^2(Q)\times\mathbf{R}$]

$$\frac{d}{dt} = v\frac{\partial}{\partial q} + a\frac{\partial}{\partial v} + \frac{\partial}{\partial t}. \tag{3.1}$$

Then the Euler–Lagrange equations can be written as

$$[L]_{(q,\dot{q},\ddot{q})} = 0, \tag{3.2}$$

where we have defined

$$[L] := \frac{\partial L}{\partial q} - \frac{d}{dt}\left(\frac{\partial L}{\partial v}\right) = \alpha - aW. \tag{3.3}$$

Noether's theorems yield a sufficient condition for a $\delta q(q,v,t)$ to be a *dynamical symmetry transformation* (DST) of L, that is to say, to map solutions into solutions. This condition can be written as[8–10]

$$[L]\delta q + \frac{dG}{dt} = 0 \tag{3.4}$$

for certain $G(q,v,t)$. We call such a δq a *Noether transformation*. The acceleration appears linearly in (3.4), so that it splits into two relations[10–12]:

$$\alpha\,\delta q + v\frac{\partial G}{\partial q} + \frac{\partial G}{\partial t} = 0, \tag{3.5}$$

$$\frac{\partial G}{\partial v} - W\,\delta q = 0. \tag{3.6}$$

An immediate result from (3.6) is that G is an FL-projectable function since $\Gamma_\mu \cdot G = \gamma_\mu (\partial G/\partial v) = \gamma_\mu W\,\delta q = 0$. Therefore, there exists $G_h(q,p,t)$ (up to primary constraints) such that

$$G = FL^*(G_h). \tag{3.7}$$

Now we apply the operator K to G_h, bearing (2.22), (2.19), and (2.16) in mind, under the only condition (3.7). The result is

$$\begin{aligned} K\cdot G_h = {} & \chi^1_\mu \frac{\partial \lambda^\mu}{\partial v}\left(FL^*\left(\frac{\partial G_h}{\partial p}\right) - \delta q\right) \\ & + \left(\alpha\,\delta q + v\frac{\partial G}{\partial q} + \frac{\partial G}{\partial t}\right) \\ & + \alpha M\left(\frac{\partial G}{\partial v} - W\,\delta q\right). \end{aligned} \tag{3.8}$$

If G corresponds to a Noether transformation, (3.5) and (3.6) set the last two terms to zero. Moreover, assume $\delta q(q,v,t)$ to be *FL projectable*. There is $\delta q_h(q,p,t)$ (up to primary constraints) such that

$$\delta q = FL^*(\delta q_h). \tag{3.9}$$

Moreover,

$$\begin{aligned} 0 = \frac{\partial}{\partial v}(FL^*(G_h) - G) &= W\,FL^*\{q,G_h\} - \frac{\partial G}{\partial v} \\ &= W\,FL^*(\{q,G_h\} - \delta q). \end{aligned}$$

Thus there are functions $h^\mu(q,p,t)$ such that

$$\{q,G_h\} \underset{M_0}{=} \delta q_h + h^\mu \frac{\partial \phi^0_\mu}{\partial p}. \tag{3.10}$$

Redefining $G_h := G_h - h^\mu \phi_\mu^0$ we have $\{q,G_h\} \underset{M_0}{=} \delta q_h$, so that we can assume G_h and δq_h chosen in order that

$$\delta q_h = \{q,G_h\}. \tag{3.11}$$

Therefore, we conclude from (3.9) and (3.11) that (3.8) becomes

$$K \cdot G_h = 0. \tag{3.12}$$

Conversely, suppose we have $G_h(q,p,t)$ satisfying relation (3.12) and define δq_h, δq, and G as in (3.11), (3.9), and (3.7). Then we have $\partial G/\partial v = W FL^*(\partial G_h/\partial p) = W \delta q$, which is (3.6), and the identity for $K \cdot G_h$ [(3.8)] shows that (3.5) also holds; that is to say, (3.4) is satisfied. We have proven the following theorem.

Theorem 1: An infinitesimal projectable function $\delta q(q,v,t)$ is a Noether transformation if there exists $G_h(q,p,t)$ such that $K \cdot G_h = 0$ and $\delta q = FL^*\{q, G_h\}$.

Now we make use of this Lagrangian result to derive a sufficient condition for a $G_h(q,p,t)$ to generate a Hamiltonian DST in the sense that

$$\delta f = \{f,G_h\}. \tag{3.13}$$

Theorem 2: An infinitesimal function $G_h(q,p,t)$ satisfying $K \cdot G_h = 0$ generates a Hamiltonian DST.

We call such a DST a Hamiltonian Noether transformation. We have shown that $\delta q := FL^*\{q,G_h\}$ is a Lagrangian DST. Taking into account the equivalence of both formalisms,[2] we only need show $\delta(\partial L/\partial v) = FL^*\{p,G_h\}$. To this end we write the following identity, which can be obtained using (2.9) and the chain's rule:

$$\frac{\partial}{\partial v}(K \cdot G_h) = -FL^*\{p,G_h\} + \delta\frac{\partial L}{\partial v} + [L]\frac{\partial \delta q}{\partial v} + \left(\frac{\partial}{\partial t} + v\frac{\partial}{\partial q}\right)\left(\frac{\partial G}{\partial v} - W\delta q\right). \tag{3.14}$$

The last term in (3.14) vanishes because $\partial G/\partial v = W(\partial G_h/\partial p)$ and, since we transform solutions of the Euler–Lagrange equations, $[L] = 0$. Then $K \cdot G_h = 0$ implies

$$\delta\frac{\partial L}{\partial v} = FL^*\{p,G_h\}, \tag{3.15}$$

so that Theorem 2 is proven.

Let us observe that if $K \cdot G_1 = K \cdot G_2 = 0$, then $K \cdot \{G_1,G_2\} = 0$. Therefore, generators of Hamiltonian Noether transformations close under the Poisson bracket.

Finally, we want to express (3.12) in an equivalent way, which will prove to be useful in the case of gauge transformations. Application of (2.14) to (3.12) shows that $FL^*\{G_h,\phi_\mu^0\} = 0$, that is to say,

$$\{G_h,\phi_\mu^0\} \underset{M_0}{=} 0. \tag{3.16}$$

Now (2.13) leads to $FL^*(\{G_h,H\} + \partial G_h/\partial t) = 0$, which implies

$$\{G_h,H\} + \frac{\partial G_h}{\partial t} \underset{M_0}{=} 0. \tag{3.17}$$

Conversely, by (2.13), (3.16) and (3.17) imply (3.12). Therefore, the following theorem holds.

Theorem 3: A function $G_h(q,p,t)$ satisfying (3.16) and (3.17) generates a Hamiltonian DST.

It can be shown that these sufficient conditions [(3.16) and (3.17)] are in fact very close to those that are necessary.[13] Notice, also, from (3.17) that G_h is a constant of motion. Moreover, in a constrained system G_h is a first class function because it must be tangent to the final constraint manifold.

IV. HAMILTONIAN GAUGE TRANSFORMATIONS

The preceding results apply to DST in general dynamical systems. Now we consider the specific case of gauge transformations, that is to say, *DST depending on arbitrary functions and their derivatives*. Thus we are necessarily dealing with a constrained system. We will write a generator $G(q,p,t)$ of a gauge transformation in the form

$$G(q,p,t) = \sum_{k\geq 0} \epsilon^{(-k}(t) G_k(q,p), \tag{4.1}$$

where ϵ is an arbitrary function of time and $\epsilon^{(-k}(t)$ is a primitive of order k. As a result of the arbitrariness of ϵ, conditions (3.16) and (3.17) split into

$$\{\phi_\mu^0,G_k\} \underset{M_0}{=} 0, \tag{4.2}$$

$$G_0 \underset{M_0}{=} 0, \tag{4.3}$$

$$G_{k+1} + \{G_k,H\} \underset{M_0}{=} 0. \tag{4.4}$$

Relations (4.3) and (4.4) can be seen as a mechanism to construct a gauge transformation. Since G is first class, the G_k are also first class. To be precise, the G_k are first-class constraints: Let us prove this inductively; it is obvious for G_0[(4.3)]. Suppose we have chosen H to be first class (which is always possible; for instance, the $H^{(f+1)}$ reached in Ref. 2). Then if G_k is a first-class constraint, $\{G_k,H\}$ is as well. Therefore, (4.4) implies that G_{k+1} is also a first-class constraint. Notice, also, that $G_{k+1} + \{G_k,H\}$ is a primary first-class constraint.

The algorithm can be applied in the following way (see, also, Ref. 14, which proposes an algorithm to construct the gauge generator when no second class constraints are present): H is a first-class Hamiltonian and

$$G_0 = \text{primary first-class constraint}, \tag{4.5}$$

$$G_{k+1} = -\{G_k,H\} + \text{primary first-class constraints}. \tag{4.6}$$

One must play with this indeterminacy in order to let the test (4.2) hold. It is worth observing that the simpler form of a primary first-class constraint may not be suitable to begin (4.5).

There is no guarantee that this algorithm has a solution; however, it is reached in usual computations. Moreover, in these cases one can choose $G_k = 0$ for $k \geqslant f+1$ (if the stabilization algorithm ends at the fth step). For this reason the generator is usually written as

$$G = \sum_{k=0}^{f} \epsilon^{(k} G_{f-k}. \tag{4.7}$$

V. APPLICATIONS

A. The Polyakov string

The Lagrangian density of the Polyakov string is given by[15,16]

$$\mathscr{L} = (\sqrt{-g}/2) g^{\alpha\beta}\partial_\alpha x^\mu \partial_\beta x_\mu$$
$$= (-1/2\sqrt{-g})(g_{11}\dot{x}^2 - 2g_{01}(\dot{x}x') + g_{00}x'^2). \quad (5.1)$$

The canonical momenta are

$$p_\mu = \frac{\partial \mathscr{L}}{\partial \dot{x}^\mu} = \frac{-1}{\sqrt{-g}}(g_{11}\dot{x}_\mu - g_{01}x'_\mu),$$
$$\Pi^{\alpha\beta} = \frac{\partial \mathscr{L}}{\partial \dot{g}_{\alpha\beta}} = 0, \quad (5.2)$$

so that we obtain the canonical Hamiltonian density

$$\mathscr{H}_c = -(\sqrt{-g}/g_{11})H + (g_{01}/g_{11})T, \quad (5.3)$$

where $H = \frac{1}{2}(p^2 + x'^2)$ and $T = (px')$. We also obtain the primary constraints

$$\Pi_{00} = \Pi_{01} = \Pi_{11} = 0 \quad (5.4)$$

whose stability gives

$$\dot{\Pi}_{00} = \{\Pi_{00}, H_c\} = (-1/2\sqrt{-g})H,$$
$$\dot{\Pi}_{01} = \{\Pi_{01}, H_c\} = (g_{01}/\sqrt{-g}g_{11})H - (1/g_{11})T, \quad (5.5)$$
$$\dot{\Pi}_{11} = \{\Pi_{11}, H_c\} = \frac{-1}{\sqrt{-g}}\left(\frac{g_{00}}{2g_{11}} - \frac{g}{g_{11}^2}\right)H + \frac{g_{01}}{g_{11}^2}T.$$

Thus H and T are independent secondary constraints. As a result of the algebra

$$\{H(\sigma), H(\sigma')\} = T(\sigma)\partial_\sigma \delta(\sigma - \sigma') - T(\sigma')\partial_{\sigma'} \times \delta(\sigma - \sigma'),$$
$$\{H(\sigma), T(\sigma')\} = H(\sigma)\partial_\sigma \delta(\sigma - \sigma') - H(\sigma')\partial_{\sigma'} \times \delta(\sigma - \sigma'), \quad (5.6)$$
$$\{T(\sigma), T(\sigma')\} = T(\sigma)\partial_\sigma \delta(\sigma - \sigma') - T(\sigma')\partial_{\sigma'}(\sigma - \sigma'),$$

no tertiary constraints appear and we are left with five (Π_{00}, Π_{01}, Π_{11}, H, and T) first-class constraints.

We have three primary first-class constraints, so we expect three independent gauge transformations. The algorithm for constructing a canonical gauge generator starts by selecting a combination of primary first-class constraints. In order to simplify the expressions and taking into account that the three primary constraints give only two secondary constraints, let us consider the following combinations:

$$\varphi_W = g_{00}\Pi_{00} + g_{01}\Pi_{01} + g_{11}\Pi_{11},$$
$$\varphi_1 = (2\sqrt{-g}g_{01}/g_{00})\Pi_{01} + (2\sqrt{-g}g_{11}/g_{00})\Pi_{11}, \quad (5.7)$$
$$\varphi_2 = [(2g_{01}^2 - g_{00}g_{11})/g_{00}]\Pi_{01} + (2g_{01}g_{11}/g_{00})\Pi_{11},$$

which are such that

$$\dot{\varphi}_W = \{\varphi_W, H_c\} = 0,$$
$$\dot{\varphi}_1 = \{\varphi_1, H_c\} = H, \quad \dot{\varphi}_2 = \{\varphi_2, H_c\} = T. \quad (5.8)$$

Thus we see that the generator starting with φ_W has only one piece:

$$G_W = \int d\sigma\, \epsilon_W(\sigma)(g_{00}\Pi_{00} + g_{01}\Pi_{01} + g_{11}\Pi_{11}). \quad (5.9)$$

Now let us consider φ_1 and apply the algorithm

$$G_0(\sigma) = \varphi_1(\sigma),$$
$$G_1(\sigma) + \{G_0(\sigma), H_c\} = \int d\sigma'(\alpha\varphi_W(\sigma') + \beta\varphi_1(\sigma') + \gamma\varphi_2(\sigma')),$$

where $\alpha = \alpha(\sigma,\sigma')$, etc. Then

$$G_1(\sigma) = -H(\sigma) + \int d\sigma'(\alpha\varphi_W(\sigma') + \beta\varphi_1(\sigma') + \gamma\varphi_2(\sigma')).$$

The next step is

$$G_2(\sigma) + \{G_1(\sigma), H_c\} = \text{primary first-class constraints}$$

or

$$G_2(\sigma) = \{H(\sigma), H_c\} - \int d\sigma'(\beta H(\sigma') + \gamma T(\sigma')) + \text{primary first-class constraints.}$$

We need to compute

$$\{H(\sigma), H_c\} = -2T\partial_\sigma\left(\frac{\sqrt{-g}}{g_{11}}\right) + 2H\partial_\sigma\left(\frac{g_{01}}{g_{11}}\right) - \frac{\sqrt{-g}}{g_{11}}\partial_\sigma T + \frac{g_{01}}{g_{11}}\partial_\sigma H.$$

Thus we realize that we can finish the algorithm with the choice

$$\beta(\sigma,\sigma') = 2\partial_\sigma(g_{01}/g_{11})\delta(\sigma - \sigma') + (g_{01}/g_{11})(\sigma)\partial_\sigma \delta(\sigma - \sigma'),$$
$$\gamma(\sigma,\sigma') = -2\partial_\sigma(\sqrt{-g}/g_{11})\delta(\sigma - \sigma') - (\sqrt{-g}/g_{11})(\sigma)\partial_\sigma \delta(\sigma - \sigma'),$$
$$\alpha(\sigma,\sigma') = 0.$$

Then the generator has two pieces and after integrating by parts can be written as

$$G_A = \int d\sigma\Big[\dot{\epsilon}_A\varphi_1 + \epsilon_A\left(-H + \partial_\sigma\left(\frac{g_{01}}{g_{11}}\right)\varphi_1 - \partial_\sigma\left(\frac{\sqrt{-g}}{g_{11}}\right)\varphi_2\right) + \dot{\epsilon}_A\left(-\frac{g_{01}}{g_{11}}\varphi_1 + \frac{\sqrt{-g}}{g_{11}}\varphi_2\right)\Big]. \quad (5.10)$$

As a result of our choice (5.7), the consistency condition (4.2) is trivially satisfied because H, T do not depend on g's. Starting with φ_2 we could have constructed

$$G_B = \int d\sigma'\Big[\dot{\epsilon}_B\varphi_2 + \epsilon_B\left(-T - \partial_\sigma\left(\frac{\sqrt{-g}}{g_{11}}\right)\varphi_1 + \partial_\sigma\left(\frac{g_{01}}{g_{11}}\right)\varphi_1\right) + \dot{\epsilon}_B\left(\frac{\sqrt{-g}}{g_{11}}\varphi_1 - \frac{g_{01}}{g_{11}}\varphi_2\right)\Big]. \quad (5.11)$$

The action of the three generators G_W, G_A, and G_B on the fields $g_{\alpha\beta}(\sigma)$, $x_\mu(\sigma)$ yields

$$\delta g_{00} = \epsilon_W g_{00},$$

$$\delta g_{01} = \epsilon_W g_{01} + \dot{\epsilon}_A \frac{2\sqrt{-g} g_{01}}{g_{00}} + \epsilon_A \left(\frac{2g_{01}^2 - g_{00} g_{11}}{2g_{00}\sqrt{-g}} g'_{00} - g'_{01} \frac{g_{01}}{\sqrt{-g}} + g'_{11} \frac{g_{00}}{2\sqrt{-g}} \right) - \dot{\epsilon}'_A \sqrt{-g} + \dot{\epsilon}_B \frac{2g_{01}^2 - g_{00} g_{11}}{g_{00}} + \epsilon_B \left(\frac{g_{01}}{g_{00}} g'_{00} - g'_{11} \right) - \dot{\epsilon}'_B g_{01},$$

$$\delta g_{11} = \epsilon_W g_{11} + \dot{\epsilon}_A \frac{2\sqrt{-g} g_{11}}{g_{00}} + \epsilon_A \left(\frac{g_{01} g_{11}}{g_{00}\sqrt{-g}} g'_{00} - \frac{2g_{11}}{\sqrt{-g}} g'_{01} + \frac{g_{01}}{\sqrt{-g}} g'_{11} \right) + \dot{\epsilon}_B \frac{2g_{01} g_{11}}{g_{00}} + \epsilon_B \left(\frac{g_{11}}{g_{00}} g'_{00} - g'_{11} \right) - 2\dot{\epsilon}'_B g_{11}, \tag{5.12}$$

$$\delta x_\mu = \epsilon_A p_\mu - \epsilon_B x'_\mu.$$

Our canonical gauge transformations do not have the nice, well-known form

$$\delta x_\mu = \epsilon^\alpha \delta_\alpha x_\mu,$$

$$\delta g_{\alpha\beta} = \Lambda g_{\alpha\beta} + \epsilon^\gamma \partial_\gamma g_{\alpha\beta} + \partial_\alpha \epsilon^\gamma g_{\gamma\beta} + \partial_\beta \epsilon^\gamma g_{\alpha\gamma}. \tag{5.13}$$

This fact was really expected because Eq. (5.13) are not *FL* projectable: They contain the velocities $\dot{g}_{\alpha\beta}$. However, a change in the arbitrary parameters can always make the connection. In our case the change is given by

$$\epsilon_A = (\sqrt{-g}/g_{11})\epsilon^0, \quad \epsilon_B = -\epsilon^1 - (g_{01}/g_{11})\epsilon^0,$$

$$\epsilon_W = \Lambda + \epsilon^0(\dot{g}_{00}/g_{00}) + 2\dot{\epsilon}^0 + \epsilon^1(g'_{00}/g_{00}) + 2\dot{\epsilon}^1(g_{01}/g_{00}). \tag{5.14}$$

Substitution of (5.14) in (5.12) gives the covariant form (5.13). Notice that relations (5.14) involve non-*FL*-projectable functions, as must occur. Also, notice the fact that the first-class constraints T, H satisfy a nontrivial algebra, making the first-class primary constraints φ_1, φ_2 enter the generator in a definite way and giving the correct gauge transformations, so that the canonical gauge generator is not simply an arbitrary combination of first-class constraints. The first class Hamiltonian we have used is simply the canonical Hamiltonian because no second-class constraints are present. The procedure is less trivial in the example in Sec. V B.

B. The CP_2^{n-1} model.

The Lagrangian density is[17]

$$\mathscr{L} = (D_\mu Z_\alpha)^*(D^\mu Z_\alpha) - \lambda(Z_\alpha^* Z_\alpha - n/2g), \tag{5.15}$$

where $D_\mu = \partial_\mu + iA_\mu$ and $g^{\mu\sigma} = \mathrm{diag}(\pm)$. Here A_μ is a two-dimensional auxiliary gauge field and λ is a field which enforces the condition $Z_\alpha^* Z_\alpha = n/2g$ on the n complex fields Z_α.

The infinitesimal gauge invariance of the theory is given by

$$\delta Z_\alpha = -i\theta Z_\alpha, \quad \delta Z_\alpha^* = i\theta Z_\alpha^*, \quad \delta A_\mu = \partial_\mu \theta, \quad \delta\lambda = 0, \tag{5.16}$$

where $\theta = \theta(x^0, x^1) = \theta(\tau,\sigma)$ is an arbitrary parameter. The canonical momenta are

$$\Pi_\mu = \frac{\partial \mathscr{L}}{\partial \dot{A}^\mu} = 0, \quad \Pi_\lambda = \frac{\partial \mathscr{L}}{\partial \dot{\lambda}} = 0,$$

$$\Pi_\alpha = \frac{\partial \mathscr{L}}{\partial \dot{Z}_\alpha} = \dot{Z}_\alpha^* - iZ_\alpha^* A_0, \quad \Pi_\alpha^* = \frac{\partial \mathscr{L}}{\partial \dot{Z}_\alpha^*} = \dot{Z}_\alpha + iZ_\alpha A_0. \tag{5.17}$$

Thus we obtain the canonical Hamiltonian density

$$\mathscr{H}_c = \Pi_\alpha \Pi_\alpha^* - iA_0(\Pi_\alpha Z_\alpha - \Pi_\alpha^* Z_\alpha^*) + \partial_1 Z_\alpha^* \partial_1 Z_\alpha + iA_1(Z_\alpha \partial_1 Z_\alpha^* - Z_\alpha^* \partial_1 Z_\alpha) + A_1^2 Z_\alpha^* Z_\alpha + \lambda(Z_\alpha^* Z_\alpha - n/2g) \tag{5.18}$$

and the primary constraints

$$\Pi_0 = 0, \quad \Pi_1 = 0, \quad \Pi_\lambda = 0. \tag{5.19}$$

Then the primary Hamiltonian density is

$$\mathscr{H}_p = \mathscr{H}_c + v_0\Pi_0 + v_1\Pi_1 + v_\lambda \Pi_\lambda, \tag{5.20}$$

where $v_\mu = \dot{A}_\mu$ and $v_\lambda = \dot{\lambda}$. After the stability algorithm is performed it turns out that the theory contains two first-class constraints,

$$\varphi_1 = \Pi_0, \quad \varphi_2 = i(\Pi_\alpha Z_\alpha - \Pi_\alpha^* Z_\alpha^*) - \partial_1 \Pi_1 \tag{5.21a}$$

and six second-class constraints,

$$\chi_1 = Z_\alpha \Pi_\alpha + Z_\alpha^* \Pi_\alpha^*,$$

$$\chi_2 = \lambda Z_\alpha^* Z_\alpha - \Pi_\alpha \Pi_\alpha^* + \partial_1 Z_\alpha^* \partial_1 Z_\alpha + iA_1(Z_\alpha^* \partial_1 Z_\alpha - Z_\alpha \partial_1 Z_\alpha^*) - A_1^2 Z_\alpha^* Z_\alpha, \tag{5.21b}$$

$$\chi_3 = \Pi_\lambda,$$

$$\chi_4 = Z_\alpha^* Z_\alpha - n/2g,$$

$$\chi_5 = \Pi_1,$$

$$\chi_6 = iZ_\alpha \partial_1 Z_\alpha^* - iZ_\alpha^* \partial_1 Z_\alpha + 2A_1 Z_\alpha^* Z_\alpha.$$

Because the primary constraints Π_1, Π_λ have become second class, the arbitrary functions $v_1 = \dot{A}_1$, $v_\lambda = \dot{\lambda}$ have been canonically determined:

$$v_1 = (g/n)(iZ_\alpha \partial_1 \Pi_\alpha + i\Pi_\alpha^* \partial_1 Z_\alpha - iZ_\alpha^* \partial_1 \Pi_\alpha^* - i\Pi_\alpha \partial_1 Z_\alpha + 2A_1(Z_\alpha \Pi_\alpha + Z_\alpha^* \Pi_\alpha^*)) - \partial_1 A_0 \equiv f_1 \tag{5.22a}$$

and

$$v_\lambda = \frac{-2g}{n}\left(\{\chi_2, H_c\} + \int d\sigma' v_1(\sigma')\{\chi_2, \chi_5\} \right) \equiv f_\lambda. \tag{5.22b}$$

Thus the first-class Hamiltonian density is

$$\mathscr{H} = \mathscr{H}_c + f_1\Pi_1 + f_\lambda \Pi_\lambda, \tag{5.23}$$

which incorporates the second-class primary constraints in the correct way. Now we can begin the algorithm with

$$G_0(\sigma) = \Pi_0(\sigma)$$

and

$$G_1(\sigma) = -\{G_0(\sigma),H\} + \text{primary first-class constraints}$$
$$= -i(\Pi_\alpha Z_\alpha - \Pi^*_\alpha Z^*_\alpha)$$
$$+ \partial_1 \Pi_1 + \text{primary first-class constraints.}$$

It can be checked that $\{-i(\Pi_\alpha Z_\alpha - \Pi^*_\alpha Z^*_\alpha) + \partial_1 \Pi_1, H\} = 0$. Thus the algorithm ends at this stage and the gauge generator is

$$G = \int d\sigma [\dot{\theta}\Pi_0 - i\theta(\Pi_\alpha Z_\alpha - \Pi^*_\alpha Z^*_\alpha) - \partial_1 \theta \Pi_1], \tag{5.24}$$

which gives the correct gauge transformations. In this condition (4.2) is trivially satisfied in a natural way.

ACKNOWLEDGMENT

This work has been partially supported by CAICYT Project No. AE 87-0016-3.

[1]C. Batlle, J. Gomis, J. M. Pons, and N. Román-Roy, Lett. Math. Phys. **13**, 17 (1987).
[2]C. Batlle, J. Gomis, J. M. Pons, and N. Román-Roy, J. Math. Phys. **27**, 2953 (1986).
[3]J. M. Pons, J. Phys. A: Math. Gen. **21**, 2705 (1988).
[4]X. Gràcia and J. M. Pons, preprint UB-ECM-PF 7/87, Universitat de Barcelona.
[5]X. Gràcia, Universitat de Barcelona, preprint, April 1988.
[6]R. Abraham and J. E. Marsden, *Foundations of Mechanics* (Addison-Wesley, Reading, MA, 1978), 2nd ed.
[7]J. F. Cariñena, C. López, and N. Román-Roy, J. Geom. Phys. **4**, 315 (1987).
[8]E. Noether, Nachr. Ges. Wiss. Gött. **2**, 235 (1918).
[9]E. L. Hill, Rev. Mod. Phys. **23**, 253 (1951).
[10]J. Gomis, K. Kaminura, and J. M. Pons, Prog. Theor. Phys. **71**, 1051 (1984).
[11]K. Kamimura, Nuovo Cimento B **68**, 33 (1982).
[12]J. Gomis, K. Kamimura, and J. M. Pons, Europhys. Lett. **2**, 187 (1986).
[13]X. Gràcia and J. M. Pons, Ann. Phys. **187**, 355 (1988).
[14]L. Castellani, Ann. Phys. **143**, 357 (1982).
[15]L. Brink, P. Di Vecchia, and P. Howe, Phys. Lett. B **65**, 471 (1976).
[16]C. Batlle, J. Gomis, and J. M. Pons, Phys. Rev. D **34**, 2430 (1986).
[17]E. Gozzi and A. Guha, J. Math. Phys. **24**, 1213 (1983).

n^k almost-tangent structures and the Hamiltonization of higher-order field theories

Manuel de León
CECIME, Consejo Superior de Investigaciones Científicas, Serrano 123, 28006 Madrid, Spain

Paulo R. Rodrigues
Departamento de Geometria, Instituto de Matematica, Universidade Federal Fluminense, 24000 Niteroi, Rio de Janeiro, Brazil

(Received 17 May 1988; accepted for publication 11 January 1989)

In this paper a new interpretation of the Hamiltonization of higher-order Lagrangian field theories is given, introducing a new class of tensor fields of type (1, 1) that extends almost-tangent structures to the tangent bundle of n^k velocities.

I. INTRODUCTION

This paper is the third contribution to a geometrical study of higher-order Lagrangian and Hamiltonian formalisms in many independent variables (see de León and Rodrigues[1,2] for notations and results). In Ref. 1 we introduced an intrinsical version of the Poincaré–Cartan form and in Ref. 2 we examined a possible relation between extremals defined by Langrangian and Hamiltonian variational problems of higher order. As expected, there is no way to establish a Legendre transformation between both formalisms which is at least a local diffeomorphism: This was examined in terms of the geometrical approach of higher-order variational theory.

Here we will adopt a different procedure without using the variational theory. As remarked in the Introduction of Ref. 1, we were inspired by almost-tangent geometry to give an intrinsic definition of the Poincaré–Cartan form. Here we will introduce a new class of tensor fields of type (1, 1) given by a family of endomorphisms J_a, $1\leqslant a\leqslant n$ of $T(T_n^k M)$, where $T_n^k M$ is the tangent bundle of n^k velocities of M. The intrinsic formulation of higher-order theories is then clearly obtained. For instance, the definition of the Poincaré–Cartan form involves an operator J_β, where β is a multi-index

$$(\beta_1,...,\beta_n) \text{ and } J_\beta = J_1^{\beta_1}\cdots J_n^{\beta_n}.$$

Almost-tangent structures were introduced by Clark and Bruckheimer[3] and Eliopoulos[4] around 1960. Klein[5] showed that almost-tangent structures have an essential role in Lagrangian theories like that of the role of symplectic structures in Hamiltonian theories (for example, every tangent bundle of a finite differentiable manifold is endowed with a canonical almost-tangent structure).

The paper is organized as follows. In Sec. II we recall some definitions and results about tangent bundles of n^k velocities. In Secs. III and IV we examined $\bar{\pi}_r^k$ semibasic forms and the Legendre transformation defined by such forms. Our construction gives a geometrical interpretation of the concept of regularity proposed in Ref. 2. We conclude the paper by studying the Hamiltonization proposed by Podolski and co-workers[6–8] for an electromagnetic theory with second-order field variables.

We would like to remark that our study is developed on the bundle $J^k(R^n,M) = R^n \times T_n^k M$. Nevertheless, if we consider the more general situation of k-jet prolongations $J^k(N,E)$ of locally trivial fiber bundles (E,π,N), then Morimotos's[9] theory fails since we have no directions on the base manifold N for lifting tensor fields vertically. In such a case we adopt Saunder's[10] approach using the operators S_ω parametrized by closed one-forms on N to construct global Poincaré–Cartan forms. It is easy to see that if $N = R^n$, $E = R^n \times M$, then $S_{dx_a} = \bar{J}_a^*$, $1\leqslant a\leqslant n$, where the star stands for adjoint operators acting on forms.

II. PRELIMINARIES

Let M be a manifold of dimension m and R^n the Euclidean space with the coordinates $(x_1,...,x_n) = (x_a)$. We donote by $T_n^k M$ the tangent bundle of n^k velocities, i.e., $T_n^k M = J_0^k(R^n,M)$ is the manifold of all jets of order k of C^∞ mapping from R^n to M at the origin $0\in R^n$. Let us denote by $\eta(n,k)$ the set of all n-tuples $\alpha = (\alpha_1,...,\alpha_n)$ of non-negative integers such that $|\alpha| = \alpha_1 + \cdots + \alpha_n \leqslant k$. We set

$$\alpha+\beta = (\alpha_1,...,\alpha_n) + (\beta_1,...,\beta_n) = (\alpha_1+\beta_1,...,\alpha_n+\beta_n),$$

for all $\alpha,\beta\in\eta(n,k)$,

$$(a) = (0,...,1,...,0),$$

with 1 in the ath place.

If (z^A) are local coordinates on M, then we denote by (z_α^A) the induced coordinates on $T_n^k M$ defined by

$$z_\alpha^A(j_0^k\sigma) = \left(\frac{1}{\alpha!}\right)\left(\frac{\partial}{\partial x^\alpha}\right)(z^A\circ\sigma)\Big|_{x=0},$$

where $\alpha! = (\alpha_1)!\cdots(\alpha_n)!$ and $\partial/\partial x^\alpha = \partial^{|\alpha|}/(\partial x_1)^{\alpha_1}\cdots(\partial x_n)^{\alpha_n}$.

Alternatively, we shall use the coordinates (y_α^A) on $T_n^k M$, with $y_\alpha^A = (\alpha!\, z_\alpha^A)$.

As we have seen in Ref. 1 there exist a family J_β, $0\leqslant|\beta|\leqslant k$ of canonical tensor fields of type (1, 1) on $T_n^k M$ locally given by

$$J_\beta = \sum_{|\alpha|=0}^{k-|\beta|}\left(\frac{\partial}{\partial z_{\alpha+\beta}^A}\right)\otimes(dz_\alpha^A).$$

III. SEMIBASIC FORMS

Let us denote by

$$\rho_r^k: R^n\times T_n^k M\to R^n\times T_n^r M, \quad \pi_r^k: T_n^k M\to T_n^r M$$

 0022-2488/89/061351-03$02.50

the obvious projections $(r<k)$, $\rho_r^k = Id_{R^n} \times \pi_r^k$. We set

$$\bar{\pi}_r^k\colon R^n \times T_n^k M \to T_n^r M$$

for the projection given by $pr_2 \circ \rho_r^k$, where

$$pr_2 : R^n \times T_n^r M \to T_n^r M.$$

Definition: Let Λ be a one-form on $R^n \times T_n^k M$. Then Λ is called $\bar{\pi}_r^k$ semibasic if it vanishes on all vertical vector fields with respect to the fibration

$$\bar{\pi}_r^k : R^n \times T_n^k M \to T_n^r M,$$

$$(x_a, y_\alpha^A)_{|\alpha| \leqslant k} \to (y_\alpha^A)_{|\alpha| \leqslant r}.$$

It is easy to see that Λ is $\bar{\pi}_r^k$ semibasic if and only if it is locally expressed by

$$\Lambda = \sum_{|\alpha|=0}^{r} \Lambda_A^\alpha (x_a, y_\beta^B)(dy_\alpha^A), \quad 1 \leqslant A, \quad B \leqslant m, \quad 0 \leqslant |\beta| \leqslant k.$$

We recall that the $\bar{\pi}_r^k$ vertical bundle $V = \ker T\bar{\pi}_r^k$ is locally spanned by the vectors fields $\langle (\partial/\partial x_a), \ (\partial/\partial y_\alpha^A);$ $r+1 \leqslant |\alpha| \leqslant k \rangle$.

Theorem: Let

$$q_{T_n^r M}\colon T^*(T_n^r M) \to T_n^r M$$

be the canonical projection of the cotangent bundle $T^*(T_n^r M)$ onto $T_n^r M$ and $Id \times q_{T_n^r M}$ the obvious induced projection

$$R^n \times T^*(T_n^r M) \to R^n \times T_n^r M.$$

If Λ is a $\bar{\pi}_r^k$ semibasic form on $R^n \times T_n^k M$, then there is a mapping

$$D : R^n \times T_n^k M \to R^n \times T^*(T_n^r M)$$

such that $(Id \times q_{T_n^r M}) \circ D = \rho_r^k$.

Proof: In fact, $D(x,v) = (x,p)$, where $p \in T^*_{\pi_r^k(v)}(T_n^r M)$ is determined by

$$p(X) = \Lambda_{(x,v)}(\bar{X}),$$

where $X \in T_{\pi_r^k(v)}(T_n^r M)$, $\bar{X} \in T_{(x,v)}(R^n \times T_n^k M)$, and $T\pi_r^k(\bar{X}) = X$. Then we have

$$D(x_a, y_\alpha^A)_{0<|\alpha|\leqslant k} = (x_a, \Lambda_A^\alpha)_{0<|\alpha|\leqslant r}. \qquad \square$$

Corollary: If λ is the Liouville form on $T^*(T_n^r M)$, then $(pr_2 \circ D)^* \lambda = \Lambda$, where pr_2: $R^n \times T^*(T_n^r M) \to T^*(T_n^r M)$ is the canonical projection on the second factor.

Remark: Similarly, we can consider π_r^k semibasic one-forms on $T_n^k M$ and prove that a π_r^k semibasic one-form Λ on $T_n^k M$ defines a mapping D: $T_n^k M \to T^*(T_n^r M)$ such that $q_{T_n^r M} \circ D = \pi_r^k$.

Then if Λ is locally given by

$$\Lambda = \sum_{|\alpha|=0}^{r} \Lambda_A^\alpha (y_\beta^B) dy_\alpha^A,$$

we easily obtain that D is given by

$$D : (y_\alpha^A)_{0<|\alpha|\leqslant k} \to (y_\alpha^A, \Lambda_A^\alpha)_{0<|\alpha|\leqslant r}.$$

Clearly, we have $D^* \lambda = \Lambda$.

IV. THE LEGENDRE TRANSFORMATION

Let us first recall the following construction of fibered bundles. Suppose that (E_a, π_a, N) are fibered bundles, $1 \leqslant a \leqslant n$, and consider the Cartesian product $(E_1 \times \cdots \times E_n, \pi_1 \times \cdots \times \pi_n, N \times \cdots \times N)$ of such bundles. Let Δ: $N \to N \times \cdots \times N$ be the diagonal map $x \to (x,\ldots,x)$. Then one obtains a fibered bundle denoted by $E_1 + \cdots + E_n$ such that

$$\begin{array}{ccc} E_1 + \cdots + E_n & \to & E_1 \times \cdots \times E_n \\ \downarrow & & \downarrow \pi_1 \times \cdots \times \pi_n \\ N & \xrightarrow{\Delta} & N \times \cdots \times N \end{array}$$

is commutative. If each (E_a, π_a, N) is a vector bundle, then $E_1 + \cdots + E_n = E_1 \oplus \cdots \oplus E_n$ is the Whitney sum.

Now, in Ref. 1 we have shown that the Poincaré–Cartan form is defined by

$$\Omega_L = \left[\sum_{|\beta|=0}^{k-1} (-1)^{|\beta|} \left(\frac{1}{(\beta + (a))!} \right) d_{T_\beta}(d_{J_{\beta+(a)}} L) \right] \wedge \omega_a + L\omega. \tag{1}$$

If we use the fact that

$$\bar{J}_{\beta+(a)} = J_{\beta+(a)} - J_\beta C_{ba} \otimes dx_b,$$

then (1) takes the form

$$\Omega_L = \left[\sum_{|\beta|=0}^{k-1} (-1)^{|\beta|} \left(\frac{1}{(\beta + (a))!} \right) d_{T_\beta}(d_{J_{\beta+(a)}} L) \right] \wedge \omega_a - E_L \omega, \tag{2}$$

where the remaining terms are incorporated in the energy E_L of L after the substitution of $\bar{J}_{\beta+(a)}$. At the local level we have

$$\begin{aligned} \Omega_L &= \sum_{|\beta|=0}^{k-1} p_A^{\beta+(a)} \theta_\beta^A \wedge \omega_a + L\omega \\ &= \sum_{|\beta|=0}^{k-1} p_A^{\beta+(a)} dy_\beta^A \wedge \omega_a - E_L \omega, \end{aligned} \tag{3}$$

where E_L is given by

$$E_L = \sum_{|\beta|=0}^{k-1} p_A^{\beta+(a)} y_{\beta+(a)}^A - L. \tag{4}$$

If we set

$$\Lambda_a = \sum_{|\beta|=0}^{k-1} (-1)^{|\beta|} \left(\frac{1}{(\beta + (a))!} \right) d_{T_\beta}(d_{J_{\beta+(a)}} L), \quad 1 \leqslant a \leqslant n,$$

then Λ_a is locally given by

$$\Lambda_a = \sum_{|\beta|=0}^{k-1} p_A^{\beta+(a)} dy_\beta^A.$$

Hence one has that Λ_a is a $\bar{\pi}_{k-1}^{2k-1}$ semibasic one-form on $R^n \times T_n^{2k-1} M$. From the above theorem there is a fibered morphism $D_a = \mathrm{Leg}_a$ from $R^n \times T_n^{2k-1} M$ to $R^n \times T^*(T_n^{k-1} M)$ such that

$$(Id \times q_{T_n^{k-1} M}) \circ \mathrm{Leg}_a = \rho_{k-1}^{2k-1}, \quad 1 \leqslant a \leqslant n,$$

which is locally given by

$$(x_b, y_\alpha^A)_{0<|\alpha|\leqslant 2k-1} \to (x_b, y_\alpha^A, p_A^{\alpha+(a)})_{0<|\alpha|\leqslant k-1}, \quad 1 \leqslant b \leqslant n.$$

Now, if we take $E_a = R^n \times T^*(T_n^{k-1} M)$, π_a

$= Id \times q_{T_n^{k-1}M}$, and $N = R^n \times T_n^{k-1}M$ one obtains n vector bundles with the fiber $T_n^{k-1}R^m$, $m = \dim M$. Therefore, we may define a fibered mapping

$$R^n \times T_n^{2k-1}M \xrightarrow{\text{Leg}} (R^n \times T^\star(T_n^{k-1}M)) \oplus \cdots \oplus (R^n \times T^\star(T_n^{k-1}M)) \tag{5}$$

defined by

$$\text{Leg}(x,v) = (\text{Leg}_1(x,v),\ldots,\text{Leg}_n(x,v)).$$

We call Leg the Legendre transformation of L. For example, if $k = 2$, Leg is locally given by

$$(x_b, y_\alpha^A) \to ((x_b, y_\beta^A, p_A^{\beta+(1)}),\ldots,(x_b, y_\beta^A, p_A^{\beta+(n)})),$$
$$0 \leqslant |\alpha| \leqslant 3, \quad 0 \leqslant |\beta| \leqslant 1,$$

where

$$p_A^{(a)} = \left(\frac{\partial L}{\partial y_{(a)}^A}\right) - d_{T_b}\left(\frac{\partial L}{\partial y_{(a)+(b)}^A}\right),$$
$$p_A^{(a)+(b)} = \left(\frac{\partial L}{\partial y_{(a)+(b)}^A}\right). \tag{6}$$

As $R^n \times T_n^{2k-1}M$ and the Whitney sum on the rhs of (5) are fibered over $R^n \times T_n^{k-1}M$ we may examine the rank of each Leg_a (i.e., the rank of the functions $p_A^{\alpha+(a)}$, $0 \leqslant |\alpha| \leqslant k-1$) with respect to the variables y_α^A, $k \leqslant |\alpha| \leqslant 2k-1$. If we suppose that for every $a = 1,\ldots,n$ the rank is maximal, then Leg is a submersion. This construction gives a geometrical interpretation of the concept of *regularity* proposed in Ref. 2. If Leg is a submersion, then taking a section s of Leg, one obtains the Hamiltonization of the theory by defining $H = E_L \circ s$.

Remark: The reader is invited to compare the present procedure with the one proposed by Shadwick.[11]

V. THE HAMILTONIZATION OF PODOLSKI

In what follows we will set $(a_1) + \cdots (a_n) = (a_1,\ldots,a_n)$. We will take $k = 2$, $n = 4$, and $m = 4$. The Podolski formalism (see, for example, de León and Rodrigues,[12] p. 145 or Podolski and Schwed[8]) starts with a Lagrangian $L = L(y_\alpha^A)$, $0 \leqslant |\alpha| \leqslant 2$, where the independent variables $(x_1, x_2, x_3, x_4 = ict)$ are implicitly defined in L. The Podolski Hamiltonization is obtained by the introduction of the coordinates $q_A = y^A$. The momenta are defined by

$$p_{A/1} = \frac{\partial L}{\partial \dot{q}_A} - \left(\frac{\partial}{\partial t}\right)\left(\frac{\partial L}{\partial \ddot{q}_A}\right) - \left(\frac{\partial}{\partial x_j}\right)\left(\frac{\partial L}{\partial \dot{q}_{A,j}}\right), \tag{7}$$
$$p_{A/2} = \frac{\partial L}{\partial \ddot{q}_A},$$

where the dots are time derivatives and $1 \leqslant j \leqslant 3$. The Hamiltonian is then defined by

$$H = p_{A/1}\dot{q}_A + p_{A/2}\ddot{q}_A - L. \tag{8}$$

In what follows it is convenient to use the relations

$$\frac{\partial}{\partial x_4} = (ic)^{-1}\left(\frac{\partial}{\partial t}\right), \quad \dot{q}_A = (ic)y_{(4)}^A,$$
$$\ddot{q}_A = (ic)^2 y_{(44)}^A, \qquad \dot{q}_{A,j} = (ic)y_{(4j)}^A,$$

$$p_{A/1} = (ic)^{-1}\left[\left(\frac{\partial L}{\partial y_{(4)}^A}\right) - \sum_{b=1}^{4}\left(\frac{\partial}{\partial x_b}\right)\left(\frac{\partial L}{\partial y_{(4b)}^A}\right)\right],$$
$$P_{A/2} = ((ic)^2)^{-1}\left(\frac{\partial L}{\partial y_{(44)}^A}\right).$$

Thus using our previous notations (6) we have

$$(ic)p_{A/1} = p_A^{(4)}, \quad (ic)^2 p_{A/2} = p_A^{(44)}$$

and the Hamiltonian is

$$H = p_A^{(4)} y_{(4)}^A + p_A^{(44)} y_{(44)}^A - L. \tag{9}$$

Now, we may interpret (8) or (9) as follows. If H is only a function of $(y^A, y_{(4)}^A, p_A^{(4)}, p_A^{(44)})$, then we are in the presence of a second-order theory, with the Legendre transformation defined on the tangent bundle of order 3 with values on the cotangent bundle of TM:

$$\text{Leg: } T_1^3 M \to T^\star(TM)$$

(since the Hamiltonization is only performed with a single independent variable). The map Leg is the Legendre transformation previously studied by the present authors in the context of higher-order particle mechanics (see de León and Rodrigues,[12] p. 108). However, Podolski considered the Hamiltonian (9) as being of the type

$$H = H(y^A, y_{(u,v)}^A, y_{(u,4)}^A, p_A^{(4)}, p_A^{(44)}), \quad 1 \leqslant u,v \leqslant 3. \tag{10}$$

In such a case the Hamiltonization is obtained by pulling back the original Lagrangian L defined on $T_4^2 M$ to $T_4^3 M$ via the canonical projection ρ_2^3. The Legendre transformation for $(\rho_2^3)^\star L$ is then a map

$$\text{Leg: } T_4^5 M \to T^\star(T_4^2 M) \oplus \cdots \oplus T^\star(T_4^2 M) \text{ (four-times)}, \tag{11}$$

which is not a submersion. In fact, we consider only the fourth term in the Whitney sum on the rhs of (11), for which the local coordinates are $(y_\alpha^A, p_A^{\alpha+(4)})$, $0 \leqslant |\alpha| \leqslant 2$. Then we consider a section s_4 defined on some submanifold of $T^\star(T_4^2 M)$, locally characterized by constraint relations of the type

$$p_A^{(a)} = p_A^{(ab)} = 0, \quad 1 \leqslant a, \quad b \leqslant 3.$$

ACKNOWLEDGMENT

One of the authors (PRR) gratefully acknowledges the hospitality provided by CECIME, Consejo Superior de Investigaciones Científicas, Spain, where this work has been performed.

The work of PRR was partially supported by CNPq-Brazil, Proc. MA 30.1115/79.

[1] M. de León and P. R. Rodrigues, Lett. Math. Phys. **14**, (4), 353 (1987).
[2] M. de León and P. R. Rodrigues, to be published in Bol. Acad. Galega de Ciencias.
[3] R. Clark and M. Bruckheimer, C. R. Acad. Sci. Paris **251**, 627 (1960).
[4] H. Eliopoulos, C. R. Acad. Sci. Paris **255**, 1563 (1962).
[5] J. Klein, Ann. Inst. Fourier (Grenoble) **12**, 1 (1962).
[6] B. Podolski, Phys. Rev. **62**, 68 (1942).
[7] B. Podolski and C. Kikuchi, Phys. Rev. **65**, 228 (1944).
[8] B. Podolski and P. Schwed, Rev. Mod. Phys. **20**, 40 (1948).
[9] A. Morimoto, *Prolongations of geometric structures*, Lect. Notes, Math. Inst., Nagoya Univ. (Nagoya, 1969).
[10] D. J. Saunders, J. Phys. A: Math. Gen. **20**, 339 (1987).
[11] W. F. Shadwick, Lett. Math. Phys. **6**, 409 (1982).
[12] M. de León and P. R. Rodrigues, *Generalized Classical Mechanics and Field Theory*, North-Holland Math. Ser. 112 (North-Holland, Amsterdam, 1985).

The nonexistence of the path-space measure for the Dirac equation in four space-time dimensions

Tomasz Zastawniak[a)]
Mathematics Department, University of Warwick, Coventry, Warwickshire CV4 7AL, United Kingdom

(Received 22 April 1988; accepted for publication 18 January 1989)

It is proved that the path-space measure for the Dirac equation in four space-time dimensions does not exist. The origin of the nonexistence of the measure turns out to be the dependence of the solution on the first derivative of the initial condition.

I. INTRODUCTION

Consider the Cauchy problem for the Dirac equation describing the evolution of a free relativistic particle in two space-time dimensions,

$$\begin{aligned}&[\partial_t + \alpha\partial_x]\psi(t,x) = -im\beta\psi(t,x),\\ &\psi(0,x) = \phi(x),\end{aligned} \tag{1}$$

where ψ and ϕ are $\mathbb{C}^2$-valued functions, and $\alpha,\beta\in\mathbb{C}^{2\times 2}$ are constant matrices satisfying the relations $\alpha^2 = \beta^2 = 1$ and $\alpha\beta + \beta\alpha = 0$. There exists a $\mathbb{C}^{2\times 2}$-valued Borel measure ν_t on the space of paths $C_t = \{X\in C([0,t],\mathbb{R}): X(t) = 0\}$, such that the solution of Cauchy problem (1) is given by the formula

$$\psi(t,x) = \int_{C_t} d\nu_t(X)\phi(x + X(0)). \tag{2}$$

There are several methods of constructing the measure ν_t. The first to define ν_t was Ichinose,[1] who applied a Riesz-type representation theorem. Blanchard *et al.*[2] obtained a representation of the solution to (1) in terms of the probabilistic expectation and a Markovian stochastic process associated with Dirac equation (1). This representation leads immediately to a definition of the measure ν_t, and to formula (2). Zastawniak[3] gave another equivalent definition of ν_t, which was obtained by considering the expansion of the solution ψ in powers of the mass m of the particle. It turns out[4] that the measure ν_t is concentrated on the set of so-called zigzag paths, i.e., such functions $X\in C_t$ that the derivative $X'(s)$ exists for all but a finite number of $s\in[0,t]$ and equals ± 1 times the speed of light (which is 1 in our units). This can be regarded as a formulation of the phenomenon of the Zitterbevegung of a Dirac particle in terms of the path-space measure. The integral with respect to the matrix-valued path-space measure ν_t has got all the properties suggested by Feynman in his brief description[5] of the path integral for the Dirac equation. Moreover, if the particle is subjected to an external electromagnetic field, the Dirac equation becomes

$$\begin{aligned}&[\partial_t - iV(t,x)]\psi(t,x)\\ &\quad + \alpha[\partial_x - iA(t,x)] = -im\beta\psi(t,x).\end{aligned}$$

With V, A: $[0,\infty)\times\mathbb{R}\to\mathbb{R}$ being arbitrary (sufficiently regular) electromagnetic potentials, the solution ψ to the initial problem $\psi(0,x) = \phi(x)$ is given by the formula[6]

$$\begin{aligned}\psi(t,x) = \int_{C_t} d\nu_t(X)\exp\Big[i\int_0^t V(s,x + X(s))ds\\ + i\int_0^t A(s,x + X(s))dX(s)\Big]\phi(x + X(0)),\end{aligned}$$

which resembles the famous Feynman–Kac–Ito formula.[7]

It is crucial for the construction of the measure ν_t that there exists an L^∞ estimate of the solution to Cauchy problem (1). According to Ichinose,[8] the above results seemed not to generalize easily to the four-dimensional case for the lack of a theorem on the L^∞-well posedness of the Cauchy problem for the Dirac equation in four space-time dimensions.[9] The aim of the present paper is to prove that, in fact, such a generalization is impossible. To be more precise, let us write down the initial Cauchy problem for the Dirac equation in four space-time dimensions,

$$\begin{aligned}&(\partial_t + \alpha\cdot\nabla)\psi(t,x) = -im\beta\psi(t,x),\\ &\psi(0,x) = \phi(x).\end{aligned} \tag{3}$$

Here ϕ and ψ are $\mathbb{C}^4$-valued functions, and $\alpha\cdot\nabla = \alpha_1\partial_{x_1} + \alpha_2\partial_{x_2} + \alpha_3\partial_{x_3}$, where $\alpha_1,\alpha_2,\alpha_3,\beta\in\mathbb{C}^{4\times 4}$ are constant matrices satisfying the relations $\beta^2 = 1$, $\alpha_k\beta + \beta\alpha_k = 0$, and $\alpha_k\alpha_l + \alpha_l\alpha_k = \delta_{k,l}$ for $k,l = 1,2,3$. To generalize the results of Refs. 1–3, one should look for a 4×4 matrix-valued measure ν_t that is defined on the path-space $C_t = \{X\in C([0,t],\mathbb{R}^3): X(t) = 0\}$ and satisfies a formula analogous to (2). It is proved in the next section that such a measure ν_t does not exist.

It will be convenient to use the so-called Weyl representation of the matrices α_k and β,

$$\alpha_k = \begin{bmatrix}\sigma_k & 0\\ 0 & -\sigma_k\end{bmatrix},\quad \beta = \begin{bmatrix}0 & 1\\ 1 & 0\end{bmatrix}, \tag{4}$$

where 1 and 0 stand for the identity and null 2×2 matrices, respectively, and for $k = 1,2,3$, σ_k are the Pauli matrices. The Cauchy problem (3) for the Dirac equation in the Weyl representation can be rewritten in the form

$$\begin{aligned}&(\partial_t + j\sigma\cdot\nabla)\psi_j(t,x) = -im\psi_{-j}(t,x),\\ &\psi_j(0,x) = \phi_j(x),\end{aligned} \tag{5}$$

where $j = \pm 1$, and ψ_j, ϕ_j are $\mathbb{C}^2$-valued functions.

If the mass parameter $m = 0$, problem (5) can be understood as a system of two Cauchy problems for two independent differential equations, called the Weyl equations. The problems are indexed by $j = \pm 1$, and have the form

[a)] On leave from Institute of Mathematics, Jagiellonian University, Reymonta 4, 30-059 Kraków, Poland.

 0022-2488/89/061354-05$02.50

$$(\partial_t + j\sigma\cdot\nabla)\psi(t,x) = 0,$$
$$\psi(0,x) = \phi(x), \tag{6}$$

where both the solution ψ and the initial condition ϕ are $\mathbb{C}^2$-valued functions. The next section will start with a discussion of the Weyl equation, in which case we shall show that there is no corresponding 2×2 matrix-valued path-space measure. The proof for the Weyl equation shows clearly the reason why such a measure does not exist. Afterwards, using an expansion of the solution in powers of the mass m, we shall extend the argument to the Dirac equation.

The theorem on the nonexistence of the path-space measure for the Dirac (and Weyl) equation is similar to the well-known result[10] that the path-space measure does not exist for the Schrödinger equation. Our proof starts similarly as that for the Schrödinger equation. Assuming that a measure ν_t that satisfies (2) exists, we can show that

$$\sup\left\{\left|\int_{C_t} d\nu_t(X)\phi(X(0))\right|:\ \phi\in C_0^\infty(\mathbb{R}^3,\mathbb{C}^4),\|\phi\|_\infty\leqslant 1\right\} = \infty, \tag{7}$$

$\|\cdot\|_\infty$ being the supremum norm. But this is impossible, since the integral in (7) can be estimated in terms of the total variation of the measure ν_t and the supremum norm of ϕ. We shall however see that the proof that the supremum in (7) is infinite is completely different from the known proof for the Schrödinger equation. The main difference is that the solution of the Dirac (and Weyl) equation apparently involves the first derivative of the initial condition. It is the dependence of the solution on the derivative of the initial condition what gives rise to the infinity of the supremum in (7). Obviously, such a proof would not work for the Schrödinger equation. In this case the solution does not depend on the derivative of the initial condition, so the origin of the infinity in (7) and of the nonexistence of the path-space measure must be different. Essentially, for the Schrödinger equation, the infinity comes from the oscillating exponential which appears in the integral kernel of the evolution operator.

Let us note that for the Dirac equation in two-dimensional space-time, the solution does not depend on the derivative of the initial condition, either. One can write the expansion of the solution in powers of the mass. It turns out to have the same form as expansion (14) given in the next section, but formula (10) defining $W_j(t)$ must be replaced by the simple expression $[W_j(t)\phi](x) = \phi(x - jt)$. There are no derivatives of the initial condition in this expression. The supremum in (7) can easily be shown to be finite and, as we know, the corresponding path-space measure does exist. Incidentally, in the two space-time dimensional case, the expansion in powers of the mass can be used to construct the measure.[11]

Before we proceed to the next section, we would like to fix the notation we are going to use. For $z\in\mathbb{C}^n$, let $|z| = \max\times\{|z_i|:\ i = 1,2,3,...,n\}$ be the maximum norm, and let $\|\phi\|_\infty = \sup\{|\phi(x)|:\ x\in\mathbb{R}^3\}$ denote the supremum norm of a bounded function φ from $\mathbb{R}^3$ to $\mathbb{C}^n$. For a $\mathbb{C}$-valued measure ν on a σ-algebra Σ, the total variation of ν is given by the formula

$$\mathrm{var}(\nu) = \sup\left\{\sum_{i=1}^n |\nu(A_i)|:\ A_1,A_2,...,A_n \text{ are disjoint sets from } \Sigma\right\}.$$

For a $\mathbb{C}^{n\times n}$-valued measure ν, the total variation is defined as the maximum of the total variations of the n^2 entries of ν,

$$\mathrm{var}(\nu) = \max\{\mathrm{var}(\nu_{i,j}):\ i,j = 1,2,...,n\}.$$

II. THE NONEXISTENCE OF THE MEASURE

The following theorem states the nonexistence of the path-space measure for the Weyl equation.

Theorem 1: There is no such $\mathbb{C}^{2\times 2}$-valued measure ν_t on the path-space $C_t = \{X\in C([0,t],\mathbb{R}^3):\ X(t) = 0\}$ that for all smooth functions with compact support $\varphi\in C_0^\infty\times(\mathbb{R}^3,\mathbb{C}^2)$, the formula

$$\psi(t,x) = \int_{C_t} d\nu_t(X)\phi(x + X(0)) \tag{8}$$

holds, ψ being the solution to the Cauchy problem (6) for the Weyl equation with the initial condition ϕ.

Proof: We shall prove, that if there exists a $\mathbb{C}^{2\times 2}$-valued measure ν_t on the path-space C_t such that the solution to the Cauchy problem (6) is given by equality (8), then

$$\sup\left\{\left|\int_{C_t} d\nu_t(X)\phi(X(0))\right|:\ \phi\in C_0^\infty(\mathbb{R}^3,\mathbb{C}^2),\|\phi\|_\infty\leqslant 1\right\} = \infty. \tag{9}$$

On the other hand, each of the four entries of the 2×2 matrix-valued measure ν_t is a $\mathbb{C}$-valued measure, hence its total variation is bounded.[12] The estimate

$$\left|\int_{C_t} d\nu_t(X)\phi(X(0))\right| \leqslant 2\,\mathrm{var}(\nu_t)\|\phi\|_\infty$$

holds. The factor 2 comes from the multiplication of a vector in $\mathbb{C}^2$ by a matrix in $\mathbb{C}^{2\times 2}$ under the integral sign. Obviously, the estimate is in contradiction with (9).

Now, we shall prove that the supremum in (9) is infinite. For $j = \pm 1$ and $t\in\mathbb{R}$, let us introduce an operator $W_j(t)$ from $C^2(\mathbb{R}^3,\mathbb{C}^2)$ to $C^1(\mathbb{R}^3,\mathbb{C}^2)$ such that

$$[W_j(t)\phi](x) = (\partial_t - j\sigma\cdot\nabla)t\int_S \phi(x + tn)d\lambda(n), \tag{10}$$

where S is the unit sphere in $\mathbb{R}^3$, λ denotes the normal (i.e., divided by 4π) Lebesgue measure on S, and n stands for a unit vector belonging to S. If $\phi\in C^2(\mathbb{R}^3,\mathbb{C}^2)$ is the initial condition of the Cauchy problem (6) for the Weyl equation, then the function $\psi(t,x) = [W_j(t)\phi](x)$ turns out to be its solution. Indeed, it follows immediately from (10) that $\psi(0,x) = [W_j(0)\phi](x) = \phi(x)$, and by the Poisson formula,[13] the function $u:[0,\infty)\times\mathbb{R}^3\to\mathbb{C}^2$, defined by the equality

$$u(t,x) = t\int_S \phi(x + tn)d\lambda(n),$$

satisfies the Cauchy problem for the wave equation

$$(\partial_t^2 - \Delta)u(t,x) = 0,$$
$$u(0,x) = 0,$$
$$\partial_t u(0,x) = \phi(x).$$

The relation between ψ and u is $\psi(t,x) = (\partial_t - j\sigma\cdot\nabla)u(t,x)$. Since $(\partial_t + j\sigma\cdot\nabla)(\partial_t - j\sigma\cdot\nabla) = (\partial_t^2 - \Delta)$, it follows that ψ satisfies the Cauchy problem (6).

Consider a sequence of initial conditions $\phi^k: \mathbb{R}^3 \to \mathbb{C}^2$ such that for $k = 1,2,3,...$, and $x\in\mathbb{R}^3$,

$$\phi^k(x) = \exp(ik|x|^2)\zeta(|x|)v, \tag{11}$$

where $v = (1,1)^T \in \mathbb{C}^2$, and ζ: $\mathbb{R}\to\mathbb{R}$ is a function of class C^∞ such that $|\zeta(r)| \leqslant 1$ for all $r\in\mathbb{R}$, $\zeta(r) = 1$ if $|r|$ is not greater than t, and $\zeta(r) = 0$ if $|r|$ is not less than $2t$. Obviously, $\phi^k \in C_0^\infty(\mathbb{R}^3,\mathbb{C}^2)$ and $\|\phi^k\|_\infty \leqslant 1$. Moreover, if the Euclidean norm of $x\in\mathbb{R}^3$ is not greater than t, then $\nabla\phi^k(x) = 2ikx\phi^k(x)$. Using formula (10) we get for any $t\geqslant 0$, $j = \pm 1$, and $k = 1,2,3,...$,

$$[W_j(t)\phi^k](0) = (1 + 2ikt)\exp(ikt^2)v. \tag{12}$$

Let $\psi^k(t,x) = [W_j(t)\phi^k](x)$ denote the solution to Cauchy problem (6) with the initial condition ϕ^k. It follows from (12) that $|\psi^k(t,0)| \to \infty$ as $k\to\infty$ and so, if a measure ν_t satisfies formula (8), then the supremum in (9) is infinite. Theorem 1 is proved. □

We are now ready to prove the theorem on the nonexistence of the path-space measure for the Dirac equation in four space-time dimensions.

Theorem 2: There does not exist a $\mathbb{C}^{4\times 4}$-valued measure ν_t on the space of paths $C_t = \{X\in C([0,t],\mathbb{R}^3): X(t) = 0\}$ such that for any smooth function with compact support $\phi\in C_0^\infty(\mathbb{R}^3,\mathbb{C}^4)$, the formula

$$\psi(t,x) = \int_{C_t} d\nu_t(X)\phi(x + X(0)) \tag{13}$$

holds, where ψ is the solution to the Cauchy problem (3) for the Dirac equation in four space-time dimensions.

Proof: Since all representations of matrices α_j and β are unitarily equivalent,[14] it suffices to prove the theorem under the assumption that α_j and β are given by formulas (4), i.e., in the Weyl representation.

First of all let us show that for any functions ϕ_+ and ϕ_- in $C^2(\mathbb{R}^3,\mathbb{C}^2)$, the following series converges for all $t\geqslant 0$, $x\in\mathbb{R}^3$, and $j = \pm 1$:

$$\psi_j(t,x) = \sum_{n=0}^{\infty}(-im)^n \int_0^t ds_n \int_0^{s_n} ds_{n-1}\cdots\int_0^{s_2} ds_1 \times\left[W_j\left(t - 2\sum_{l=1}^{n}(-1)^{n+l}s_l\right)\phi_{j(-1)^n}\right](x) \tag{14}$$

and the functions ψ_+ and ψ_-, defined by (14), belong to $C^1([0,\infty)\times\mathbb{R}^3,\mathbb{C}^2)$, and satisfy the Cauchy problem (5) for the Dirac equation in the Weyl representation.

Let K be a compact set contained in $[0,\infty)\times\mathbb{R}^3$. Then the set $A_K = \cup\{x + [-t,t]n: (t,x)\in K, n\in S\}$ is a compact subset of $\mathbb{R}^3$. Using definition (10) of the operator $W_j(t)$, it is not difficult to prove that there exists a constant $M_K > 0$ such that

$$\sup_{\substack{(t,x)\in K\\ -t<s<t}} |[W_j(s)\phi](x)| \leqslant M_K \sup_{x\in A_K}[|\phi(x)| + |D\phi(x)|], \tag{15}$$

$$\sup_{\substack{(t,x)\in K\\ -t<s<t}} |\partial_t[W_j(s)\phi](x)| \leqslant M_K \sup_{x\in A_K}[|D\phi(x)| + |D^2\phi(x)|], \tag{16}$$

$$\sup_{\substack{(t,x)\in K\\ -t<s<t}} |\nabla[W_j(s)\phi](x)| \leqslant M_K \sup_{x\in A_K}[|D\phi(x)| + |D^2\phi(x)|], \tag{17}$$

$D\phi$ and $D^2\phi$ being the first and second differentials of ϕ, respectively. Let us observe that if $0 < s_1 < s_2 < ... < s_n < t$, then $-t < t - 2\Sigma_{l=1}^n(-1)^{n+l}s_l < t$. Hence it follows from estimate (15), that for all $(t,x)\in K$, the nth term of expansion (14) is bounded by $M_K R m^n T^n/n!$, where $R = \sup\{|\phi_j(x)| + |D\phi_j(x)|: x\in A_K, \quad j = \pm 1\}$, and $T = \sup\{t: (t,x)\in K\}$. Therefore expansion (14) converges uniformly for $(t,x)\in K$. Analogously, estimates (16) and (17) give the uniform convergence for $(t,x)\in K$ of the series of derivatives with respect to t and x of the terms of expansion (14). It follows that ψ_+ and ψ_- belong to $C^1([0,\infty)\times\mathbb{R}^3,\mathbb{C}^2)$ and for all $(t,x)\in[0,\infty)\times\mathbb{R}^3$ and $j = \pm 1$, the following series of derivatives are convergent to the derivatives of ψ_j:

$$\begin{aligned}\partial_t\psi_j(t,x) &= \sum_{n=1}^{\infty}(-im)^n\int_0^t ds_{n-1}\int_0^{s_{n-1}} ds_{n-2}\cdots\int_0^{s_2} ds_1\left[W_j\left(-t - 2\sum_{l=1}^{n-1}(-1)^{n+l}s_l\right)\phi_{j(-1)^n}\right](x)\\ &\quad + \sum_{n=0}^{\infty}(-im)^n\int_0^t ds_n\int_0^{s_n} ds_{n-1}\cdots\int_0^{s_2} ds_1\,\partial_t\left[W_j\left(t - 2\sum_{l=1}^{n}(-1)^{n+l}s_l\right)\phi_{j(-1)^n}\right](x)\\ &= -im\sum_{n=0}^{\infty}(-im)^n\int_0^t ds_n\int_0^{s_n} ds_{n-1}\cdots\int_0^{s_2} ds_1\left[W_j\left(-t + 2\sum_{l=1}^{n}(-1)^{n+l}s_l\right)\phi_{-j(-1)^n}\right](x)\\ &\quad + \sum_{n=0}^{\infty}(-im)^n\int_0^t ds_n\int_0^{s_n} ds_{n-1}\cdots\int_0^{s_2} ds_1\,\partial_t\left[W_j\left(t - 2\sum_{l=1}^{n}(-1)^{n+l}s_l\right)\phi_{j(-1)^n}\right](x),\end{aligned} \tag{18}$$

and

$$\sigma\cdot\nabla\psi_j(t,x) = \sum_{n=0}^{\infty}(-im)^n\int_0^t ds_n\int_0^{s_n} ds_{n-1}\cdots\int_0^{s_2} ds_1\,\sigma\cdot\nabla\left[W_j\left(t - 2\sum_{l=1}^{n}(-1)^{n+l}s_l\right)\phi_{j(-1)^n}\right](x). \tag{19}$$

From the proof of Theorem 1 we have $(\partial_t + j\sigma\cdot\nabla)[W_j(t+s)\phi_{\pm j}] = 0$. From definition (10), we can deduce easily that $W_j(-t) = W_{-j}(t)$. Thus we get from (14), (15), and (16) that indeed the functions ψ_+ and ψ_- solve the Cauchy problem (5) for the Dirac equation in the Weyl representation.

Next, we consider a sequence of initial conditions $\phi_j^k \in C_0^\infty(\mathbb{R}^3,\mathbb{C}^2)$, where $j = \pm 1$, $k = 1,2,3,...$, and $\phi_+^k = \phi_-^k = \phi^k$, ϕ^k being the function defined by formula (11). Let $\psi^k = (\psi_+^k, \psi_-^k)$ be the solution to Cauchy problem (5) for the Dirac equation with the initial condition $\phi^k = (\phi_+^k, \phi_-^k)$. From (14) and the uniqueness of the solution to the Dirac equation, we get

$$\psi_j^k(t,0) = \sum_{n=0}^{\infty} (-im)^n \int_0^t ds_n \int_0^{s_n} ds_{n-1} \cdots \int_0^{s_2} ds_1 \times \left[W_j\left(t - 2\sum_{l=1}^{n} (-1)^{n+l} s_l\right) \phi_{j(-1)^n}^k \right](0). \tag{20}$$

Let us observe that for $n \geqslant 1$, the nth term of expansion (20) is bounded by $(t+1)t^{n-1}m^n/(n-1)!$. Indeed, setting $u = t - 2\Sigma_{l=1}^n (-1)^{n+l}s_l$ and using formula (12), we have

$$\begin{aligned}
&\left| (-im)^n \int_0^t ds_n \int_0^{s_n} ds_{n-1} \cdots \int_0^{s_2} ds_1 \right. \\
&\quad \left. \times \left[W_j\left(t - 2\sum_{l=1}^{n} (-1)^{n+l} s_l\right) \phi_{j(-1)^n}^k \right](0) \right| \\
&\quad \leqslant \frac{m^n}{2} \int_0^t ds_n \int_0^{s_n} ds_{n-1} \cdots \int_0^{s_3} ds_2 \left| \int_a^b du(1+2imu) \right. \\
&\quad \left. \times \exp(imu^2) \right| \\
&\quad \leqslant m(t+1)\frac{(mt)^{n-1}}{(n-1)!}.
\end{aligned} \tag{21}$$

We put here $a = t - 2\Sigma_{l=2}^n (-1)^{n+l}s_l$ and $b = t - 2\Sigma_{l=3}^n (-1)^{n+l}s_l$ for simplicity. The last inequality in (21) holds, since $|a| \leqslant t$, $|b| \leqslant t$, and

$$\begin{aligned}
&\left| \int_a^b du(1+2imu)\exp(imu^2) \right| \\
&\quad \leqslant \left| \int_a^b du \exp(imu^2) + \int_a^b du \frac{d\exp(imu^2)}{du} \right| \\
&\quad \leqslant |a| + |b| + 2 \leqslant 2(t+1).
\end{aligned}$$

It follows that the sum over $n \geqslant 1$ in expansion (20) is for all $k = 1,2,3,...$ bounded by $m(t+1)e^{mt}$. By formula (12), the leading term (for $n = 0$) in expansion (20) is $[W_j(t)\phi_j^k](0) = (1+2ikt)\exp(ikt^2)v$, and so it tends in norm to infinity as $k \to \infty$. Therefore $|\psi^k(t,0)| \to \infty$ as $k \to \infty$. The same argument as for the Weyl equation (see the proof of Theorem 1) shows that this is in contradiction with the existence of a path-space measure ν_t satisfying formula (13). The proof is completed. □

Remark: Let us note that in the proofs of Theorems 1 and 2, we used only values of the solutions $\psi^k(t,\cdot)$ calculated at one fixed point $x = 0$. We are allowed to do that, since all the functions $\psi^k(t,\cdot)$ are continuous (in fact they are smooth functions with compact supports).

III. CONCLUDING REMARKS

Despite the fact that for the Dirac equation in two space-time dimensions the path integral can be understood as an integral with respect to a well-defined path-space measure, it has been proved in Sec. II that this is no longer true in four space-time dimensions. This motivates the need for a definition of a relativistic path integral which is not based on the measure theory. The analogy with the Schrödinger equation, for which the path-space measure does not exist either, suggests that one could apply methods developed for the Schrödinger equation. However, since there are different reasons why the path-space measure does not exist for the Dirac and Schrödinger equations, the analogy should be expected to be very limited.

As an example let us take the method of oscillatory integrals, which gives a very useful definition of the path integral for the Schrödinger equation.[15] Since this method is based on the specific form of the integral kernel of the evolution operator for the Schrödinger equation involving oscillating exponentials, it is unlikely that it would work for the Dirac equation.

The path integral for the Schrödinger equation can also be defined by means of the analytic continuation of the path integral for the diffusion equation, i.e., the integral with respect to the Wiener measure. A relativistic analogy of this method was suggested by Kac *et al.*,[16] who observed that the path integral for the Dirac equation in two space-time dimensions can be regarded as the analytic continuation in mass of a path integral for the telegrapher's equation. Let us note however, that the proof of Theorem 2 still works if the non-negative mass parameter m is replaced by any complex number z, which means that in four space-time dimensions there is no relativistic analogy (in the sense of Kac *et al.*[16]) of the Wiener measure. Therefore it seems impossible in four space-time dimensions to use the analytic continuation to define the relativistic path integral along the same lines as Kac *et al.*[16] did in the two-dimensional case. Naturally, this does not exclude the possibility that a definition based on another scheme of analytic continuation may be available.

Last but not least, there is the method of approximation of the path integral by integrals over finite-dimensional spaces of polygonal paths. In connection with this method, we would like to mention two different schemes of approximation of the solution to the Dirac equation in four dimensions: the first one attributable to Suarez,[17] and the other one by Jacobson.[18] The idea of the latter one is closer to Feynman's original concepts.[19] Both of the schemes involve polygonal paths whose speeds exceed the speed of light. This is a common feature of the approximation schemes for the Dirac equation, but its physical interpretation is rather unclear.

ACKNOWLEDGMENTS

It is a pleasure to thank for the warm hospitality the Mathematics Institute, The University of Warwick, and St. Johns's College, Oxford, where this paper was written. I would like to express my deep gratitude to Professor K. D. Elworthy for generous help and for friendly discussions from which I have greatly benefited. I wish also to thank Professor A. Pliś for his continuous interest in my work.

[1]T. Ichinose, Duke Math. J. **51**, 1 (1984); T. Ichinose and H. Tamura, J. Math. Phys. **25**, 1810 (1984); T. Ichinose, Physica A **124**, 419 (1984); T. Ichinose, Proc. Jpn. Acad. A **58**, 290 (1982); T. Ichinose and H. Tamura, Prog. Theor. Phys. Suppl. **92**, 144 (1987); J. Math. Phys. **29**, 103 (1988).
[2]P. Blanchard, P. Combe, M. Sirugue, and M. Sirugue-Collin, "Jump processes related to the two-dimensional Dirac equation," Research Center Bielefeld–Bochum–Stochastic (BiBos), preprint no. 117 (1985); "Probabilistic solution of the Dirac equation in presence of an external electromagnetic field," Research Center Bielefeld–Bochum–Stochastic (BiBos), preprint no. 44 (1984); "Path integral representation for the solution of the Dirac equation in presence of an electromagnetic field," Research Center Bielefeld–Bochum–Stochastic (BiBoS), preprint no. 66 (1984); "The Zitterbevegung of a Dirac electron in a central field," Research Center Bielefeld–Bochum–Stochastic, (BiBoS), preprint no. 199 (1985).
[3]T. Zastawniak, "Path integrals and Probabilistic Representations for the Dirac Equation in Two and Four Space-Time Dimensions," Ph.D. thesis, Jagiellonian University, Kraków, 1987; to appear in Bull. Pol. Acad. Sci. Math. **37**, No. 7–8 (1989).
[4]See last source in Ref. 1 and both sources in Ref. 3.
[5]R. P. Feynman and A. R. Hibbs, *Quantum Mechanics and Path Integrals* (McGraw-Hill, New York, 1965), Problem 2-6, p. 34, Remark 2-6, p. 38; R. P. Feynman, Phys. Rev. **80**, 440 (1950); G. V. Riazanov, Zh. Eksp. Teor. Fiz. **33**, 1437 (1958) [Sov. Phys. JETP 6, 1107 (1958)]; G. Rosen, *Formulations of Classical and Quantum Dynamical Theory* (Academic, New York and London, 1969), Appendix E, p. 118.
[6]For details and proofs see Refs. 1–3.
[7]See, e.g., B. Simon, *Functional Integration and Quantum Physics* (Academic, New York and London, 1979), Chap. V, Formulas (15.1) and (15.2).
[8]See first source in Ref. 1, Section 4, Remark 4; see also B. Gaveau, J. Funct. Anal. **58**, 314 (1984), where the lack of the L^∞ bound is avoided by working in momentum space.
[9]Ichinose as well as Blanchard *et al.* obtained a path-integral representation for a relativistic particle in four space-time dimensions in the presence of an extenal central field by separating the radial and angular variables. The representation, however, involves path integrals in two space-time dimensions only; see second source in Ref. 1 and last source in Ref. 2.
[10]See M. Reed and B. Simon, *Methods of Modern Mathematical Physics, Vol. II: Analysis, Self-Adjointness* (Academic, New York, 1975), Chap. X, Problem 64, and Notes to Sec. X.11.
[11]See Ref. 3.
[12]See, e.g., N. Dunford and J. T. Schwartz, *Linear Operators, Vol. I: General Theory* (Interscience, New York, 1958), Chapter III; A. J.Weir, *Integration and Measure, Vol. II: General Integration and Measure* (Cambridge U.P., London, 1974), Chap. 16.
[13]See, e.g., L. Bers, F. John, and M. Schechter, *Partial Differential Equations* (Interscience, New York, 1964), Pt. I, Chap. 2, Sec. 2.
[14]See, e.g., A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1963), Vol. II, Chap. XX, § 10.
[15]See, e.g., S. Albeverio and R. Høegh-Krohn, *Mathematical Theory of Feynman Path Integrals, Lecture Notes in Mathematics*, Vol. 523 (Springer, Berlin, 1976).
[16]R. Gaveau, T. Jacobson, M. Kac, and L. S. Schulman, Phys. Rev. Lett. **53**, 419 (1984).
[17]R. Suarez, J. Phys. A **17**, 1003 (1984).
[18]T. Jacobson, J. Phys. A **17**, 2433 (1984).
[19]See second source in Ref. 5.

Stochastic field theory at finite temperature and equilibrium states in quantum statistical mechanics

Pratul Bandyopadhyay
Indian Statistical Institute, Calcutta 700035, India

Brahmananda Bhattacharya
Fakir chand College, Diamond Harbour, W.B., India

Pradip Ghosh
Maharaja Manindra Chandra College, Calcutta 700002, India

(Received 5 February 1988; accepted for publication 21 December 1988)

Stochastic field theory at finite temperature, when it is formulated from stochastic mechanics, which incorporates Brownian motion in both the external, as well as the internal, space of a particle, and the external and internal fields, and which are written as thermal doublets, gives rise to finite temperature quantum field theory in Minkowski space. The Haag–Hugenholtz–Winnik, (HHW) formalism of modular conjugation and the Tomita–Takesaki theory of modular Hilbert algebra then follow from this finite temperature quantum field theoretical formalism. It is shown that there is an inherent discrete Z_2 symmetry among the fields of the doublet, and the equilibrium condition is ensured by this symmetry. When this symmetry is spontaneously broken, thermodynamical equilibrium is destroyed.

I. INTRODUCTION

In the previous paper,[1] we have tried to show that in stochastic field theory at finite temperature, when it is constructed from stochastic mechanics where it is taken that there are universal Brownian motions in the external space as well as in the internal space of the particles, the field functions associated with the internal space variables may be related with the modular conjugation operation of the HHW formalism as well as the Tomita–Takesaki theory of modular Hilbert algebras. This helps us to study thermofield dynamics as formulated by Takahashi and Umezawa,[2] where a new tilde operation is introduced in terms of the stochastic fields associated with the internal variables. In fact, in Ref. 1, it has been shown that thermofield dynamics can be reformulated in terms of stochastic field theory at finite temperature, and the real-time Green's functions can be derived when we write the field functions associated with the external space-time and internal space-time as thermal doublets.

Recently Niemi and Semenoff[3] have formulated finite temperature quantum field theories in Minkowski space (real time) using Feynman path integrals. It has been shown that at nonzero temperature, a new field arises which plays the role of a ghost field, and a new discrete Z_2 symmetry arises. This thermal Z_2 symmetry actually relates the time-reversal invariance of the two fields—the physical field and its conjugate ghost field. Thermodynamical equilibrium is destroyed when this Z_2 symmetry is broken.

In this paper, we shall show that the Niemi–Semenoff formalism is equivalent to thermofield dynamics when we interpret the latter with the stochastic field theory at finite temperature involving fields in the external and internal space, and the thermal Z_2 symmetry for the equilibrium condition is essentially related with the tilde operation associating the stochastic external and internal fields. In this sense, Z_2 symmetry introduced by Niemi and Semenoff is found to be isomorphic with the modular conjugation operation of Haag, Hugenholtz, and Winnik, and it finds its relevance in the Tomita–Takesaki theory of modular Hilbert algebras. In this way the real-time formalism of finite temperature quantum field theory is found to be associated with the equilibrium conditions in quantum statistical mechanics.

It may be recalled here that when the stochastic nonlocal field is written as a product function of the external and internal fields, the correlation functions we obtain involve imaginary time. The correlation function involving only external fields can be obtained when we set that the internal time variable vanishes and the internal space variables are integrated out. Thus a conventional Euclidean field theory can be constructed from stochastic fields. However, as pointed out by Guerra and Ruggiero,[4] this Euclidean field theoretic formalism has a real-time interpretation in the sense that a stochastic field can be generated from stochastic oscillators in real time, as is considered in stochastic mechanics.[5] In Ref. 1, we have shown that this inherent real-time formalism can be explicitly recovered when we write the external and internal stochastic fields as thermal doublets. Indeed it is essential also at finite temperature to write them as doublets, as the temperature effect on the external and internal fields may be different. In Ref. 1, it has been shown that an anisotropic feature in the internal variable gives rise to a fermion. So it may happen that at finite temperature this anisotropic feature is destroyed, which means that the thermal doublet will have different statistics. Thus we find that this thermal doublet formalism for the external and internal fields is a necessary requirement to study thermal effect and equilibrium condition. This helps us to relate the stochastic field theoretic formalism with thermofield dynamics of Takahashi and Umezawa, as well as with the Niemi–Semenoff formalism of Z_2 symmetry, and a unified picture emerges for finite temperature quantum field theory.

In Sec. II, we shall recapitulate certain results as derived

 0022-2488/89/061359-06$02.50

in Ref. 1 to show the association of stochastic fields with thermofield dynamics. In Sec. III, we shall study the equilibrium condition in terms of this stochastic field theory at finite temperature, and we shall point out its association with the Haag–Hugenholtz–Winnik (HHW) formalism of modular conjugation and the Tomita–Takesaki theory of modular Hilbert algebras. In Sec. IV, we shall show the equivalence of this equilibrium condition with the Niemi–Semenoff formalism of Z_2 symmetry.

II. STOCHASTIC FIELD THEORY AT FINITE TEMPERATURE AND THERMOFIELD DYNAMICS

We have recapitulated certain results of Ref. 1 for completeness. Nelson[5] developed a quantization procedure which is based on Brownian motion processes evolving in real time. But the major difficulty in Nelson's procedure is that we do not know how to have a relativistic generalization of this and achieve the quantization of a Fermi field. In a recent paper,[6] it has been shown that in Nelson's formalism, the relativistic generalization as well as the quantization of a Fermi field can be achieved when an anisotropy in the internal space of a particle is introduced and it is taken into account that there are Brownian motions in both the external and internal space. In view of this, we denote the configuration variables as $Q(t,\xi_0)$, where ξ_0 is the fourth component (real) of the internal four-vector ξ_μ, which is considered to be the attached vector to the space-time point x_μ. We assume that $Q(t,\xi_0)$ is a separable function and can be denoted as

$$Q(t,\xi_0) = q(t)q(\xi_0). \tag{1}$$

The process $Q(t,\xi_0)$ is assumed to satisfy the stochastic differential equations

$$dQ_i(t,\xi_0) = b_i(Q(t,\xi_0),t,\xi_0)dt + d\omega_i(t), \tag{2}$$

$$dQ_i(t,\xi_0) = b'_i(Q(t,\xi_0),t,\xi_0)d\xi_0 + d\omega_i(\xi_0), \tag{3}$$

where $b_i(Q(t,\xi_0),t,\xi_0)$ and $b'_i(Q(t,\xi_0),t,\xi_0)$ correspond to certain velocity fields and $d\omega_i$ are independent Brownian motions. It is assumed that $d\omega_i(t)$ $[d\omega_i(\xi_0)]$ does not depend on $Q(S,S')$ for $S\leqslant t(S'\leqslant\xi_0)$, and the expectations have the following values:

$$\begin{aligned}
&\langle d\omega_i(t)\rangle = 0,\\
&\langle d\omega_i(t)d\omega_j(t')\rangle\\
&\quad = (\hbar/m)\delta_{ij}\delta(t-t')dt\,dt',\\
&\langle d\omega_i(\xi_0)\rangle = 0,\\
&\quad \langle d\omega_i(\xi_0)d\omega_j(\xi'_0)\rangle\\
&\quad = (\hbar/\pi^0)\delta_{ij}\delta(\xi_0-\xi'_0)d\xi_0\,d\xi'_0,
\end{aligned} \tag{4}$$

where $\hbar$ is Planck's constant divided by 2π and π^0 is a suitable constant having the dimension of m. The description is asymmetrical in both "external" and "internal" time but we can also write

$$dQ_i(t,\xi_0) = b^*_i(Q(t,\xi_0),t,\xi_0)dt + d\omega^*_i(t), \tag{5}$$

$$dQ_i(t,\xi_0) = b'^*_i(Q(t,\xi_0),t,\xi_0)d\xi_0 + d\omega^*_i(\xi_0), \tag{6}$$

where ω^* has the same properties as ω except that $d\omega^*_i(t)$ $[d\omega^*_i(\xi_0)]$ are independent of $Q(S,S')$ with $S\geqslant t$, $S'\geqslant\xi_0$.

From the stochastic equations considered here, the following moments can be derived.

$$\begin{aligned}
&\langle Q_i(t,\xi_0)\rangle = 0,\\
&\langle Q_i(t,\xi_0)Q_j(t',\xi'_0)\rangle\\
&\quad = (\hbar/2m\omega)(\hbar/2\pi^0\omega')\delta_{ij}e^{-\omega(t-t')}e^{-\omega'(\xi_0-\xi'_0)}\\
&\qquad (t>t',\xi_0>\xi'_0).
\end{aligned} \tag{7}$$

Indeed this follows from the fact that $Q_i(t,\xi_0)$ can be written in a separable way $q_i(t)q_i(\xi_0)$ as we have assumed earlier, and so we can utilize the results for the moments of the q_i derived by Abers and Lee[7] as well as by Moore[8]

$$\begin{aligned}
&\langle q_i(t)\rangle = 0,\\
&\langle q_i(t)q_j(t')\rangle = (\hbar/2m\omega)\delta_{ij}e^{-\omega(t-t')} \quad (t>t').
\end{aligned} \tag{8}$$

These results can be extended to the variables $q_i(\xi_0)$ in an analogous way:

$$\begin{aligned}
&\langle q_i(\xi_0)\rangle = 0,\\
&\langle q_i(\xi_0)q_j(\xi'_0)\rangle = (\hbar/2\pi^0\omega')\delta_{ij}e^{-\omega'(\xi_0-\xi'_0)}\\
&\qquad (\xi_0>\xi'_0).
\end{aligned} \tag{9}$$

Let $\{e_i(\vec{x})\}$ denote the complete orthonormal set of eigenfunctions of the three-dimensional Laplacian $-\Delta$,

$$\Delta e_i(\vec{x}) = -k_i^2 e_i(\vec{x}). \tag{10}$$

Also we denote $\{e_j(\vec{\xi})\}$ as the complete orthonormal set of eigenfunctions of the three-dimensional Laplacian $-\Delta'$ in terms of the variables ξ_i, so that

$$\Delta' e_j(\vec{\xi}) = -\vec{\pi}_j^2 e_j(\vec{\xi}).$$

Now we can construct a stochastic field which can be expressed as an orthonormal expansion in terms of $q_i(t)$, $e_i(\vec{x})$, $q_j(\xi_0)$, $e_j(\vec{\xi})$,

$$\phi(x,t,\xi) = \sum_{i,j} q_i(t)e_i(\vec{x})q_j(\xi_0)e_j(\vec{\xi}). \tag{11}$$

Now from the moments of $q_i(t)$, $q_j(\xi_0)$, we can determine the moments of $\phi(x,t,\xi)$,

$$\begin{aligned}
&\langle\phi(x,t,\xi)\rangle = 0,\\
&\langle\phi(x,t,\xi)\phi(x',t',\xi')\rangle\\
&\quad = \frac{1}{(2\pi)^3}\int d^3\vec{k}\,e^{i\vec{k}(\vec{x}-\vec{x}')}g(t-t')\\
&\qquad \times\frac{1}{(2\pi)^3}\int d^3\vec{\pi}\,e^{i\vec{\pi}(\vec{\xi}-\vec{\xi}')}g(\xi_0-\xi'_0)\\
&\quad = \frac{1}{(2\pi)^4}\int\frac{d^4k\,e^{i\{k,(x-x')\}}}{(k,k)+m^2}\cdot\frac{1}{(2\pi)^4}\int\frac{d^4\pi\,e^{i\{\pi,(\xi-\xi')\}}}{(\pi,\pi)+\pi^{02}},
\end{aligned} \tag{12}$$

where $g(t-t')$ and $g(\xi_0-\xi'_0)$ are given by Eqs. (8) and (9), (A,B) denotes an Euclidean product and the units have been chosen to be $\hbar = m = \pi^0 = 1$.

It is noted that in the limit $\xi_0 = \xi'_0 = 0$, integrating over the internal space variable $\vec{\xi}$, the correlation function just reduces to that of the scalar field

$$\langle\phi(x,t)\delta(x',t')\rangle = \frac{1}{(2\pi)^4}\int\frac{d^4k\,e^{i\{k,(x-x')\}}}{(k,k)+m^2}. \tag{13}$$

This is the Euclidean Markov field result which has been obtained from Nelson's real-time formalism of Brownian motion, and in this sense it gives rise to the equivalence of these two formalisms as advocated by Guerra and Ruggiero.[4] Again it has been shown that when an anisotropic feature of the internal space-time corresponding to the variable ξ_μ is taken into account, implicitly so that the two opposite helicities give rise to the fermion and the antifermion, we can obtain the fermionic propagator in Euclidean space-time.[6]

Now introducing temperature dependent moments[8]

$$\langle d\omega_i(t)d\omega_j(t')\rangle_{T\neq 0} = \frac{1}{\beta m}\delta_{ij}\sum_{n=-\alpha}^{\alpha} e^{i\omega_n(t-t')},$$

$$\langle d\omega_i(\xi_0)d\omega_j(\xi_0')\rangle_{T\neq 0} = \frac{1}{\beta\pi^0}\delta_{ij}\sum_{n=-\alpha}^{\alpha} e^{i\omega_n(\xi_0-\xi_0')}, \tag{14}$$

with $\omega_n = 2\pi n/\beta\hbar$, which has been derived from the consideration of the KMS condition, we can obtain the moments

$$\langle q(t)q(t')\rangle_{T\neq 0} = \frac{1}{\beta m}\sum_{n=-\alpha}^{\alpha}\frac{\exp[i\omega_n(t-t')]}{\omega^2+\omega_n^2}, \tag{15}$$

$$\langle q(\xi_0)q(\xi_0')\rangle_{T\neq 0} = \frac{1}{\beta\pi^0}\sum_{n=-\alpha}^{\alpha}\frac{\exp[i\omega_n(\xi_0-\xi_0')]}{\omega^2+\omega_n^2}. \tag{16}$$

These lead us to the correlation function of the stochastic fields for a particular mode $n=1$ when we take $\xi_0=\xi_0'=0$ and integrate over the space variable $\vec{\xi}$,

$$\langle q(x,t)\phi(x',t')\rangle_{T\neq 0} = \frac{1}{(2\pi)^4}\int\frac{d^4k\,e^{i(k,(x-x'))}}{(k,k)+m^2}2\pi\delta(\omega_1-k_0), \tag{17}$$

using the normalization condition $\hbar = m = \pi^0 = \beta = 1$ (see Ref. 1).

When an anisotropy is introduced in the internal space, so that the two opposite orientations give rise to the fermion and antifermion, we get the correlation function of the spinor field for $n=1$ (see Ref. 1),

$$\langle\bar{\phi}(x,t)\bar{\phi}(x',t')\rangle_{T\neq 0} = \frac{1}{(2\pi)^4}\int\frac{d^4k\,e^{i(k,(x-x'))}}{\gamma_\mu k_\mu + m}2\pi\delta(\omega_1-k_0). \tag{18}$$

As it is noted here, the statistics of the particle depend on the internal space-time variable ξ. Indeed the internal helicity arising from the anisotropic feature of the internal space corresponds to the fermion number. Again, when there is no anisotropy in the internal space, we get a boson. Now temperature should definitely affect the internal motion, and as such it may happen that at high temperature the anisotropic feature of the internal space will be destroyed and the fermion may be transformed into a boson. That is, a massive extended body depicting a fermion can have such a phase transition. However, this does not mean that fermion number conservation will be violated, as Lorentz invariance does not allow such a process to occur.[9] The only effect of such a phase transition will be that a thermal pair of opposite statistics will emerge as zero energy modes at the critical temperature leading to a nonequilibrium state. In view of this, at finite temperature we should write the stochastic nonlocal field $\phi(x,\xi)$, which is assumed to satisfy the condition of separability $\phi(x,\xi)=\phi(x)\phi(\xi)$ as a thermal doublet $\binom{\phi(x)}{\phi(\xi)}$. It may be noted that though x and ξ represent two different spaces, the external motion is a manifestation of the internal motion, and as such a mapping of x and ξ is possible. In that case x may be represented in the functional form $x(\xi)$ and the simplest form of mapping can be taken as $x=c\xi$, where c is a suitable constant. In view of this mapping, there should be a mapping of $\phi(x)$ and $\phi(\xi)$ also. We can assume that $\phi(\xi)=\tilde{\lambda}\phi^+(x)=\tilde{\phi}^+(x)$, where λ is a suitable parameter. Thus the thermal doublet $\binom{\phi(x)}{\phi(\xi)}$ can be written as $\binom{\phi(x)}{\tilde{\phi}^+(x)}$. So we can consider that there exists a conjugate Hilbert space $\tilde{H}$ associated with the Hilbert space H such that $\tilde{H}$ is the set H with scalar multiplication $\lambda,\xi\to\bar{\lambda}\xi$, where $\lambda\in C$ and $\xi\in H$, and with scalar product $\xi,\eta\to(\xi,\eta)$ with $\xi,\eta\in H$, where $\xi,\eta\to(\xi,\eta)$ is the scalar product. In effect H is the Hilbert space associated with the external space, and $\tilde{H}$ is the conjugate Hilbert space associated with the internal space.

In Ref. 1 it has been shown that this helps us to reformulate thermofield dynamics as proposed by Takahashi and Umezawa in terms of stochastic fields. For example, in case of bosons, we can write

$$\phi(x)=\begin{pmatrix}\phi(x)\\ \tilde{\phi}^+(x)\end{pmatrix}=\int_x\frac{d^4p}{(2\pi)^3}\theta(p_0)\delta(p^2-m^2)\left\{\begin{pmatrix}a_+(\vec{p})\\ \tilde{a}_+ +(\vec{p})\end{pmatrix}e^{-ipx}+\begin{pmatrix}a_-^+(\vec{p})\\ \tilde{a}_-(\vec{p})\end{pmatrix}e^{ipx}\right\} \tag{19}$$

and

$$\langle T\phi(x)\phi^+(y)\rangle = \langle 0(\beta)\left|T\left[\begin{pmatrix}\phi(x)\\ \tilde{\phi}^+(x)\end{pmatrix}(\phi^+(y),\tilde{\phi}(y))\right]\right|0(B)\rangle$$
$$= i\int\frac{d^4p}{(2\pi)^4}e^{-ip(x-y)}V_B(|\vec{p}|,\beta)\begin{pmatrix}(p^2-m^2+i0)^{-1} & 0\\ 0 & (-p^2-m^2-i0)^{-1}\end{pmatrix}V_B^+ \tag{20}$$

where V_B are the coefficients of Bogoliubov transformations

$$V_B(|\vec{p}|,\beta)=\begin{pmatrix}\cosh\theta(|\vec{p}|,\beta) & \sinh\theta(|\vec{p}|,\beta)\\ \sinh\theta(|\vec{p}|,\beta) & \cosh\theta(|\vec{p}|,\beta)\end{pmatrix}$$
$$=\begin{pmatrix}(\sqrt{1-e^{-\beta\epsilon(\vec{p})}})^{-1} & e^{-\beta\epsilon(\vec{p})/2}/\sqrt{1-e^{-\beta\epsilon(\vec{p})}}\\ e^{-\beta\epsilon(\vec{p})/2}/\sqrt{1-e^{-\beta\epsilon(\vec{p})}} & (\sqrt{1-e^{-\beta\epsilon(\vec{p})}})^{-1}\end{pmatrix} \tag{21}$$

with

$$\epsilon(\vec{p})=\sqrt{\vec{p}^2+m^2}.$$

The "total" Lagrangian becomes

$$\begin{aligned}\bar{L}_\phi = L_\phi - \tilde{L}_\phi &= \partial_\mu\phi^+\,\partial^\mu\phi - m^2\phi^+\phi\\ &\quad - \partial_\mu\tilde{\phi}^+\,\partial^\mu\tilde{\phi} + m^2\tilde{\phi}^+\tilde{\phi}\\ &= \partial_\mu\phi^+\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\partial_\mu\phi\\ &\quad - m^2\phi^+\begin{pmatrix}1 & 0\\ 0 & -1\end{pmatrix}\phi.\end{aligned} \tag{22}$$

These relations can easily be generalized to fermionic cases also.[1] From these relations we find that we can define a temperature dependent vacuum

$$|0(\beta)\rangle=\sqrt{1-e^{-\beta\epsilon}}\,\exp(e^{-\beta\epsilon/2}a^+\tilde{a}^+)\,|0\rangle,$$

where

$$\begin{aligned}a|0(\beta)\rangle &= e^{-\beta\epsilon/2}\tilde{a}^+|0(\beta)\rangle,\\ \tilde{a}|0(\beta)\rangle &= e^{-\beta\epsilon/2}a^+|0(\beta)\rangle.\end{aligned} \tag{23}$$

The thermal average of an operator A can now be written as a vacuum expectation value

$$\langle A\rangle=\mathrm{Tr}\,e^{-\beta H}A/\mathrm{Tr}\,e^{-\beta H}=\langle 0(\beta)|A\,|0(\beta)\rangle. \tag{24}$$

This is the basis of thermofield dynamics, and it is found here to be related with stochastic fields involving external and internal space.

III. EQUILIBRIUM STATE IN QUANTUM STATISTICAL MECHANICS AND STOCHASTIC FIELD THEORY AT FINITE TEMPERATURE

The correlation functions derived for the stochastic fields at finite temperature (17) and (18) have been derived from those of stochastic oscillators as given in Eq. (14). These relations have been assumed taking into consideration the KMS condition. So the KMS condition is built into this mechanism, which ensures the equilibrium state. However, it should be noted that Eq. (20) suggests that apart from in the external space, we should have the KMS condition in the internal space also. This is a major consequence of this formalism. This implies that when we write the internal field function as a tilde function in the external space through the mapping $\phi(\xi)=\lambda\phi^+(x)=\tilde{\phi}^+(x)$, the KMS condition should have its specific form in the case of tilde fields also. That is, if the state ω satisfies the KMS condition with respect to α_t, then the extension $\tilde{\omega}$ of ω satisfies the KMS condition with respect to $\tilde{\alpha}_t$, which is the canonical extension of α_t. This makes the connection between statistical mechanics and the Tomita–Takesaki theory of modular Hilbert algebras and the Haag–Hugenholtz–Winnik formalism get a proper physical meaning.

To depict a thermodynamic equilibrium state, a Gibbs state is given by

$$\omega(A)=\mathrm{Tr}\,e^{-\beta H}A/\mathrm{Tr}\,e^{-\beta H}=\langle A\rangle. \tag{25}$$

The KMS condition states that for any A and B belonging to the operator algebra U (C^* algebra) describing the system considered, there exists a function $F_{AB}(z)$ of a complex number z, holomorphic in $0<Emz<\beta$, continuous in $0\leqslant Emz\leqslant\beta$, and satisfying the relations

$$\begin{aligned}&F_{AB}(t)=\omega(A\alpha_t(B)),\\ &F_{AB}(t+i\beta)=\omega(\alpha_t(B)A),\quad \text{for } t\in R,\end{aligned} \tag{26}$$

where

$$\alpha_t(B)=e^{iHt}Be^{-iHt}.$$

Also we have the condition that

$$\omega(A^+A)=0,\quad \text{for } A\in U \text{ implies } A=0. \tag{27}$$

Equation (26) is a consequence of the invariance of the trace operation under cyclic permutations of the operator product. Equation (27) implies that if we consider the GNS representation

$$\omega(A)=(\Omega_\omega,\pi_\omega(A)\Omega_\omega), \tag{28}$$

the cyclic vector Ω_ω is also a separating vector.

In Ref. 1, we have shown in detail how the tilde operation of stochastic fields corresponding to the stochastic fields of the internal space gives rise to the HHW modular conjugation algebra and the Tomita–Takesaki theory of modular Hilbert algebra. The main reasoning behind this is that in this thermal doublet formalism of stochastic fields with the tilde mapping requires that the "total" Lagrangian should be written as

$$\bar{L}=L-\tilde{L}, \tag{29}$$

as has been shown in Eq. (22). This implies that the "total" Hamiltonian H is given by

$$\bar{H}=H-\tilde{H}, \tag{30}$$

where the temperature dependent vacuum satisfies

$$\bar{H}\,|0(\beta)\rangle=(H-\tilde{H})|0(\beta)\rangle=0. \tag{31}$$

Moreover, we can construct and involution operator J satisfying

$$\tilde{H}=JHJ, \tag{32}$$

with $J^2=1$. This operator J relates in effect the transformation of the external and internal fields through the relations

$$\begin{aligned}\tilde{a}&=JaJ,\\ \tilde{a}^+&=Ja^+J.\end{aligned} \tag{33}$$

Indeed as long as the problem of unboundedness of the operators does not cause serious trouble, we can have the HHW formalism of modular conjugation and the Tomita–Take-

saki theory of modular Hilbert algebras. These aspects have been discussed in detail in Ref. 1.

Thus we find that we can have a field theoretic origin of the HHW algebra and the Tomita–Takesaki formalism when we take into account the thermal behavior of stochastic fields involving the external and internal space written as thermal doublets to depict the thermal aspects of the quantum statistical mechanics. The KMS condition is ensured by the correlation functions of the stochastic fluctuations, and as such stochastic field theory at finite temperture, in effect, represents thermodynamic equilibrium states.

IV. EQUILIBRIUM CONDITION AND NIEMI–SEMENOFF Z_2 SYMMETRY

Niemi and Semenoff[3] have formulated a finite temperature quantum field theory in Minkowski space which introduces a new ghost field, and the equilibrium condition is found to be governed by Z_2 symmetry between the physical field and the ghost field. This Z_2 symmetry effectively corresponds to time reversal symmetry. This arises from the extension of the time integration from the real axis segment $[-T,T]$ to an integration over the contour in the complex time plane. This contour starts at $-T$ and runs along the real-time axis to $+T$. From $+T$ the contour continues along the imaginary time direction to $+T-i\beta/2$, from $+T-i\beta/2$ parallel to the real-time axis to $-T-i\beta/2$, and finally parallel to the imaginary time axis to $-T-i\beta$. The generating functional can then be used to compute Green's functions, with arguments on the real T axis, and the contribution related to the arguments on how the imaginary T axis can be absorbed into the normalization. This leads to the result that the propagator acquires a matrix structure. There are two kinds of fields ϕ_1 and ϕ_2, where ϕ_1 appears on external lines and represents the physical field, and ϕ_2 appears as a new field corresponding to the time-reversal field of ϕ_1 and thus represents a ghost field. Evidently ϕ_1 and ϕ_2 are linked by the discrete Z_2 symmetry.

Now when we recast this result into our formalism of stochastic field theory, we note that we can associate the tilde field in the thermal doublet as the ghost field ϕ_2, and as such it represents the stochastic field in the internal space. It can be shown that this field in the internal space may be taken effectively to represent the time-reversal partner of the field in the external space. Indeed, from the stochastic differential equations (2) and (5), we can write the current velocity

$$V=\tfrac{1}{2}(b+b^*), \tag{34}$$

and the osmotic velocity

$$U=\tfrac{1}{2}(b-b^*). \tag{35}$$

Similarly, from the equations corresponding to the internal space variables (3) and (6), we can write the internal current velocity

$$v^1=\tfrac{1}{2}(b'+b'^*), \tag{36}$$

and the internal osmotic velocity

$$U'=\tfrac{1}{2}(b'-b'^*). \tag{37}$$

Now in Nelson's formalism of stochastic quantization, it has been taken that when ψ is written as α^{R+iS}, the current velocity V is associated with $\partial S/\partial x$, and the osmotic velocity u is associated with $\partial R/\partial x$. But when Nelson's procedure is generalized to have relativistic generalization and the quantization of Fermi fields,[6] the internal current velocity is related with $\partial R/\partial\xi$. This can again be associated with $\partial R/\partial x$ through the mapping of x and ξ. Now the osmotic velocity in this case represents the situation when the sample path terminates at the point x at time t, which actually corresponds to the motion in the negative time axis. Since the term $\partial R/\partial x$ corresponds to this velocity, and the term $\partial R/\partial\xi$ represents the internal current velocity, we can associate the field in the internal space with the time-reversal field. This suggests that the mapping $\phi(x)$ and $\phi(\xi)$ is isomorphic to the Z_2 symmetry corresponding to time-reversal invariance. This leads to the fact that $\phi(x)$ and $\tilde{\phi}^+(x)$ have this Z_2 symmetry through the mapping $\phi(\xi)=\tilde{\phi}^+(x)=\lambda\phi^+(x)$, $\lambda\in C$. This suggests that the tilde operation of Takahashi and Umezawa actually corresponds to the discrete symmetry operation Z_2, and is thus equivalent to the Niemi–Semenoff formalism.

Now to study the equilibrium condition, we have argued that in this formalism of stochastic field theory, we should have the KMS condition in the internal space also apart from that in the external space. This leads to the physical relevance of the HHW modular conjugation algebra and the Tomita–Takesaki theory of modular Hilbert algebra. As the external and internal fields may be associated with Z_2 symmetry corresponding to time-reversal invariance, we find that the necessity of the existence of the KMS condition in the internal space, in addition to that in the external space, is a manifestation of the fact that Z_2 symmetry is the necessary condition for the equilibrium state as proposed by Niemi and Semenoff. This equilibrium is destroyed by the violation of the KMS condition in the internal space through the thermal change in the isotropic or anisotropic nature of this space, leading to a change in statistics of the thermal pair, and thus in that case we will have the spontaneous breakdown of Z_2 symmetry.

V. DISCUSSION

We have studied here the equilibrium condition in terms of the stochastic field theory at finite temperature involving fields in the external and internal space, and have shown that the HHW formalism of modular conjugation as well as the Tomita–Takesaki theory of modular Hilbert algebra find a physical relevance in terms of stochastic field in the internal space. Besides the ghost field of Niemi and Semenoff, the Z_2 symmetry requirement for the equilibrium condition becomes equivalent to this formalism when the ghost field is associated with the field in the internal space. When the equilibrium is destroyed, the isotropic or the anisotropic feature of the internal space gets changed, leading to the change in statistics of the thermal doublet, and as such Z_2 symmetry is violated. This situation may arise when supersymmetry is restored at high temperture from a physical state of broken supersymmetry at low temperature.

It may be noted that when the equilibrium condition is studied from the viewpoint of stochastic fields, the correlation functions involved have been considered taking into ac-

count the KMS condition, and as such the KMS condition is built into this formalism. Thus it may appear surprising to consider that the Z_2 symmetry, just as the KMS condition in the internal space, is an *essential* condition for equilibrium, as the correlation function in the external space itself has also the implicit KMS condition in that space. Indeed, as the thermofield dynamics of Takahashi and Umezawa, as well as the Niemi–Semenoff theory of finite temperature quantum field theory at finite temperature, necessitate an extra field which is identified here with the field involving internal space, it is evident that the KMS condition in the external space itself is not sufficient to study the equilibrium condition. This becomes transparent from the fact that a stochastic field involving only the external space faces serious trouble as the two-point correlation function at $T=0$, $\langle\phi(x,t)\phi(x',t')\rangle$, involving only external space-time variables, is not Lorentz invariant (rotationally invariant) when it is derived from a finite temperature correlation function taking the limit $\beta\to\alpha$.[8] Indeed, the stochastic fluctuations operating at $T\neq 0$ still have their residual effect at $T=0$ through the moment of the component oscillators. However, when the moments of the stochastic fields are determined incorporating two fields, one in the external variables and the other in the internal variables, this Lorentz invariance may be restored through CPT invariance, as the Z_2 symmetry manifested in this two-field formalism implies time-reversal invariance, which again becomes equivalent to CP symmetry. Thus the consistency requirement of stochastic fields requires the two-field formalism, which helps us to associate the stochastic fields at finite temperature with the finite temperature quantum field theory in Minkowski space. This also helps us to study the equilibrium condition giving a physical insight into the HHW formalism of modular conjugation and the Tomita–Takesaki theory of modular Hilbert algebra.

[1]P. Bandyopadhyay, B. Bhattacharya, and P. Ghosh, "Stochastic Field Theory of Finite Temperature and Thermofield Dynamics" (submitted for publication).
[2]Y. Takahashi and H. Umezawa, Collect. Phenom. **2**, 151 (1974).
[3]A. J. Niemi and G. W. Semenoff, Ann. Phys. **152**, 105 (1983).
[4]F. Guerra and P. Ruggiero, Phys. Rev. Lett. **31**, 1022 (1973).
[5]E. Nelson, Phys. Rev. **150**, 1079 (1966); *Dynamical Theory of Brownian Motion* (Princeton U.P., Princeton, NJ, 1967).
[6]P. Bandyopadhyay and K. Hajra, J. Math. Phys. **28**, 711 (1987)
[7]E. S. Abers and B. W. Lee, Phys. Lett. C **9** (1973).
[8]S. M. Moore, J. Math. Phys. **21**, 2102 (1980).
[9]P. Bandyopadhyay, Hadronic J. 7, 1706 (1984).

Inelastic two-composite-particle system scattering theory at high energy

Zhang Yu-shun
Center of Theoretical Physics, CCAST (World Lab.), Beijing, People's Republic of China and Institute of High Energy Physics, Academia Sinica, Beijing, People's Republic of China

(Received 9 January 1988; accepted for publication 11 January 1989)

In this paper, the methods of nucleons outside the closed shell (NOCS) and nucleons inside the closed shell (NICS) is introduced into the composite-particle systems, and Glauber's multiple scattering theory is used to divide the multiple scattering amplitude (between two composite particle systems) into NOCS–NOCS, NOCS–NICS, NICS–NOCS, and NICS–NICS scattering amplitudes. In order to apply these concepts to our research, an effective approximate method is used for each scattering amplitude. For example, an independent particle model approximation has been used for the NOCS–NOCS scattering amplitude. As for the NICS–NICS and NICS–NOCS scattering amplitudes, the geometrical model and Bohr's collective model are adopted, respectively. It can be shown that the high-energy scattering S matrix elements for composite-particle systems present a clear physical picture and render convenience to calculation.

I. INTRODUCTION

In nuclear physics, the concept of the shell model has been widely used in the discussion of various physical problems. The existence of the closed shell makes it possible to divide a composite-particle system into nucleons outside the closed shell (NOCS) and nucleons inside the closed shell (NICS).[1] It is convenient to discuss the single particle effect and the collective effect separately if the concept of the shell model is introduced in the scattering theory, so that the physical properties can be more satisfactorily approached. So far, the methods of NOCS and NICS have not been adopted in Glauber's scattering theory.[2] In this paper we try to discuss the problems of high energy scattering of composite particle systems within the framework of Glauber's theory.

This paper is organized as follows. In Sec. II, the dividing NOCS and NICS is described, in Sec. III, the S matrix of a particle and NICS cluster is described, in Sec. IV, the S matrix of the NOCS–NICS and collective model are discussed, in Sec. V, the S matrix of NOCS–NOCS is described, in Sec. VI, the S matrix of NICS-NICS and the geometrical model are discussed, and in Sec. VII discussions and conclusions from the work are given.

II. THE DIVIDING INTO THE NOCS AND NICS

According to Glauber's theory,the multiple scattering amplitude between the composite particle systems A and B, i.e., for the $A+B\to A^*+B^*$ scattering process, is given by[3]

$$F_{fi}(\mathbf{q})=\Theta(\mathbf{q})\frac{ik}{2\pi}\int d\mathbf{b}\, e^{i\mathbf{q}\cdot\mathbf{b}} \times\int \mathbf{dx}\,\mathbf{dy}\,\psi^*_{A_f}(\mathbf{x})\psi^*_{B_f}(\mathbf{y})\Gamma\psi_{A_i}(\mathbf{x})\psi_{B_i}(\mathbf{y}), \quad (2.1)$$

where $\Theta(\mathbf{q})$ is the correction factor for the center of mass; k, the momentum of the incident particle system; $\mathbf{q}$, the momentum transfer; $\mathbf{b}$, the impact parameter; $\psi^*_{B_f}(\mathbf{y})$, $\psi^*_{A_f}(\mathbf{x})$ and $\psi_{B_i}(\mathbf{y})$, $\psi_{A_i}(\mathbf{x})$ are the final and the initial states of the target and incident particle system, respectively, where $\mathbf{X}$ stands for $\mathbf{X}_i\cdots X_A$, the coordinates of NOCS and $\mathbf{X}_{A+1}\cdots\mathbf{X}_{N_A}$, the coordinates of NICS, and the coordinate $\mathbf{y}$ may be defined in the same manner;

$$d\mathbf{x}=d\mathbf{x}_1\cdots d\mathbf{x}_A\, d\mathbf{x}_{A+1}\cdots d\mathbf{x}_{N_A} =\prod_{i=1}^{A} d\mathbf{x}_i \prod_{j=A+1}^{N_A} d\mathbf{x}_j,$$

$$d\mathbf{y}=d\mathbf{y}_1\cdots d\mathbf{y}_B\, d\mathbf{y}_{B+1}\cdots d\mathbf{y}_{N_B} =\prod_{\alpha=1}^{B} d\mathbf{y}_\alpha \prod_{\beta=B+1}^{N_B} d\mathbf{y}_\beta.$$

The total profile function Γ is given by

$$\Gamma=1-\prod_{k=1}^{N_A}\prod_{\gamma=1}^{N_B}[1-\Gamma_{k\gamma}(\mathbf{b}+\mathbf{S}_k-\boldsymbol{\sigma}_\gamma)], \quad (2.2)$$

where $\mathbf{S}_K$ and $\boldsymbol{\sigma}_\gamma$ are the projections of the particle coordinate $\mathbf{X}_K$ and $\mathbf{Y}_\gamma$ on the plane perpendicular to $\mathbf{k}$, respectively, and $\Gamma_{k\gamma}(\mathbf{b})$ is the two-body profile function. The relation between this function and the two-body scattering amplitude is

$$\Gamma_{k\gamma}(\mathbf{b})=\frac{1}{2\pi ik'}\int d\mathbf{q}\, e^{-i\mathbf{q}\cdot\mathbf{b}}\cdot f_{k\gamma}(\mathbf{q}). \quad (2.3)$$

In high-energy collision, if the spin effect is neglected, the two-body scattering amplitude is usually written as

$$f_{k\gamma}(\mathbf{q})=(ik'/4\pi)a(1-i\rho)e^{-\beta^2q^2/2}, \quad (2.4)$$

where a is the total cross section of the two-body scattering, ρ, the ratio of the real to the imaginary part of the forward amplitude, β, the slope parameter, and k', the momentum of a incident particle.

In general, the NOCS is more active and easier to be excited, but the character of the NICS has more collectivity. Therefore, when two-composite-particle system collision $A+B\to A^*+B^*$ is to be considered, both A and B must be

 0022-2488/89/061365-08$02.50

divided into two parts, the NOCS and the NICS. Then the multiple scattering amplitudes of the two composite particle systems, Eq. (2.1), can also be divided into NOCS–NOCS, NOCS–NICS, NICS–NOCS, and NICS–NICS multiple scattering amplitudes that can be treated by different effective approximate methods, respectively. So the total profile function in Eq. (2.2) can be written as

$$\begin{aligned}&\Gamma(\mathbf{b},\mathbf{s}_1\cdots\mathbf{s}_{N_A},\boldsymbol{\sigma}_1\cdots\boldsymbol{\sigma}_{N_B})\\&\quad=1-[1-\Gamma_1(\mathbf{b},\mathbf{s}_1\cdots\mathbf{s}_A,\boldsymbol{\sigma}_1\cdots\boldsymbol{\sigma}_B)]\\&\qquad\cdot[1-\Gamma_2(\mathbf{b},\mathbf{s}_1,\cdots\mathbf{s}_A,\boldsymbol{\sigma}_{B+1}\cdots\boldsymbol{\sigma}_{N_B})]\\&\qquad\cdot[1-\Gamma_3(\mathbf{b},\mathbf{s}_{A+1}\cdots\mathbf{s}_{N_A},\boldsymbol{\sigma}_1\cdots\boldsymbol{\sigma}_B)]\\&\qquad\cdot[1-\Gamma_4(\mathbf{b},\mathbf{s}_{A+1}\cdots\mathbf{s}_{N_A},\boldsymbol{\sigma}_{B+1}\cdots\boldsymbol{\sigma}_{N_B})],\end{aligned}\tag{2.5}$$

where

$$\begin{aligned}&\Gamma_1(\mathbf{b},\mathbf{s}_1\cdots\mathbf{s}_A,\boldsymbol{\sigma}_1\cdots\boldsymbol{\sigma}_B)\\&\quad=1-\prod_{i=1}^{A}\prod_{\alpha=1}^{B}[1-\Gamma_{i\alpha}(\mathbf{b}+\mathbf{s}_i-\boldsymbol{\sigma}_\alpha)],\\&\Gamma_2(\mathbf{b},\mathbf{s}_1\cdots\mathbf{s}_A,\boldsymbol{\sigma}_{B+1}\cdots\boldsymbol{\sigma}_{N_B})\\&\quad=1-\prod_{i=1}^{A}\prod_{\beta=B+1}^{N_B}[1-\Gamma_{i\beta}(\mathbf{b}+\mathbf{s}_i-\boldsymbol{\sigma}_\beta)],\\&\Gamma_3(\mathbf{b},\mathbf{s}_{A+1}\cdots\mathbf{s}_{N_A},\boldsymbol{\sigma}_1\cdots\boldsymbol{\sigma}_B)\\&\quad=1-\prod_{j=A+1}^{N_A}\prod_{\alpha=1}^{B}[1-\Gamma_{j\alpha}(\mathbf{b}+\mathbf{s}_j-\boldsymbol{\sigma}_\alpha)],\end{aligned}$$

$$\begin{aligned}&\Gamma_4(\mathbf{b},\mathbf{s}_{A+1}\cdots\mathbf{s}_{N_A},\boldsymbol{\sigma}_{B+1}\cdots\boldsymbol{\sigma}_{N_B})\\&\quad=1-\prod_{j=A+1}^{N_A}\prod_{\beta=B+1}^{N_B}[1-\Gamma_{j\beta}(\mathbf{b}+\mathbf{s}_j-\boldsymbol{\sigma}_\beta)],\end{aligned}\tag{2.6}$$

the profile functions of interactions between the NOCS–NOCS, NOCS–NICS, NICS–NOCS, and NICS-NICS are described, respectively. Similarly, the wave functions of the systems A and B are divided into the wave function of NOCS, part u, and NICS, part v, respectively. They are

$$\begin{aligned}&\psi_{A_i}(\mathbf{x})=u_{A_i}(\mathbf{x}_1\cdots\mathbf{x}_A)v_{A_i}(\mathbf{x}_{A+1}\cdots\mathbf{x}_{N_A})\equiv u_{A_i}v_{A_i},\\&\psi_{B_i}(\mathbf{y})=u_{B_i}(\mathbf{y}_1\cdots\mathbf{y}_B)v_{B_i}(\mathbf{y}_{BH}\cdots\mathbf{y}_{N_B})\equiv u_{B_i}v_{B_i},\\&\psi_{A_f}=u_{A_f}(\mathbf{x}_1\cdots\mathbf{x}_A)v_{A_f}(\mathbf{x}_{A+1}\cdots\mathbf{x}_{N_A})\equiv u_{A_f}v_{A_f},\\&\psi_{B_f}(\mathbf{y})=u_{B_f}(\mathbf{y}_1\cdots\mathbf{y}_B)v_{B_f}(\mathbf{y}_{B+1}\cdots\mathbf{y}_{N_B})\equiv u_{B_f}v_{B_f}.\end{aligned}\tag{2.7}$$

The forward peak and the diffraction pattern are the features of the high-energy particle–particle collision. Since the interaction time between particle–particle is very short, the particles passing through the target nucleus generally do not come into collision with the same nucleon in target nucleus several times. Therefore, the contribution of the intermediate states is not important in the multiple collision process, and can be neglected. Substituting Eqs. (2.5) and (2.7) into the integrand of Eq. (2.1), one obtains

$$\begin{aligned}&\int d\mathbf{x}\,d\mathbf{y}\,\psi^*_{A_f}(\mathbf{x})\psi^*_{B_f}(\mathbf{y})\Gamma\psi_{A_i}(\mathbf{x})\psi_{B_i}(\mathbf{y})\\&\quad=\delta_{fi}-\sum\int\prod_{i=1}^{A}d\mathbf{x}_i\prod_{\alpha=1}^{B}d\mathbf{y}_\alpha\,u^*_{A_f}u^*_{B_f}[1-\Gamma_1(\mathbf{b},\mathbf{s}_1\cdots\mathbf{s}_A,\boldsymbol{\sigma}_1\cdots\boldsymbol{\sigma}_B]u_{A_i}u_{B_i}\\&\qquad\times\int\prod_{i=1}^{A}d\mathbf{x}_i\prod_{\beta=B+1}^{N_B}d\mathbf{y}_\beta\,u^*_{A_i}v^*_{B_f}[1-\Gamma_2(\mathbf{b},\mathbf{s}_1\cdots\mathbf{s}_A,\boldsymbol{\sigma}_{B+1}\cdots\boldsymbol{\sigma}_{N_B})]u_{A_i}v_{B_i}\\&\qquad\times\int\prod_{j=A+1}^{N_A}d\mathbf{x}_j\prod_{\alpha=1}^{B}d\mathbf{y}_\alpha\,v^*_{A_f}u^*_{B_i}[1-\Gamma_B(\mathbf{b},\mathbf{s}_{A+1}\cdots\mathbf{s}_{N_A},\boldsymbol{\sigma}_1\cdots\boldsymbol{\sigma}_B)]v_{A_i}u_{B_i}\\&\qquad\times\int\prod_{j=A+1}^{N_A}d\mathbf{x}_j\prod_{\beta=B+1}^{N_B}d\mathbf{y}_\rho\,v^*_{A_i}v^*_{B_i}[1-\Gamma_4(\mathbf{b},\mathbf{s}_{A+1}\cdots\mathbf{s}_{N_A},\boldsymbol{\sigma}_{B+1}\cdots\boldsymbol{\sigma}_{N_B})]v_{A_i}v_{B_i}.\end{aligned}\tag{2.8}$$

The first integral in the second term on the right-hand side of Eq. (2.8) is the S matrix which represents the interaction between NOCS of A and that of B; the second integral is the S matrix between NOCS of A and NICS of B; the third integral is the S matrix between NICS of A and NOCS of B; and the fourth integral is the S matrix between NICS of A and that of B. On the basis of Eq. (2.8), in this paper we shall discuss these S matrices and treat them with different approximate methods. These methods can be used to discuss the particle scattering problems.

III. THE S MATRIX OF A PARTICLE AND NICS CLUSTER

According to Eq. (2.8), the scattering amplitude of a particle scattered by NICS cluster may be written as

$$\begin{aligned}&f_{(s)f_i}(\mathbf{q})\\&\quad=\frac{ik}{2\pi}\int d\mathbf{b}\,e^{i\mathbf{q}\cdot\mathbf{b}}\\&\qquad\times{}_s\left\langle f\left|\left[1-\prod_{j=1}^{\infty}[1-\Gamma_j(\mathbf{b}-\mathbf{s}_j)]\right]\right|i\right\rangle_s\\&\quad=\frac{ik}{2\pi}\int d\mathbf{b}\,e^{i\mathbf{q}\cdot\mathbf{b}}\\&\qquad\times{}_s\left\langle f\left|\left[1-\exp\left(-\frac{i}{\hbar w}\sum_{j=1}^{\infty}\int dz\,V(\mathbf{r}-\mathbf{x}_j)\right)\right]\right|i\right\rangle_s,\end{aligned}\tag{3.1}$$

where $V(\mathbf{r}-\mathbf{x}_j)$ is the interaction potential between the free particles and the jth particle in NICS; $|i\rangle_s$, $|f\rangle_s$ are the initial

and final states in NICS. Suppose the Hamiltonian of the NICS is $\mathscr{H}_s$, i.e.,

$$\mathscr{H}_s|i\rangle_s = \varepsilon_i|i\rangle_s. \tag{3.2}$$

Here we assume that the NICS occupies a certain space $\mathscr{D}$ and then divide $\mathscr{D}$ into infinitesimal scattering objects. Here we regard the NICS as a scattering medium with definite transparency. Thus a particle scattered by the NICS can be regarded as scattered by infinitesimal scattering objects. And assume that the volume of each small unit is ε^3; as shown in Fig. 1. We introduce $v(\mathbf{r}-\mathbf{x}_j)$ to describe the interaction density of the $V(\mathbf{r}-\mathbf{x}_j)$, then

$$V(\mathbf{r}-\mathbf{x}_j) = \varepsilon^3 v(\mathbf{r}-\mathbf{x}_j) \tag{3.3}$$

and

$$\lim_{j\to\infty}\sum_j V(\mathbf{r}-\mathbf{x}_j) = \lim_{j\to\infty}\sum_j \varepsilon^3 v(\mathbf{r}-\mathbf{x}_j) = \int_{\mathscr{D}} d\mathbf{x}\, v(\mathbf{r}-\mathbf{x}) = V(\mathbf{r}), \tag{3.4}$$

where $V(\mathbf{r})$ is the potential with respect to NICS; thus Eq. (3.1) is reduced to

$$f_{(s)f_i}(\mathbf{q}) = \frac{ik}{2\pi}\int d\mathbf{b}\, e^{i\mathbf{q}\cdot\mathbf{b}} \times {}_s\left\langle f\left|\left[1-\exp\left(\frac{-i}{\hbar v}\int dz V(\mathbf{r})\right)\right]\right|i\right\rangle_s, \tag{3.5}$$

the S matrix of the b representation is

$$S_{i-s}(\mathbf{b}) = {}_s\left\langle f\left|\exp\left(\frac{-i}{\hbar v}\int dz\, V(\mathbf{r})\right)\right|i\right\rangle_s = \int v_f^*(\mathbf{x}_1\cdots\mathbf{x}_n\cdots)\exp\left(\frac{-i}{\hbar v}\int dz\, V(\mathbf{r})\right) \times v_i(\mathbf{x}_1\cdots\mathbf{x}_n\cdots)\prod_i d\mathbf{x}_i. \tag{3.6}$$

Use $v_i(\mathbf{x}_1\cdots\mathbf{x}_n\cdots)$, $v_f(\mathbf{x}_1\cdots\mathbf{x}_n\cdots)$ to describe these collective states, denoted by $|c;i\rangle$, $|c;f\rangle$, i.e.,

$$\mathscr{H}_s|c;n\rangle = \varepsilon_n|c;n\rangle, \tag{3.6'}$$

then

$$v_i(\mathbf{x}_1\cdots\mathbf{x}_n\cdots) = \langle\mathbf{x}_1\cdots\mathbf{x}_n\cdots|c;i\rangle, \quad v_f(\mathbf{x}_1\cdots\mathbf{x}_n\cdots) = \langle\mathbf{x}_1\cdots\mathbf{x}_n\cdots|c;f\rangle. \tag{3.7}$$

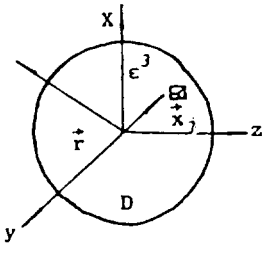


FIG. 1. Dividing NICS into infinitesimal objects and the volume of each small unit is ε^3.

Because of the orthogonal completeness of the basis vector

$$\int|\mathbf{x}_1\cdots\mathbf{x}_n\cdots\rangle\langle\mathbf{x}_1\cdots\mathbf{x}_n\cdots|\prod_i d\mathbf{x}_i = 1, \quad \langle\mathbf{x}_1\cdots\mathbf{x}_n'\cdots|\mathbf{x}_1\cdots\mathbf{x}_n\cdots\rangle = \delta(\mathbf{x}_1-\mathbf{x}_1')\cdots\delta(\mathbf{x}_n-\mathbf{x}_n')\cdots, \tag{3.8}$$

thus

$$S_{1-s}(\mathbf{b}) = \int\langle c;f|\mathbf{x}_1\cdots\mathbf{x}_n\cdots\rangle\exp\left(\frac{-i}{\hbar v}\int dz\, V(\mathbf{r})\right) \times\langle\mathbf{x}_1\cdots\mathbf{x}_n\cdots|c;i\rangle\prod_i d\mathbf{x}_i = \left\langle c;f\left|\exp\left(\frac{-i}{\hbar v}\int dz\, V(\mathbf{r})\right)\right|c;i\right\rangle. \tag{3.9}$$

The multiple scattering amplitude of a particle scattered by the NICS cluster is given by

$$f_{(s)f_i}(\mathbf{q}) = \frac{ik}{2\pi}\left\langle c;f\left|\times\int d\mathbf{b}\, e^{i\mathbf{q}\cdot\mathbf{b}}\left[1-\exp\left(\frac{-i}{\hbar v}\int dz\, V(\mathbf{r})\right)\right]\right|c;i\right\rangle = \left\langle c;f\left|\hat{f}_{(s)}(\mathbf{q})\right|c;i\right\rangle, \tag{3.10}$$

where

$$\hat{f}_{(s)}(\mathbf{q}) = \frac{ik}{2\pi}\int d\mathbf{b}\, e^{i\mathbf{q}\cdot\mathbf{b}}\left[1-\exp\left(\frac{-i}{\hbar v}\int dz\, V(\mathbf{r})\right)\right] \tag{3.11}$$

is called the scattering operator. When the $\hat{f}_{(s)}(\mathbf{q})$ operates on the collective state $|c;i\rangle$, it transits to the collective excited state $|c;f\rangle$.

(1) As a result of collective motion, $\mathscr{D}$ will vibrate, or rotate. If the target nucleus is spherically symmetric, but is susceptible to vibration around that spherical shape, R may be expressed as

$$R = R_0\left(1+\sum_{\lambda\mu}\alpha_{\lambda\mu}Y_{\lambda\mu}(\theta,\psi)\right), \tag{3.12}$$

where $R_0 = \gamma_0 A'^{1/3}$, A' is the mass number of the target. Assume that the NICS interaction to which an incident particle is subjected is described by an optical model potential $V(\mathbf{r})$. Then we insert (3.12) into (3.11) and expand the latter in powers of

$$\sum_{\lambda\mu}\alpha_{\lambda\mu}Y_{\lambda\mu}$$

to get[4]

$$V(\mathbf{r}) \simeq V_0(r) - \sum_{\lambda\mu}\frac{d}{dr}V_0(r)\cdot\alpha_{\lambda\mu}Y_{\lambda\mu}(\theta,\varphi). \tag{3.13}$$

Now we consider a nucleus with a set of quadrupole vibration.

Following Buck,[5] the relation between α_μ and phonon operator can be written as

$$\alpha_\mu = \gamma_2(1/\sqrt{5})(a_\mu + (-1)^\mu\cdot a^*_{-\mu}), \tag{3.14}$$

assuming that the nuclear excitation energy may be neglected compared with the projectile energy, in which case the scattering operator (3.11) may be rewritten as

$$\hat{f}_{(v)}(\mathbf{q}) = \frac{ik}{2\pi}\int d\mathbf{b}\, e^{i\mathbf{q}\cdot\mathbf{b}}\Big\{1-\exp\Big[-X_0(b) - \sum_{\mu}\Big(X_{z\mu}a_{\mu} + X^{*}_{z\mu}a^{*}_{\mu}\Big)\Big]\Big\}, \tag{3.15}$$

where

$$\chi_0(b) = \frac{i}{\hbar v}\int_{-\infty}^{\infty} dz\, V_0(r), \tag{3.16a}$$

$$\chi_{z\mu}(\mathbf{b}) = \frac{ir_z}{\hbar v\sqrt{5}}\int_{-\infty}^{\infty} dz\frac{d}{dr}V_0(r)\cdot Y_{z\mu}(\theta,\varphi). \tag{3.16b}$$

To evaluate the integral for $\chi_{z\mu}$ one must take account of the Z dependence of $Y_{z\mu}$. If $(d/dr)V_0$ depends only on $r=(b^2+z^2)^{1/2}$, the $Y_{2\pm1}$ terms will integrate to zero since they are odd in z. Further, assuming that the leading contribution will come from near $z\simeq 0$,[6] we thus obtain

$$\chi_{zs}(b) = \frac{ir_2}{\hbar v\sqrt{5}}\frac{1}{2}\left(\frac{5}{4\pi}\right)^{1/2}\int_{-\infty}^{\infty} dz\frac{d}{dr}V_0(r) \equiv -X_1(b), \tag{3.17a}$$

$$\chi_{2\pm2}(b) = \frac{ir_2}{\hbar v\sqrt{5}}\frac{1}{2}\left(\frac{15}{8\pi}\right)^{1/2} e^{\pm i2\varphi}\int_{-\infty}^{\infty} dz\frac{d}{dr}V_0(r) = \sqrt{\tfrac{3}{2}}e^{\pm i2\varphi}\chi_1(b). \tag{3.17b}$$

Define

$$g = -a_0 + \sqrt{\tfrac{3}{2}}(a_2e^{-i2\varphi} + a_{-2}e^{i2\varphi}). \tag{3.18}$$

We can write for the scattering operator

$$\hat{f}_{(v)}(\mathbf{q}) = \frac{ik}{2\pi}\int \mathbf{db}\, e^{i\mathbf{q}\cdot\mathbf{b}} e^{-X_0(b)}\cdot e^{-X_1(b)(g+g^{+})}. \tag{3.19}$$

Here we have dropped the 1 that contributes only to forward elastic scattering and assumed $x_1 = x_1^{*}$.

To find this it is convenient to use the Baker–Campbell–Hausdorff formula[7]

$$e^{-X_1(b)(g+g^{+})} = e^{-X_1(b)g^{+}}\cdot e^{-X_1(b)g}\cdot e^{2X_1^2(b)} \tag{3.20}$$

since $[g^{+},g] = -4$, then

$$\hat{f}_{(r)}(\mathbf{q}) = \frac{ik}{2\pi}\int d\mathbf{b}\, e^{i\mathbf{q}\cdot\mathbf{b}}\cdot e^{-X_0(b)}\cdot e^{2X_1^2(b)}\cdot e^{-X_1(b)g^{+}}\cdot e^{-X_1(b)g}. \tag{3.21}$$

(2) On the other hand, if the target is an (axially symmetric) deformed nucleus[4]

$$R = R_0\Big(1+\sum_{\lambda}\beta_\lambda Y_{\lambda 0}(\theta')\Big), \tag{3.22}$$

where the angle θ' refers to the body fixed system. β is the conventional nuclear deformation parameter. $\beta>0$ and $\beta<0$ refer, respectively, to prolate and oblate deformation. We assume that $V(\mathbf{r})$ is the optical-model potential. If the potential is again expanded in powers of $\Sigma_\lambda\beta_\lambda Y_{\lambda 0}(\theta')$, one gets exactly the same expression as in (1), namely, Eqs. (3.13), except that $\alpha_{\lambda\mu}$ and $Y_{\lambda\mu}(\theta,\varphi)$ are replaced, respectively, by β_λ and

$$Y_{\lambda\theta}(\theta') = \sum_{\mu} D^{\lambda}_{\mu 0}(\theta_i)\, Y_{\lambda\mu}(\theta\varphi).$$

Here $D^{\lambda}_{\mu 0}(\theta_i)$ is a rotation matrix and θ_i stands for the Euler angles between the body-fixed and the space-fixed coordinates. Then[8]

$$V(\mathbf{r}) \simeq V_0(r) - \sum_{\lambda\mu}\beta_\lambda V_{\lambda\mu}, \tag{3.23}$$

where

$$V_{\lambda\mu} = \frac{d}{dr}V_0(r)D^{\lambda}_{\mu 0}(\theta_i)\, Y_{\lambda\mu}(\theta,\varphi). \tag{3.23'}$$

In the rotational case, the scattering operator (3.11) may be rewritten as

$$\hat{f}_{(R)}(\mathbf{q}) = \frac{ik}{2\pi}\int d\mathbf{b}\, e^{i\mathbf{q}\cdot\mathbf{b}}\Big\{1-\exp\Big[-\Lambda_0(b) - \sum_{\lambda\mu}\beta_\lambda\omega_{\lambda\mu}(b,\varphi,\theta_i)\Big]\Big\}, \tag{3.24}$$

where

$$\Lambda_0(b) = \frac{i}{\hbar v}\int_{-\infty}^{\infty} dz\, V_0(r), \tag{3.25a}$$

$$\omega_{\lambda\mu}(b) = \frac{i}{\hbar v}\int dz\frac{d}{dr}V_0(r)D^{\lambda}_{\mu 0}(\theta_i)\, Y_{\lambda\mu}(\theta,\varphi),$$

for an axially symmetric nucleus with quadrupole deformation $\lambda=2$, and we now make a further hypothesis, $Z=0$, so that $Y_{2\mu}$ has a very simple representation; then we have

$$\omega_{20}(b) = \frac{i}{\hbar v}\int dz\frac{d}{dr}V_0(r)\cdot\frac{1}{4}\left(\frac{5}{4\pi}\right)^{1/2} \equiv \Lambda_1(b),$$

$$\omega_{2\pm2}(b) = \frac{i}{\hbar v}\int dz\frac{d}{dr}V_0(r)\frac{3}{8\pi}\left(\frac{5}{4\pi}\right)^{1/2}\cdot e^{\mp i2\Phi'} = \frac{3}{2\pi}\Lambda_1(b)e^{\mp i2\Phi'}, \tag{3.25b}$$

where θ_i may be specified in terms of polar angles (θ,Φ). Here $\Phi' = \Phi - \varphi$, then Eq. (3.24) can be reduced to

$$\hat{f}_{(R)}(q) = ik\int b\, db\, J_0(qb)e^{-\Lambda_0(b)}\cdot e^{\beta_2\Lambda_1(b)t(\Phi')}, \tag{3.26}$$

where $t(\Phi') = 1 + (3/\pi)\cos 2\Phi'$.

IV. THE S MATRIX OF THE NOCS–NICS AND COLLECTIVE MODEL

The NOCS are part of the particles bound in the composite particle system. Now we discuss the scattering matrix of the NOCS in B and the NICS in A. In Ref. 9, the symbol $u_i(y_1\cdots y_B)$ is used to represent the wave function of the NOCS in B and the scattering matrix of the elastic channel of NOCS–NICS is only discussed. In this paper, we will further discuss the scattering matrix containing the inelastic chan-

nel and NICS can be transmitted to the excited collective state. The S matrix of the NOCS and NICS may be written as

$$S_{v-s}(\mathbf{b}) = \int v_f^*(\mathbf{x}_1\cdots\mathbf{x}_n\cdots)u_i^*(\mathbf{y}_1\cdots\mathbf{y}_B) \times\exp\left(\frac{-i}{\hbar w}\sum_{j=1}^{\infty}\sum_{\alpha=1}^{B}\int V(\mathbf{r}-\mathbf{x}_i+\mathbf{y}_\alpha)dz\right) \cdot v_i(\mathbf{x}_1\cdots\mathbf{x}_n\cdots)u_i(\mathbf{y}_1\cdots\mathbf{y}_B) \times\prod_{i=1}^{\infty}d\mathbf{x}_i\prod_{\alpha=1}^{B}d\mathbf{y}_\alpha. \tag{4.1}$$

At high energy, it is convenient to use the probability density to describe the state of NOCS, i.e.,[10]

$$u_i^*u_i = \prod_{\alpha=1}^{B}\rho(\mathbf{y}_\alpha), \tag{4.2}$$

when the number of NICS is infinite; we shall consider that there exists to excite the collective state $|c;f\rangle$ of NICS. Hence Eq. (4.1) may be rewritten as

$$S_{v-s}(\mathbf{b}) = \left\langle c;f\left|\left[\int\prod_{\alpha=1}^{B}\rho(\mathbf{y}_d)d\mathbf{y}_\alpha \times\exp\left(\frac{-i}{t,v}\sum_{\alpha=1}^{B}\int_{\mathscr{S}}\mathbf{dx}\,dz\,v(\gamma-\mathbf{x}+\mathbf{y}_\alpha)\right)\right]\right|c;i\right\rangle. \tag{4.3}$$

In Eq. (3.11), the amplitude operator $\hat{f}_{(s)}(\mathbf{q})$ with respect to the profile function is

$$\hat{\Gamma}_{(s)}(\mathbf{b}) = 1-\exp\left(\frac{-i}{\hbar w}\int_{\mathscr{D}}d\mathbf{x}\,dz\,v(\mathbf{r}-\mathbf{x})\right) = \frac{1}{2\pi ik'}\int d\mathbf{q}'\,e^{-\mathbf{q}'\cdot\mathbf{b}}\hat{f}_{(s)}(\mathbf{q}'). \tag{4.4}$$

Substituting Eq. (4.4) into (4.3) yields

$$S_{v-s}(\mathbf{b}) = \left\langle c;f\left|\left[1-\frac{1}{2\pi ik'}\int d\mathbf{q}'\,d\mathbf{y}\,\rho(\mathbf{y}) \times e^{-i\mathbf{q}'\cdot(\mathbf{s}_y+\mathbf{b})}\cdot\hat{f}_{(s)}(\mathbf{q}')\right]^B\right|c;i\right\rangle. \tag{4.5}$$

Let the form factor $S_B(\mathbf{q})$ of NOCS be

$$S_B(\mathbf{q}) = \int d\mathbf{y}\,\rho(\mathbf{y})\cdot e^{-i\mathbf{q}\cdot\mathbf{y}}. \tag{4.6}$$

For the α particle $S_B(q) = A_l\,e^{(-q^2/4k_1^2)} - A_2\,e^{(q^2/4k_2^2)}$, $A_l = k_2^3/(k_2^3 - c\cdot k_2^3)$, $A_2 = c\cdot k_1^3/(k^3 c\cdot k_1^3)$, where c, k_1 and k_2 are parameters,[11] so that

$$S_{v-s}(\mathbf{b}) = \langle c;f|\left[1-\frac{1}{2\pi ik'}\int d\mathbf{q}'\,S_\alpha(\mathbf{q}') \times\hat{f}_{(s)}(\mathbf{q}')e^{-i\mathbf{q}'\cdot b}\right]^B|c;i\rangle, \tag{4.7}$$

where B is the number of NOCS; $\hat{f}_{(s)}(\mathbf{q})$ is the amplitude operator of the interaction between a free particle and NICS cluster. When $\hat{f}_{(s)}(\mathbf{q})$ operates on the collective state $|c;i\rangle$, it transits to $|c;f\rangle$. If $\hat{f}_{(s)}(\vec{q})$ and the form factor $s_\alpha(\mathbf{q})$ are given, we can obtain the S matrix of the NOCS–NICS, which contains the inelastic channel.

For the vibrational nucleus, if the target nucleus is spherically symmetric, $\hat{f}_{(v)}(\mathbf{q})$ is given by Eq. (3.21). Substituting Eq. (3.21) into Eq. (4.7), and then integrating for $d\mathbf{q}'$, note that $g|0\rangle = 0$. And for the α particle,

$$\left[1-\frac{1}{2\pi ik'}\int d\mathbf{q}'\,S_\alpha(\mathbf{q}')e^{-i\mathbf{q}'\cdot\mathbf{b}}\hat{f}_{(v)}(\mathbf{q}')\right]^4 = \sum_{j=1}^{4}C_4^j\left[1-\int b'\,db'\,e^{-X_0(b')}\cdot e^{2X_1^2(b')} \times e^{-X_1(b')g^+}\Pi(bb')\right]^j, \tag{4.8}$$

where

$$\Pi(bb') = 2A_1k_1^2e^{-k_1^2(b-b')^2} - 2A_2k_2^2\cdot e^{-k_2^2(b-b')^2},$$

so the S matrix from the ground state (no phonon) to the excited state (N phonon) is

$$S_{\alpha-s}(b) = \left\langle N\left|\sum_{j=1}^{4}C_4^j \times[\tilde{A}_0(b)+\tilde{A}_1(b)+\cdots]^j\right|0\right\rangle, \tag{4.9}$$

where

$$\tilde{A}_0(b) = \int b'\,db'\cdot e^{-X_0(b')}\cdot e^{2X_1^2(b')}\Pi(bb'),$$

$$\tilde{A}_1(b) = \int b'\,db'\,e^{-X_0(b')}\cdot e^{2X_1^2(b')}(-X_1(b'))\Pi(bb')g^+.$$

For a rotational nucleus, $\hat{f}_{(R)}(\mathbf{q})$ is given by Eq. (3.27). Substituting Eq. (3.27) into Eq. (4.7), and then integrating for $d\mathbf{q}'$, we get

$$\left[1-\frac{1}{2\pi ik'}\int d\mathbf{q}'\,S_\alpha(\mathbf{q}')e^{-i\mathbf{q}'\cdot\mathbf{b}}\hat{f}_{(R)}(q')\right]^4 = \sum_{j=1}^{4}C_4^j\left[-\int b'\,db'e^{-\Lambda_0(b')}e^{\beta_2\cdot\Lambda_1(b')t(\Phi')}\Pi(bb')\right]^j. \tag{4.10}$$

The $t(\Phi')$ will depend only on $\Phi-\varphi$, the difference between the nuclear azimuthal coordinate and the projectile, by virtue of our choice of Z axis. We can also write $Y_{LM}(\theta_i) = p_{LM}(\theta)e^{iM\Phi}$. If we change variables to $\Phi' = \Phi-\varphi$ and $\theta_i' = (\theta,\Phi')$, we have

$$S_{\alpha-s}(b) = \int d\theta_i\,Y_{LM}^*(\theta_i')Y_{00}(\theta_i')\sum_{j=1}^{4}C_4^j\left[-\int b'\,db' - e^{-\Lambda_0(b')}\cdot e^{\beta_2\Lambda_1(b')t(\Phi')}\Pi(bb')\right]^j. \tag{4.11}$$

In the same way, we can obtain the S matrix element of NICS–NOCS.

V. THE S MATRIX OF NOCS–NOCS

In Eq. (2.8) the S matrix of NOCS–NOCS can be written as

$$S_{v-v}(\mathbf{b}) = \int\prod_{i=1}^{A}\prod_{\alpha=1}^{B}d\mathbf{x}_i\,d\mathbf{y}_\alpha\,u_{A_f}^*u_{B_f}^* \times[1-\Gamma_1(\mathbf{b},\mathbf{s}_1\cdots\mathbf{s}_A,\boldsymbol{\sigma}_1\cdots\boldsymbol{\sigma}_B)]u_{A_i}u_{B_i}$$

$$= \int \prod_{i=1}^{A} \prod_{\alpha=1}^{B} \alpha \mathbf{x}_i \, d\mathbf{y}_\alpha \, u^*_{A_f} u^*_{B_f} \times [1 - \Gamma_{i\alpha}(\mathbf{b} + \mathbf{s}_i - \boldsymbol{\sigma}_\alpha)] u_{A_i} u_{B_i}. \quad (5.1)$$

In Ref. 9, NOCS–NOCS scattering with the PRA method is treated, but they did not consider the excitations in systems A^* and B^* after collision. The NOCS are small in number and easier to be excited, so we may extend the RPA method to the particle excitation. The method of the double scattering amplitude approximation is used to calculate this S matrix. The motion of the NOCS can be described by the use of an independent particle model which does not lose generality. For simplicity, the statistical characteristics of NOCS are not to be considered. Then the wave functions of ingoing and outgoing systems can be written as follows:

$$u_{A_i} \equiv u_{A_i}(\mathbf{x}_1 \cdots \mathbf{x}_A) = u^{(1)}_{A_i}(\mathbf{x}_1) \cdots u^{(A)}_{A_i}(\mathbf{x}_A),$$
$$u_{B_i} \equiv u_{B_i}(\mathbf{y}_1 \cdots \mathbf{y}_B) = u^{(1)}_{B_i}(\mathbf{y}_1) \cdots u^{(B)}_{B_i}(\mathbf{y}_B),$$
$$u_{A_f} \equiv u_{A_f}(\mathbf{x}_1 \cdots \mathbf{x}_A) = u^{(1)}_{A_f}(\mathbf{x}_1) \cdots u^{(A)}_{A_f}(\mathbf{x}_A),$$
$$u_{B_f} \equiv u_{B_f}(\mathbf{y}_1 \cdots \mathbf{y}_B) = u^{(1)}_{B_f}(\mathbf{y}_1) \cdots u^{(B)}_{B_f}(\mathbf{y}_B), \quad (5.2)$$

where $u^{(i)}_A(\mathbf{x}), u^{(\alpha)}_B(y)$ $(i = 1,...,A;\ \alpha = 1,...,B)$ are the single particle wave functions in NOCS satisfying the conditions of orthogonality and normalization

$$\int d\mathbf{x} \, u^{(i)*}_A(\mathbf{x}) u^{(i)}_{A'}(\mathbf{x}) = \delta_{AA'},$$
$$\sum_A u^{(i)}_A(\mathbf{x}) u^{(i)}_A(\mathbf{x}') = \delta(\mathbf{x} - \mathbf{x}'),$$
$$\int d\mathbf{y} \, u^{(\alpha)*}_B(\mathbf{y}) u^{(\alpha)}_{B'}(\mathbf{y}) = \delta_{BB'}, \quad (5.3)$$
$$\sum_B u^{(\alpha)}_B(\mathbf{y}) u^{(\alpha)}_B(\mathbf{y}') = \delta(\mathbf{y} - \mathbf{y}').$$

By using Eqs. (5.2) and (5.3), inserting the intermediate states into Eq. (5.1), only the ground state is considered and the contribution of excitation states is neglected, then

$$S_{v-v}(\mathbf{b}) = \int \prod_{i=1}^{A} d\mathbf{x}_i \prod_{\alpha=1}^{B} d\mathbf{y}_\alpha \, u^*_{A_f} u^*_{B_f} [1 - \Gamma_{11}(\mathbf{b} + \mathbf{S}_1 - \boldsymbol{\sigma}_1)] u_{A_i} u_{B_i}$$
$$\cdot \int \prod_{i=1}^{A} a\mathbf{x}_i \prod_{\alpha=1}^{B} d\mathbf{y}_\alpha \, u^*_{A_i} u^*_{B_i} [1 - \Gamma_{12}(\mathbf{b} + \mathbf{s}_1 - \boldsymbol{\sigma}_2)] u_{A_i} u_{B_i} \cdots$$
$$\cdot \int \prod_{i=1}^{A} d\mathbf{x}_i \prod_{\alpha=1}^{B} d\mathbf{y}_\alpha \, u^*_{A_i} u^*_{B_i} [1 - \Gamma_{AB}(\mathbf{b} + \mathbf{S}_A - \boldsymbol{\sigma}_B)] u_{A_i} u_{B_i}$$
$$= \int d\mathbf{x} \, d\mathbf{y} \, u^{(1)*}_{A_f}(\mathbf{x}) u^{(1)*}_{B_f}(\mathbf{y}) [1 - \Gamma_{11}(\mathbf{b} + \mathbf{s} - \boldsymbol{\sigma})] u^{(1)}_{A_i}(\mathbf{x}) u^{(1)}_{B_i}(\mathbf{y})$$
$$\times \prod_{(i,\alpha \neq 1,1)} \int d\mathbf{x} \, d\mathbf{y} \, u^{(i)*}_{A_i}(\mathbf{x}) u^{(\alpha)*}_{B_i}(\mathbf{y}) [1 - \Gamma_{i\alpha}(\mathbf{b} + \mathbf{S} - \boldsymbol{\sigma})] u^{(i)}_{A_i}(\mathbf{x}) u^{(\alpha)}_{B_i}(\mathbf{y})$$
$$= \int d\mathbf{x} \, d\mathbf{y} \, \rho^{(1)}_{A_{f_i}}(\mathbf{x}) \rho^{(1)}_{B_{f_i}}(\mathbf{y}) [1 - \Gamma_{11}(\mathbf{b} + \mathbf{s} - \boldsymbol{\sigma})] \prod_{(i,\alpha \neq 1,1)} \int d\mathbf{x} \, d\mathbf{y} \, \rho^{(i)}_A(\mathbf{x}) \rho^{(\alpha)}_B(\mathbf{y}) [1 - \Gamma_{i\alpha}(\mathbf{b} + \mathbf{s} - \boldsymbol{\sigma})]. \quad (5.4)$$

Introduce the probability density,

$$\rho^{(i)}_{A_{f_i}}(\mathbf{x}) = u^{(i)}_{A_i}(\mathbf{x}) u^{(i)*}_{A_f}(\mathbf{x})$$
$$= \frac{1}{(2\pi)^3} \int d\mathbf{q} \, e^{-i\mathbf{q}\cdot\mathbf{x}} S^{(i)}_{A_{f_i}}(\mathbf{q}),$$
$$\rho^{(\alpha)}_{B_{f_i}}(\mathbf{y}) = u^{(\alpha)}_{B_i}(\mathbf{y}) u^{(\alpha)}_{B_f}(\mathbf{y}) \quad (5.5)$$
$$= \frac{1}{(2\pi)^3} \int d\mathbf{q} \, e^{-i\mathbf{q}\cdot\mathbf{y}} S^{(\alpha)}_{B_{f_i}}(\mathbf{q}).$$

If $f = i$, the index f_i can be neglected. Here the $S^{(i)}_{A_{f_i}}(q), S^{(\alpha)}_{B_{f_i}}(q)$ are called the form factors of individual NOCS. Using Eqs. (2.3) and (5.5) one can obtain the profile function of individual NOCS with double scattering amplitude

$$\gamma^*_{i\alpha}(\mathbf{b}) = d\mathbf{x} \, d\mathbf{y} \, \rho^{(i)}_{A_{f_i}}(\mathbf{x}) \rho^{(\alpha)}_{B_{f_i}}(\mathbf{y}) \Gamma_{i\alpha}(\mathbf{b} + \mathbf{s} - \boldsymbol{\sigma})$$
$$= \frac{1}{2\pi i k'} \int d\mathbf{q} \, e^{-i\mathbf{q}\cdot\mathbf{b}} S^{(i)}_{A_{f_i}}(-\mathbf{q}) S^{(\alpha)}_{B_{f_i}}(\mathbf{q}) f_{i\alpha}(\mathbf{q}),$$
$$\gamma_{i\alpha}(\mathbf{b}) = d\mathbf{x} \, d\mathbf{y} \, \rho^{(i)}_A(\mathbf{x}) \rho^{(\alpha)}_B(\mathbf{y}) \Gamma_{i\alpha}(\mathbf{b} + \mathbf{s} - \boldsymbol{\sigma}) \quad (5.6)$$
$$= \frac{1}{2\pi i k'} \int d\mathbf{q} \, e^{-i\mathbf{q}\cdot\mathbf{b}} S^{(i)}_A(-\mathbf{q}) S^{(\alpha)}_B(\mathbf{q}) f_{i\alpha}(\mathbf{q}).$$

Substituting Eq. (5.6) into Eq. (5.4) we can obtain the S matrix of NOCS–NOCS as follows:

$$S_{v-v}(\mathbf{b}) = [\delta^{(1)}_{A_{f_i}} \delta^{(1)}_{B_{f_i}} - \gamma^*_{11}(\mathbf{b})] \prod_{(i,\alpha) \neq (1,1)} [1 - \gamma_{i\alpha}(\mathbf{b})]$$
$$= \frac{1}{AB} \sum_{i,\alpha} [\delta^{(i)}_{A_{f_i}} \delta^{(\alpha)}_{B_{f_i}} - \gamma^*_{i\alpha}(\mathbf{b})]$$
$$\times \prod_{(i',\alpha') \neq (i,\alpha)} [1 - \gamma_{i'\alpha'}(\mathbf{b})]. \quad (5.7)$$

As $\delta^{(i)}_{A_{f_i}} = 0$, there is a NOCS excitation in A *; and at $\delta^{(\alpha)}_{B_{f_i}} = 0$ there is a NOCS excitation in B *. With the extended approximation method of RPA, the excitation in systems equals the sum of single particle excitation amplitude in NOCS. For Eq. (5.7), it can be shown that if we know all kinds of form factors of NOCS, the S matrix of NOCS–NOCS can be obtained.

VI. THE S MATRIX OF NICS–NICS AND GEOMETRICAL MODEL

In Sec. II we have pointed out that the NICS consists of infinite particles. Therefore the NICS scattered by NICS can be regarded as the multiple scattering between two infinitesimal scattering objects, as shown in Fig. 2. In Ref. 10, we have pointed out that in the NICS–NICS interaction the excitation of NICS is neglected. Then, the elastic channel S matrix is given by

$$S_{s-s}(\mathbf{b}) = \int v^*_{A_i} v^*_{B_i} \Big[\prod_j \prod_k [1 - \Gamma_{jk}(\mathbf{b} - \mathbf{s}_j + \mathbf{s}_k)] \Big] \times v_{A_i} v_{B_i} \prod_i d\mathbf{x}_i \prod_k d\mathbf{y}_k$$
$$= \int v^*_{A_i} v^*_{B_i} \Big[\exp\Big(\frac{-i}{\hbar v} \sum_{jk} \int V_s(\mathbf{r} - \mathbf{x}_j + \mathbf{y}_k) dz \Big) \Big] \times v_{A_i} v_{B_i} \prod_i d\mathbf{x}_i \prod_k d\mathbf{y}_k . \quad (6.1)$$

Let ε^3 be the volume of each infinitesimal object. Here, use $\varepsilon^3 \xi_A(\mathbf{x}_j)$ to describe the matter of each infinitesimal object, subscript A indicating that it belongs to the composite particle system A. Similarly, $\varepsilon^3 \xi_B(\mathbf{y}_k)$ belongs to the composite particle system B. ξ_A and ξ_B are the density functions in A and B, respectively; Let $V_s(\mathbf{r} - \mathbf{x}_j + \mathbf{y}_k)$ be the interaction potential between the infinitesimal object $\varepsilon^3 \xi_A(\mathbf{x}_j)$ in X_j and the infinitesimal object $\varepsilon^3 \xi_B(\mathbf{y}_k)$ in $\mathbf{y}_k$. Then the relation between V and $\xi_A(\mathbf{X}_j)$, $\xi_B(\mathbf{y}_k)$ has the following expression:

$$V_s(\mathbf{r} - \mathbf{x}_j + \mathbf{y}_k) = \varepsilon^3 \xi_A(\mathbf{x}_j) \varepsilon^3 \xi_B(\mathbf{y}_k) g(\mathbf{r} - \mathbf{x}_j + \mathbf{y}_k). \quad (6.2)$$

Here $g(\mathbf{r} - \mathbf{x}_j + \mathbf{y}_k)$ is called the interaction function, so

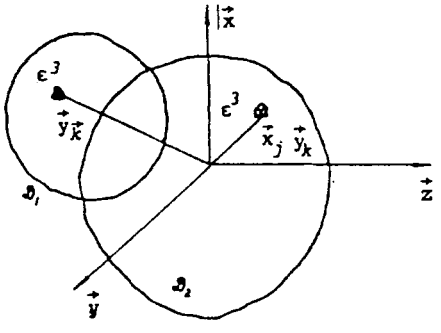


FIG. 2. Dividing two NICS into infinitesimal objects and the volume of each small unit is ε^3.

$$\lim_{j,k\to\infty} \sum_{j,k} \int dz\, V_s(\mathbf{r} - \mathbf{x}_j + \mathbf{y}_k)$$
$$= \lim_{j,k\to\infty} \sum_{j,k} \int dz\, \varepsilon^3 \xi_A(\mathbf{x}_j) \varepsilon^3 \xi_B(\mathbf{y}_k) g(\mathbf{r} - \mathbf{x}_j + \mathbf{y}_k)$$
$$= \int_{\mathscr{D}_A \mathscr{D}_B} d\mathbf{x}\, d\mathbf{y}\, dz\, \xi_A(\mathbf{x}) \xi_B(\mathbf{y}) g(\mathbf{r} - \mathbf{x} + \mathbf{y}). \quad (6.3)$$

The integral regions of $d\mathbf{x}$ and $d\mathbf{y}$ are $\mathscr{D}_A$ and $\mathscr{D}_B$, respectively, $|c_A, i\rangle$ and $|c_B, i\rangle$ are used to describe the collective motion state of A and B, respectively. So the S matrix is

$$S_{s-s}(\mathbf{b}) = \langle c_A, i | \langle c_B, i | \exp\Big[\frac{-i}{\hbar v} \int dz \int_{\mathscr{D}_A} d\mathbf{x} \int_{\mathscr{D}_B} d\mathbf{y} \times \xi(\mathbf{x})\, \xi(\mathbf{y}) g(\mathbf{r} - \mathbf{x} + \mathbf{y}) \Big] | c_A, i \rangle | c_B, i \rangle . \quad (6.4)$$

Let

$$\frac{i}{\hbar v} \int_{\mathscr{D}_A \mathscr{D}_B} dz\, d\mathbf{x}\, d\mathbf{y}\, \xi_A(\mathbf{x}) \xi_B(\mathbf{y}) g(\mathbf{r} - \mathbf{x} + \mathbf{y}) \equiv \Omega(\mathbf{b}, \alpha_{\lambda_A \mu_A}, \alpha_{\lambda_B \mu_B}), \quad (6.5)$$

which should be the function of the impact parameter $\mathbf{b}$ and $\alpha_{\lambda_A \mu_A}, \alpha_{\lambda_B \mu_B}$. From Eqs. (6.4) and (6.5) it can be seen that the S matrix of Eq. (6.4) is quite similar to that obtained from the geometry model in badron scattering. In Refs. 11 and 12, $\Omega_0(\mathbf{b})$ is the scalar, while $\Omega(\mathbf{b}, \alpha_{\lambda_A \mu_A} \alpha_{\lambda_B \mu_B})$ in Eq. (6.5) is the operator. Because the S matrix in Eq. (6.4) is only for the elastic channel, the contributions to Eq. (6.4) are made only by the scalar term $\Omega_0(b)$ and the square term of the $\alpha_{\lambda_A \mu_A}, \alpha_{\lambda_B \mu_B}$ in the expansion (6.5), and the higher-order terms can be neglected. Assume that $\mathscr{D}_A$ and $\mathscr{D}_B$ are the spherical spaces of A and B, i.e.,

$$\Omega_0(\mathbf{b}) = \frac{i}{\hbar v} \int_{\mathscr{D}_A \mathscr{D}_B} dz\, d\mathbf{x}\, d\mathbf{y}\, \xi_A(\mathbf{x}) \xi_B(\mathbf{y}) g(\mathbf{r} - \mathbf{x} + \mathbf{y}), \quad (6.6)$$

$$S_{s-s}(\mathbf{b}) = e^{-\Omega_0(\mathbf{b})}. \quad (6.7)$$

If we let $g(\mathbf{r} - \mathbf{x} + \mathbf{y})$ be $\delta(\mathbf{r} - \mathbf{x} + \mathbf{y})$, then

$$\Omega_0(\mathbf{b}) = \frac{i}{\hbar v} \int dz\, d\mathbf{x}\, \xi_A(\mathbf{x}) \xi_B(\mathbf{x} - \mathbf{r}). \quad (6.6')$$

Letting

$$\Big(\frac{i}{\hbar v} \Big)^{1/2} \int d\mathbf{x}_z\, \xi_A(\mathbf{x}) = T_A(\mathbf{s}_x), \quad (6.8)$$

have

$$\Omega_0(\mathbf{b}) = \int T_A(\mathbf{s}_x) T_B(\mathbf{s}_x - \mathbf{b}) d\mathbf{s}_x, \quad (6.9)$$

where $\mathbf{S}_x$ is the projection of the target nucleon coordinate $\mathbf{x}$ on the plane perpendicular to z; T_A is the thickness function of NICS; Eq. (6.9) is the convolution of the two thickness functions of NICS B and NICS A.

Here we assume that the interaction is the δ function and use the convolution to describe the S matrix. $\Omega_0(\mathbf{b})$ is a more complicated scalar which cannot excite the A state to

the B state, but Eq. (6.5) contains the phonon operator, i.e., $\Omega(\mathbf{b},\alpha_{\lambda_A\mu_A},\alpha_{\lambda_B\mu_B})$ is an operator, and it will operate on the phonon valcuum states. This can excite the A state to the B state.

The treatment for rotational nucleus is similar.

VII. DISCUSSIONS AND CONCLUSIONS

In the above sections, the S matrix of the two composite particle in high-energy scattering is discussed. According to Glauber's theory, the S matrix of the two composite particle system is divided into the S matrix of NOCS–NOCS, NOCS–NICS, NICS–NOCS, and NICS–NICS, and used various physically reasonable and effective approximations to treat different types of the S matrix. Here the RPA method is generalized, using the double scattering amplitude approximation to treat the S matrix of NOCS–NOCS. Therefore, the NOCS–NOCS amplitude can be calculated as long as various form factors of the NOCS are given. By the use of the sum limit method, the whole potential of NICS can be obtained, and Bohr's[1] collective coordinate approximation can be used to treat the S matrix of NOCS–NICS (NICS–NOCS). The "fold model" to describe the NICS–NICS scattering amplitude is introduced. This is corresponding to the forward scattering approximation or the case of δ force. By using the thickness function, the total phase shift can be calculated by folding the integral. The theoretical result agrees with the geometrical model, so that the collective character of NICS contains the geometrical character.

In the case of inelastic scattering, here we transfer the separate NICS into the continuous ocean, and the NICS excitation may be the phonon excitation. If the Bohr collective model is used to connect the NICS, then the S matrix of two composite particle systems scattering in high energy can be written as

$$S_{fi}(\mathbf{b}) = \sum_{\text{excite}} S_{v-v}(\mathbf{b})S_{v-s}(\mathbf{b})S_{s-v}(\mathbf{b})S_{s-s}(\mathbf{b}). \qquad (7.1)$$

The differential cross section can be written as

$$\frac{d\sigma_{fi}}{d\Omega} = \Theta^2(\mathbf{q})\frac{k^2}{4\pi^2}\left|\int d\mathbf{b}\, e^{i\mathbf{q}\cdot\mathbf{b}}(\delta_{fi} - S_{fi}(\mathbf{b}))\right|^2. \qquad (7.2)$$

The physically reasonable model should be chosen according to the character of collision systems, so that the differential cross section for every kind of mode can be calculated. Therefore, the influence of the single particle effect and collective in high energy scattering can be estimated.

It must be pointed out that in Eq. (7.1) if the scattering system does not have degree of freedom of NICS, then $S_{v-s} = S_{s-v} = S_{s-s} = 1$, the S_{v-v} is only left. If the scattering systems do not have degree of freedom of NOCS, the $S_{v-v} = S_{v-s} = S_{s-v} = 1$, the S_{s-s} is only left. Therefore, different physical considerations are needed in studying different scattering systems.

[1]A. Bohr and B. R. Mottelson, Dan. Vid. Selsk. Mat. Fys. Medd. **27**. (1953).
[2]R. J. Glauber, Lect. Theory Phys. **1**, 315 (1959).
[3]W. Czyz and L. C. Maximon, Ann. Phys. **52**, 59 (1969); V. Franco and G. K. Varam, Phys. Rev. C **18**, 349 (1978).
[4]T. Tamura, Rev. Mod. Phys. **37**, 679 (1965).
[5]B. Buck, Phys. Rev. **130**, 712 (1963).
[6]R. D. Amado, J. A. McNeil, and D. A. Sparrow, Phys. Rev. C **25**, 13 (1982).
[7]For a proof see A. Messiah, *Quantum Mechanics* (North-Holland, Amsterdam, 1965), Vol. 1, p. 442.
[8]M. L. Barlett *et al.*, Phys. Rev. C **22**, 1168 (1980).
[9]Li Yang-guo, Zhang Yu-shun, Chen Xiao-tian, and Ruan Tu-nan, Phys. Ener. Fort. Phys. Nucl. **6**, 715 (1980).
[10]Zhang Yu-shun, Li Yang-guo, Wang Wei-wei, and Ruan Tu-nan, Phys. Ener. Fort. Phys. Nucl. **6**, 55 (1982).
[11]Li Yang guo and Lo shui-yin, Auot. J. Phys. **37**, 255 (1984).
[12]T. T. Chou and C. N. Yang, Phys. Rev. **175**, 1832 (1968).
[13]D. J. Clarke and S. Y. Lo, Phys. Rev. D **10**, 1519 (1978).

On the Hamiltonian dynamics of vortex lattices

Kevin A. O'Neil
Department of Mathematical and Computer Sciences, The University of Tulsa, Tulsa, Oklahoma 74104

(Received 17 November 1988; accepted for publication 8 February 1989)

The equations of motion for infinite doubly periodic configurations of point vortices are derived. A Hamiltonian for the motion, generalizing that of Kirchhoff for finite configurations, is found and used to deduce an expression for the energy of an arbitrary vortex lattice. The energy of a lattice with periodic defects is computed. Some special stationary lattice configurations are shown to exist and integral curves for some two- and three-vortex lattice motions are exhibited.

I. INTRODUCTION

Two-dimensional point vortices can be used to model a variety of systems with pairwise logarithmic potentials, such as line charges and crystal screw dislocations, as well as to model 2-D fluid flow. Periodic configurations of vortices—vortex streets and lattices—find application in models with periodic or Dirichlet boundary conditions, as well as in models of rotating superfluids, type II superconductors, and 2-D wakes. Of particular interest are those lattices that minimize the lattice energy.

In the past, various methods have been devised to compute this energy or the 2-D Ewald potential more efficiently than by a slowly converging lattice sum. Some of these methods depend upon special lattice symmetries.[1-3] By direct integration of the energy density, Tkachenko[4] found the energy of a simple vortex lattice of arbitrary shape and showed that this energy is minimized for the triangular lattice. Recently, Campbell and others[5] derived an expression for the energy of arbitrary lattices containing more than one vortex per unit cell by generalizing a lattice summation technique of Glasser.[6] The energy is given in terms of rapidly converging infinite products.

In this paper the equations of motion for an arbitrary vortex lattice (including the limiting case of a vortex street) are derived and a new formula for the lattice energy is presented. The evolution equations, expressed in terms of Jacobi theta functions, are obtained by summing the individual vortex contributions over the lattice. Since the lattice sum does not converge absolutely, the order of summation must be specified. The nth partial sum includes the effects of all vortices within distance n of the origin. This convention produces the same result as imposing the requirement that a simple lattice rotate uniformly.

The result is a finite-dimensional dynamical system which has a Hamiltonian structure and is analogous to that of a finite-vortex configuration. This structure allows a compact expression for the energy of an arbitrary vortex lattice to be derived in a particularly simple way. The energy depends on the lattice shape and density, as well as on the vortex circulations and positions within the lattice.

In Sec. II the lattice dynamical equations are derived using the Weierstrass elliptic functions. Section III presents the Hamiltonian for these equations and the lattice energy for an arbitrary vortex lattice is computed. This formula gives the energy in terms of theta functions, which is convenient for both numerical and theoretical manipulations. As an example, the change in the lattice energy caused by the introduction of periodic defects is computed. Stationary vortex lattices are considered in Sec. IV and examples of lattice motions with only two or three vortices per unit cell appear in Sec. V.

II. LATTICE DYNAMICS

A lattice L of points in the complex plane generated by two independent vectors ω_1, ω_2 is the set $\{a\omega_1 + b\omega_2 | a,b \in \mathbf{Z}\}$. This set is doubly periodic; a singly periodic lattice is obtained by letting one generator go to infinity. The period parallelogram or unit cell of L centered at z_0 is the set $\{z_0 + s\omega_1 + t\omega_2 | s,t \in [-\frac{1}{2},\frac{1}{2})\}$. The lattice can be characterized (up to rotation) by the lattice parameter $\tau = \omega_2/\omega_1$ and the lattice density ρ, the reciprocal of the area of a unit cell of the lattice $|\mathrm{Im}(\overline{\omega}_1\omega_2)| = |\omega_1|^2|\mathrm{Im}(\tau)|$. A lattice of vortices of circulation Γ at z_0 means a vortex of circulation Γ at each point $z_0 + \omega$ for all $\omega \in L$. Conversely, given such a collection of vortices, the position z_0 is the well-defined modulo L. We think of the vortex lattice as the basic unit with which arbitrary doubly periodic configurations can be constructed.

Suppose vortex lattices of circulations Γ_j are placed at $z_j, j = 1,...,n$. For convenience, assume all positions z_j lie in the unit cell centered at the origin. The conjugate velocity of the fluid due to this vortex lattice configuration is the sum of the velocities induced by all vortices[7]:

$$\overline{V}(z) = \frac{1}{2\pi i} \sum_{\omega \in L} \sum_{j=1}^{n} \frac{\Gamma_j}{z - (z_j + \omega)}. \tag{2.1}$$

The velocity field (2.1) determines the velocity of the individual vortices since vorticity is a convected quantity. However, the sum over L is not absolutely convergent, that is, the sum depends on the order of summation. Therefore, some convention on the summation is needed.

As a physical model, we take the infinite-vortex lattice to be the limit of finite-vortex configurations consisting of those vortices within the distance R of the origin as R goes to infinity,[8,9] that is, the infinite lattice is the limit of finite-circular lattices. We will see that the shape of this limiting region has an effect on the resulting dynamics. Let the symbol Σ_R denote this limiting process:

$$\sum_R f(\omega) = \lim_{R \to \infty} \sum_{\omega \in L, |\omega| < R} f(\omega).$$

 0022-2488/89/061373-07$02.50

In this circular limit, the conjugate velocity is

$$\overline{V}(z) = \frac{1}{2\pi i}\sum \Gamma_j \sum_R \frac{1}{(z-z_j)-\omega}$$

$$= \frac{1}{2\pi i}\sum \Gamma_j F(z-z_j), \tag{2.2}$$

where $F(z) = \Sigma_R (z-\omega)^{-1}$. We can express F in terms of the Weierstrass zeta function

$$\zeta(z;\omega_1,\omega_2) = \frac{1}{z} + \sum_{0\neq\omega\in L}\left(\frac{1}{z-\omega}+\frac{1}{\omega}+\frac{z}{\omega^2}\right) \tag{2.3}$$

as follows:

$$F(z) = \zeta(z;\omega_1,\omega_2) + \alpha z, \quad \alpha = -\sum_R{}' \omega^{-2}. \tag{2.4}$$

(The prime indicates that the singular term is omitted; it is easy to show that the limit α always exists.) The function $\zeta_0(z)$ of Tkachenko[4] is the same as $F(z)$, as we shall see; we adopt this notation. The zeta function $\zeta(z) = \zeta(z;\omega_1,\omega_2)$ is an odd analytic function with simple poles of residue 1 at each lattice point. Furthermore, the zeta function is quasiperiodic: For all $z\in\mathbf{C}-L$ and all integers a, b, we have $\zeta(z+a\omega_1+b\omega_2)-\zeta(z) = a\eta_1+b\eta_2$ for certain constants η_1, η_2 called quasiperiods. Given a lattice point $\omega = a\omega_1+b\omega_2$, it is convenient to write $\eta(\omega)$ for $a\eta_1+b\eta_2$. Clearly, ζ_0 is also quasiperiodic, with the quasiperiods $\tilde{\eta}_i = \eta_i + \alpha\omega_i$. In fact, the quasiperiods of ζ_0 can be computed directly from the lattice sum.

Lemma: $\tilde{\eta}(\omega) = \pi\rho\bar{\omega}$.

Proof: By the Legendre relation $\eta_1\omega_2 - \eta_2\omega_1 = 2\pi i$, it suffices to show $\tilde{\eta}_1 = \pi\rho\bar{\omega}_1$. Since the relation is homogeneous in ω, we may rescale the lattice L and put $\omega_1 = 1$. For any $z\notin L$,

$$\tilde{\eta}_1 = \zeta_0(z+1) - \zeta_0(z)$$

$$= \lim_{R\to\infty}\left(\sum_{|\omega|<R}\frac{1}{z+1-\omega} - \sum_{|\omega|<R}\frac{1}{z-\omega}\right)$$

$$= \lim_{R\to\infty}\left(\sum_{\omega\in L\cap D_1}\frac{1}{z-\omega} - \sum_{\omega\in L\cap D_2}\frac{1}{z-\omega}\right), \tag{2.5}$$

where D_1, D_2 are the disks $|w+1|<R$, $|w|<R$, respectively. Notice that on every horizontal row of lattice points through the disks, each disk contains exactly one lattice point not included in the other. Label these points $\omega_{11},\ldots,\omega_{1r}$ (in D_1), $\omega_{21},\ldots,\omega_{2r}$ (in D_2). Of course, these ω_{ij}, as well as r, depend on R; the lattice points lie near the circle $|z| = R$ and a $R\to\infty$, $r/R\to 2\rho$.

Those lattice points belonging to $D_1\cap D_2$ generate terms that cancel in the sum, so that we may write

$$\tilde{\eta}_1 = \lim_{R\to\infty}\sum_{i=1}^{r(R)}\left(\frac{1}{z-\omega_{1i}} - \frac{1}{z-\omega_{2i}}\right). \tag{2.6}$$

Since z is fixed,

$$\tilde{\eta}_1 = \lim_{R\to\infty}\sum_{i=1}^{r(R)}\left(\frac{1}{\omega_{2i}} - \frac{1}{\omega_{1i}}\right)$$

$$= \rho\lim_{R\to\infty}\frac{1}{\rho R}\sum\left(\frac{\bar{\omega}_{2i}-\bar{\omega}_{1i}}{R}\right). \tag{2.7}$$

The limit (2.7) is evaluated as the line integral $\int x\,dy$ around the unit circle, yielding the area π. This completes the proof.

Now we can compute the conjugate velocity $\overline{V}_{j,\omega}$ of the vortex at $z_j+\omega$ by removing the flow due to that vortex from the conjugate velocity field $\overline{V}$ and evaluating at $z_j+\omega$,

$$\overline{V}_{j,\omega} = \lim_{z\to z_j+\omega}\left[\overline{V}(z) - \frac{1}{2\pi i}\frac{\Gamma_j}{z-(z_j+\omega)}\right]$$

$$= \frac{1}{2\pi i}\sum_{j\neq k}\Gamma_k\zeta_0(z_j+\omega-z_k)$$

$$+ \frac{\Gamma_j}{2\pi i}\lim_{z\to z_j+\omega}\left[\zeta_0(z-z_j) - \frac{1}{z-(z_j+\omega)}\right]$$

$$= \frac{1}{2\pi i}\sum_{j\neq k}\Gamma_k(\zeta_0(z_j-z_k)+\pi\rho\bar{\omega}) + \frac{1}{2\pi i}\Gamma_j\pi\rho\bar{\omega}$$

$$= \frac{1}{2\pi i}\sum_{j\neq k}\Gamma_k\zeta_0(z_j-z_k) + \frac{1}{2\pi i}\left(\sum\Gamma_r\right)\pi\rho\bar{\omega}. \tag{2.8}$$

The second term in (2.8) indicates an overall rotation, with the rate proportional to the sum of the circulations $\Sigma = \Sigma\Gamma_r$. Therefore, it is convenient to describe the system in a rotating coordinate system. Choose the frame rotating with rate $\rho\Sigma/2$ about the center of vorticity $(\Sigma\Gamma_r z_r)/\Sigma$; if Σ vanishes, then choose the nonrotating frame translating with the limiting velocity $-i(\rho/2)\Sigma\Gamma_r z_r$. The conjugate velocity of the vortex at $z_j+\omega$ becomes

$$\frac{1}{2\pi i}\sum_{j\neq k}\Gamma_k[\zeta_0(z_j-z_k) - \pi\rho(\bar{z}_j-\bar{z}_k)]; \tag{2.9}$$

notice that the velocity vanishes for the simple lattice ($n=1$). Since ω does not appear, the vortex velocities have period L in this rotating frame and the lattice structure is preserved. Moreover, (2.9) is well defined for vortex lattice positions modulo L. The foregoing discussion can be summarized as the following theorem.

Theorem 1: The motion of n point vortex lattices, taken as the limit of lattices in a disk and viewed from the appropriate rotating or translating coordinate system, defines a local dynamical system of n distinct particles on the torus $T = \mathbf{C}/L$, with the evolution equations

$$V_j = \dot{z}_j = \frac{i}{2\pi}\sum_{j\neq k}\Gamma_k(\bar{\zeta}_0(z_{jk}) - \pi\rho z_{jk}), \quad z_{jk} = z_j - z_k. \tag{2.10}$$

While Eqs. (2.10) are defined for vortex lattice positions $z_1,\ldots,z_n$ modulo L, the reference frame is not because (2.2) was based on z_j in the unit cell centered at the origin. The moment of vorticity $M = \Sigma\Gamma_j z_j$ is not well defined on T. However, since M is a constant of the motion there are only $(n-1)$ independent positions and velocities. Therefore, the configuration space of the vortex lattice system is $T^{n-1} - \Delta$, where the diagonal Δ is the set on which two or more vortex lattice positions coincide. Since the equations are first order, phase space is identical to configuration space.

The conjugate velocity field in the rotating frame

$$\overline{V}_{(r)}(z) = \frac{1}{2\pi i}\sum\Gamma_j(\zeta_0(z-z_j) - \pi\rho(\bar{z}-\bar{z}_j)) \tag{2.11}$$

has period L and is thus elliptic in z if the total circulation Σ

is zero. It should also be noted that the conjugate velocities are meromorphic functions of the positions when $\Sigma = 0$.

Remarks: Tkachenko[4] chose the function $\zeta_0(z)$ as the unique modification of the zeta function having the quasiperiods as shown in the lemma, which is equivalent to the rotation of a simple lattice. Goncharov[10] gives Eq. (2.4), implicitly summing over circular regions to find α, and attributes the same quasiperiods to the result. It is evident from the proof of the lemma that the quasiperiods of an infinite lattice obtained as the limit of finite lattices depend on the shape of the limiting region used. This shows the long-range nature of the intervortex interaction. For example, if a rectangular rather than circular region is used to take the limit, the quasiperiods vary with the shape of the rectangle; a square region gives the same result as a circular one. If the quasiperiods differ from the value given in the lemma, then τ is in general not constant and the lattice itself changes shape.

However, different summation conventions result in sums that differ by at most a linear function since the second derivative of (2.1) is absolutely convergent.

III. THE HAMILTONIAN AND THE TOTAL ENERGY

A Hamiltonian for the evolution equations in Sec. II is a real-valued function H which satisfies the equations (written in complex form)

$$i\Gamma_j \dot{z}_j = \left(\frac{\partial}{\partial x} + i\frac{\partial}{\partial y}\right)H =: 2\frac{\partial H}{\partial \bar{z}_j}, \quad j = 1,\ldots,n. \tag{3.1}$$

We can find a Hamiltonian for the vortex lattice system in terms of the Weierstrass sigma function $\sigma(z)$, which has a simple zero at each lattice point and satisfies the relation $\sigma'/\sigma = \zeta$. Writing $\sigma_0(z) = \sigma(z)\exp(\alpha z^2/2)$, it is easy to check that

$$H = -\frac{1}{4\pi}\sum_{j<k}\Gamma_j\Gamma_k\left[\ln|\sigma_0(z_{jk})|^2 - \pi\rho|z_{jk}|^2\right] \tag{3.2}$$

satisfies (3.1). This is a generalization of the Hamiltonian for finite-vortex configurations, $-(4\pi)^{-1}\Sigma\Gamma_j\Gamma_k \ln|z_{jk}|^2$.

For numerical work it is convenient to rewrite (3.2) in terms of the Jacobi theta function

$$\vartheta_1(z) = 2q^{1/4}[\sin \pi z - q^{2\cdot 1}\sin 3\pi z + q^{3\cdot 2}\sin 5\pi z - \cdots],$$
$$q = \exp(i\pi\tau), \quad \tau = \omega_2/\omega_1. \tag{3.3}$$

For every lattice we may choose generators such that the lattice parameter τ has the imaginary part $y \geqslant \frac{1}{2}$, from which the series (3.3) converges rapidly for z in the period parallelogram centered at the origin.

By using the relation

$$\sigma(z) = \omega_1[\vartheta_1(z/\omega_1)/\vartheta_1'(0)]\cdot\exp(\eta_1 z^2/2\omega_1) \tag{3.4}$$

one finds

$$H = -\frac{1}{2\pi}\sum_{j<k}\Gamma_j\Gamma_k\left[\ln\left|\omega_1\frac{\vartheta_1(z_{jk}/\omega_1)}{\vartheta_1'(0)}\right| - \left(\frac{\pi}{y}\right)\left(\mathrm{Im}\left(\frac{z_{jk}}{\omega_1}\right)\right)^2\right]. \tag{3.5}$$

Since the equations of motion depend only on the lattice positions modulo L, the Hamiltonian should also have this property. This is easily verified with the help of the transformation relations $\vartheta_1(z+1) = -\vartheta_1(z)$, $\vartheta_1(z+\tau) = -\vartheta_1(z)\exp(-i\pi(\tau+2z))$.

To find the vortex velocities (2.10) in terms of ϑ_1, apply (3.1):

$$V_j = \frac{1}{2\pi\bar{\omega}_1}\sum_{j\neq k}\Gamma_k\left[\left(\frac{\bar{\vartheta}_1'}{\bar{\vartheta}_1}\right)\left(\frac{z_{jk}}{\omega_1}\right) - \left(\frac{2\pi i}{y}\right)\mathrm{Im}\left(\frac{z_{jk}}{\omega_1}\right)\right]. \tag{3.6}$$

It is easy to calculate the quotient of the theta functions using (3.3). Notice that as $\tau \to i\infty$, (3.6) has as its limit the usual equations of motion for vortex streets.[9]

Given a configuration of vortex lattices, it is useful to know the kinetic energy of the fluid contained in a single period parallelogram. Since the energy in any neighborhood of a point vortex is infinite, we will fix a cutoff radius ϵ and express the total kinetic energy of the lattice in the rotating frame as the integral of the energy density $|V_{(r)}(z)|^2/2$ over a unit cell, excluding the disk of radius ϵ around each vortex. We proceed to find an asymptotic formula for this energy which is valid in the limit as the cutoff radius ϵ goes to zero. In the calculations that follow, the numbers δ, ϵ will be assumed small and terms of such order will be ignored.

First consider the case when the total circulation Σ vanishes ($n > 1$). The conjugate flow has the complex potential

$$\Phi(z) = \frac{1}{2\pi i}\sum\Gamma_j(\log\sigma_0(z - z_j) + \pi\rho\bar{z}_j z). \tag{3.7}$$

Let the region A be a period parallelogram, with the ϵ disks about the vortex positions deleted. The integral of the energy density over A is equal to $-(\frac{1}{2})\int_{\partial A}\mathrm{Im}\,\Phi d(\mathrm{Re}\,\Phi)$. Since the conjugate flow is elliptic, the potential is quasiperiodic; from the identity $\sigma_0(z+\omega)/\sigma_0(z) = \pm\exp(\pi\rho\bar{\omega}(z+\omega/2))$ it follows that

$$\Phi(z+\omega) = \Phi(z) + \rho\,\mathrm{Im}(\omega\bar{M}). \tag{3.8}$$

Hence the integrand $\mathrm{Im}\,\Phi d(\mathrm{Re}\,\Phi)$ is periodic and the integral around any period parallelogram vanishes. Evaluate over the remainder of ∂A and use the identity $\Sigma\Gamma_j\Gamma_k|z_{jk}|^2 = -2(|M|^2 - (\Sigma)\Sigma\Gamma_j|z_j|^2)$ to show that the energy E has the asymptotic form

$$E = H - \frac{1}{4\pi}\left(\sum\Gamma_j^2\right)\ln\epsilon. \tag{3.9}$$

If the total circulation is nonzero, the conjugate flow $\bar{V}_{(r)}$ does not have a potential, but can be approximated by potential flows in the following standard way. Consider the period parallelogram centered about M/Σ, or equivalently, let $M = 0$ and take the period parallelogram about the origin. In this region, distribute N^2 vortex lattices of circulation $-\Sigma/N^2$ to form a lattice L/N, with the generators ω_1/N, ω_2/N. The conjugate flow due to these added vortices is

$$-\left(\frac{\Sigma}{2\pi i N^2}\right)\sum_{\omega'\in L/N}\zeta_0(z-\omega') = -\left(\frac{\Sigma}{2\pi i N}\right)\zeta_0(Nz), \tag{3.10}$$

with the limiting form $-(\Sigma\rho/2i)\bar{z}$ away from the poles. [Here we have used the homogeneity relation $\zeta_0(z;\omega_1,\omega_2) = \lambda\zeta_0(\lambda z;\lambda\omega_1,\lambda\omega_2)$.] The resulting (neutral) configuration has a potential Φ_N and in the limit of large N

the conjugate flow uniformly approximates $\bar{V}_{(r)}$ on A_N, the complement of the disks of radius $\epsilon N^{-3/2}$ around the added vortices. Thus the energy E of the original configuration is

$$E = \lim_{N\to\infty} -\left(\frac{1}{2}\right)\int_{\partial(A\cap A_N)} \operatorname{Im}\Phi_N d(\operatorname{Re}\Phi_N). \tag{3.11}$$

A straightforward computation shows that E differs from (3.9) by integrals that do not depend on the positions z_j.

We conclude that as with finite-vortex configurations, H gives the interaction energy, or that part of the energy that depends on the vortex lattice positions. Hence the total kinetic energy differs from H by a function of the lattice generators ω_1, ω_2 and the cutoff ϵ: This difference can be found by direct integration for a simple lattice.[4] An easier method is to use the lattice symmetry and the Hamiltonian to find this energy through vortex rearrangements.

First consider a single vortex lattice, that is, a doubly periodic configuration with a single vortex in each period parallelogram. Let $E(\Gamma,l,\epsilon)$ be the energy of a unit cell of area l^2 containing a single vortex of circulation Γ and cutoff radius ϵ and let $E(1,1,\epsilon) = E_1 - (1/4\pi)\ln\epsilon$. The quantity E_1 depends only on τ. By examining the integral of the energy density, one sees that

$$E(\Gamma,l,\epsilon) = \Gamma^2 E(1,1,\epsilon/l); \tag{3.12}$$

thus it suffices to find E_1.

Place four vortex lattices of circulation $\frac{1}{4}$ with a unit cell of unit area in two different configurations: (i) All six intervortex distances z_{jk} are small compared to δ, but large compared to ϵ and (ii) the vortex positions form a lattice L' with half the periods of L, $(z_1, z_2, z_3, z_4) = (0, \omega_{1/2}, \omega_{2/2}, (\omega_1+\omega_2)/2)$. Let E_A, E_B be the energies of these two configurations, with cutoff radius ϵ. The difference $E_A - E_B$ is given by the change in the interaction energy $H_A - H_B$. The individual energies can be found by elementary considerations:

$$\begin{aligned} E_A &= E(1,1,\delta) + (\text{energy inside the } \delta \text{ disk}) \\ &= E_1 - \frac{1}{4\pi}\left(\frac{1}{4}\ln\epsilon + \sum_{j\neq k}\frac{1}{16}\ln|z_{jk}|\right), \qquad (3.13) \\ E_B &= 4E(\tfrac{1}{4},\tfrac{1}{2},\epsilon) = \tfrac{1}{4}[E_1 - (1/4\pi)\ln 2\epsilon]. \end{aligned}$$

Take the relation $E_A - E_B = H_A - H_B$ and note that the sums involving logarithms, which become singular as $\delta\to 0$, cancel (to order δ, ϵ). Then solve for E_1:

$$E_1 = \tfrac{4}{3}[-(1/16\pi)\ln 2 - H_B]. \tag{3.14}$$

The term H_B can be evaluated using various properties of the theta functions. Summing over all differences produces the logarithm of the quantity

$$\begin{aligned} &|\vartheta_1(\tfrac{1}{2})\vartheta_1(\tau/2)\vartheta_1((1+\tau)/2)|^2 \\ &\quad = |q^{-1/2}\vartheta_0(0)\vartheta_2(0)\vartheta_3(0)|^2 = \exp(\pi y)|\vartheta_1'(0)/\pi|^2. \end{aligned} \tag{3.15}$$

After some algebra, we find H_B expressed as follows:

$$H_B = (1/4\pi)(\tfrac{1}{2}\ln|\vartheta_1'(0)| - \tfrac{3}{4}\ln|\omega_1| + \tfrac{1}{4}\ln\pi). \tag{3.16}$$

Since the lattice density $\rho = 1$, we have $0 = \ln|\omega_1|^2 + \ln y$. Thus

$$E_1 = -(1/4\pi)(\tfrac{1}{3}\ln 2\pi + \tfrac{2}{3}\ln|\vartheta_1'(0)| + \tfrac{1}{2}\ln y). \tag{3.17}$$

Expression (3.17) agrees with that given by Tkachenko[4] for the energy of a single-vortex lattice (although a different definition of ϑ_1 is used).

Now it is easy to find, by a similar rearrangement argument, the energy for a period parallelogram containing several vortices. Given vortices of circulation Γ_j at positions z_j, move them into a δ disk at M/Σ. The change in energy is ΔH and the energy outside the disk is $(\Sigma\Gamma_j)^2 E(1,1,\delta)$.

Theorem 2: The energy E of a (unit area) unit cell of a vortex lattice configuration containing vortices of circulations Γ_j at positions $z_j, j = 1,\ldots,n$, with the lattice generators ω_1, ω_2, and the cutoff radius ϵ, is given by

$$E = H + \left(\sum\Gamma_i\right)^2 E_1 - \frac{1}{4\pi}\left(\sum\Gamma_i^2\right)\ln\epsilon. \tag{3.18}$$

Denoting the energy density formula of Campbell[5] by $\tilde{E}$, the following relation has been verified numerically:

$$\tilde{E} = (4\pi/n)E + (\Sigma\Gamma_i^2/2n)\ln(n\epsilon^2). \tag{3.19}$$

As an application of Theorem 2, we find the energy of a simple lattice with periodic defects. Fix a complex number z and an integer N. Take a simple lattice with period L and circulation Γ and move the vortex at $N\omega$ to $(N\omega + z)$ for each $\omega\in L$. For simplicity, assume that N is odd, so that the centers of vorticity of the original lattice and the defect lattice coincide. We shall find an expression $\Delta E_N(z)$ for the resulting change in the kinetic energy of the fluid in the N^2 unit cells of L which make up a unit cell of the new periodic configuration (in other words, the change in energy per defect). Since the energy change is from rearrangement, we may ignore the self-energy ($\ln\epsilon$) terms, change the scale by a factor $1/N$, and apply Theorem 2 to the resulting configuration of N^2 vortices, all on L/N except for the single lattice displaced by $\eta = z/N$. The change in energy for the configuration is

$$\begin{aligned} \Delta E_N(z) = &-\frac{\Gamma^2}{4\pi}\sum_{0\neq\omega\in L/N} \\ &\times\left[\ln\left|\frac{\sigma_0(\eta-\omega)}{\sigma_0(\omega)}\right|^2 - \pi\rho(|\eta-\omega|^2 - |\omega|^2)\right]. \end{aligned} \tag{3.20}$$

Consider the derivative with respect to $\bar{\eta}$:

$$\begin{aligned} &\frac{\partial}{\partial\bar{\eta}}\Delta E_N(z) \\ &\quad = -\frac{\Gamma^2}{4\pi}\sum[\bar{\zeta}_0(\eta-\omega) - \pi\rho(\eta-\omega)] \\ &\quad = -\frac{\Gamma^2}{4\pi}[N\bar{\zeta}_0(N\eta) - \bar{\zeta}_0(\eta) - (N^2-1)\pi\rho\eta]. \end{aligned} \tag{3.21}$$

Integrate, using the condition $\Delta E_N(0) = 0$, to find

$$\begin{aligned} \Delta E_N(z) = &-(\Gamma^2/4\pi)[\ln|\sigma_0(z)/N\sigma_0(z/N)|^2 \\ &-(1-N^{-2})\pi\rho|z|^2]. \end{aligned} \tag{3.22}$$

Equation (3.22) is an exact result. For large N, the change in energy per defect approaches the limiting value

$$\Delta E_\infty(z) = -(\Gamma^2/4\pi)[\ln|z^{-1}\sigma_0(z)|^2 - \pi\rho|z|^2], \tag{3.23}$$

which is a function with period L minus the interaction energy of the displaced vortex at z with a vortex at the origin. The

result (3.23) can be compared to a previous (computationally intensive) study of simple lattices with periodic vacancies.[11] It should also be noted that combinations of lattice defects of the above type are relevant to the study of the stability of vortex lattices.[12,13] For example, a simple calculation shows that the energy of small defects can be negative only if the lattice shape is such that $\rho^2 < |\alpha|^2$ holds.

IV. STATIONARY LATTICE CONFIGURATIONS

A stationary lattice configuration is one in which all vortex velocities vanish in the rotating reference frame, that is, the configuration is a relative equilibrium in the fixed frame. Finding such configurations corresponds to solving the system of equations $V_1 = \cdots = V_n = 0$, or by (3.1), finding the critical points of the Hamiltonian.

If the total circulation Σ vanishes, the vortex lattice velocities are given with respect to a moving reference frame: The critical points of H do not correspond to equilibria, but to configurations at rest in this frame. Because the rotation terms vanish, methods of algebraic geometry can be applied and the total number of configurations that are at rest in *some* moving frame can be counted.[14]

If, instead, all circulations are positive, a lower bound for the number of relative equilibria can be obtained by using an idea of Palmore.[15] The Hamiltonian goes to infinity on the diagonal Δ and all critical points of H lie on a compact subset of $T^{n-1} - \Delta$. If the Hamiltonian is a Morse function (all critical points are nondegenerate), then a lower bound for the number of these critical points and their indexes can be computed from the Betti numbers of $T^{n-1} - \Delta$.

A simpler task, which is the goal of this section, is to find a special class of relative equilibria analogous to the collinear configurations of celestial mechanics. Let a collinear vortex lattice configuration denote one in which all z_j lie on a edge or a diagonal of a unit cell, so that all differences z_{jk} are parallel to ω_1, ω_2 or $\omega_1 \pm \omega_2$.

Consider now a lattice L that has a line of symmetry; we may assume that the line is the real axis, so that L is invariant under complex conjugation: $L = \overline{L}$. There are two types of such lattices: rectangular (with ω_2/ω_1 pure imaginary) and rhombic (with $\omega_2 = \overline{\omega}_1$). Important special cases are the square lattice ($\omega_2/\omega_1 = i$) and the triangular lattice ($\omega_2/\omega_1 = (\pm 1 + i\sqrt{3})/2$).

Theorem 3: Let L have a line of symmetry and fix $n > 2$ positive circulations. Then there are at least $2(n-1)!$ collinear relative equilibrium configurations of n vortex lattices with these circulations. If L is square or triangular, there are at least $4(n-1)!$ collinear relative equilibria.

Proof: If $L = \overline{L}$, the definition of the vortex velocities as circular limits shows that all velocities are pure imaginary or real if all differences z_{jk} are real or pure imaginary, respectively. Thus the vortex velocities V_j will be perpendicular to the line containing the vortices for such a configuration, so that half of the (real) partial derivatives of H always vanish. Stationary collinear configurations are thus critical points of H restricted to the space of configurations lying along the line, $(S^1)^{n-1} - \Delta$. The diagonal divides this configuration space into components: one for each ordering of the vortices in the unit cell and $(n-1)!$ in all. If the circulations are positive, then $H \to \infty$ on the boundary of each component; thus each component contains at least one critical point (whether H is or is not a Morse function).

Thus we find at least $(n-1)!$ collinear configurations on each diagonal of a rhombic unit cell and on the two sides of the rectangular unit cell. Since a square lattice is both rectangular and rhombic and a triangular lattice is rhombic in two different ways, we obtain four lines containing collinear relative equilibria and the higher estimates apply in these cases. This completes the proof.

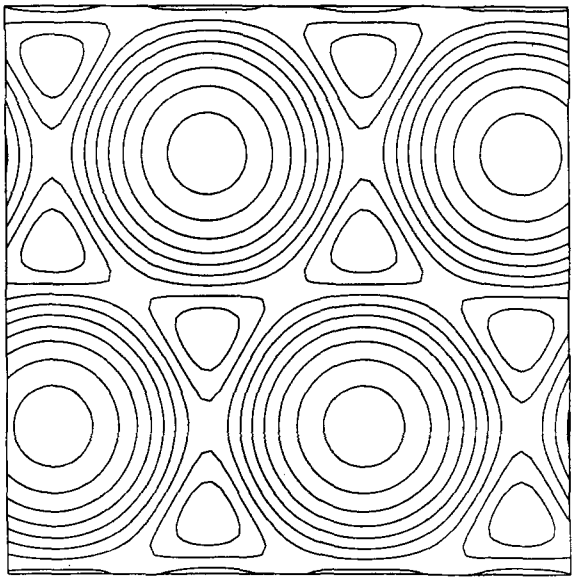


FIG. 1. Streamlines of a simple triangular lattice. The curves are interpolated from data on a 60×60 grid.

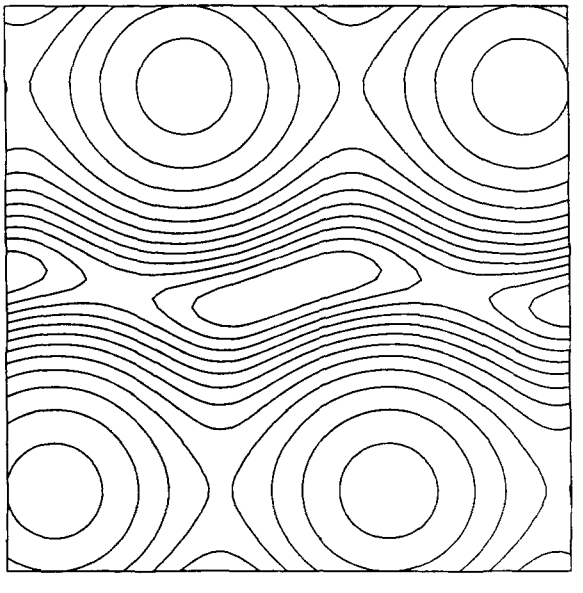


FIG. 2. Streamlines of a simple lattice, $\tau = 0.4 + 1.2i$.

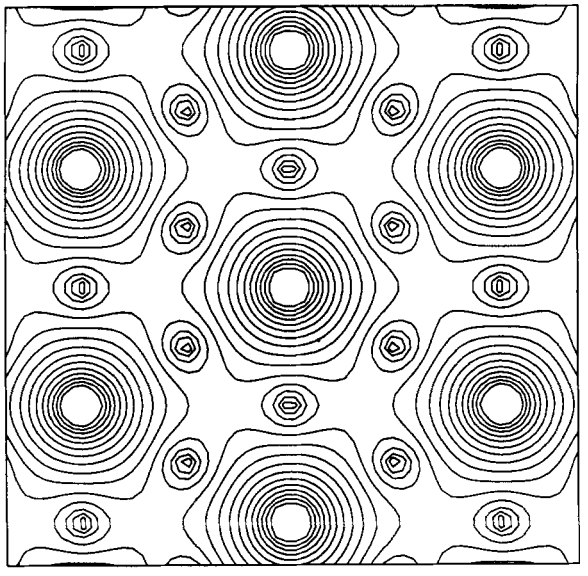


FIG. 3. The relative motion of three triangular lattices with zero net circulation. The diagram is centered around the origin and the curves trace out z_{13}. Here $(\Gamma_1,\Gamma_2,\Gamma_3) = (1,1,-2)$, $M = 0$.

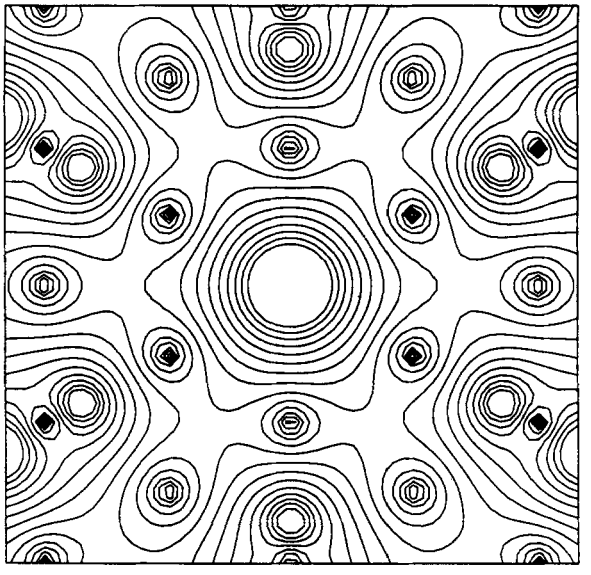


FIG. 4. Same as in Fig. 3, but with the circulations $(1,\sqrt{2},-1-\sqrt{2})$. Note the loss of periodicity.

V. LOW-DIMENSIONAL EXAMPLES

Recall that the differential equations (2.10) produce a flow on the configuration space $T^{n-1} - \Delta$, which has real dimension $2n-2$, and that the real-valued function H is conserved by the motion. Thus the level sets of H contain the integral curves of the dynamical system and have codimension one. When $n = 2$ and Σ is nonzero, these level sets will be the integral curves, independent of the circulations Γ_1, Γ_2 (since they are the streamlines of a simple lattice.)

Figures 1 and 2 show some level sets of H for $\tau = (1+i\sqrt{3})/2$ and $(0.4+1.2i)$, respectively. [In all the figures, the configurations $z_{12} = \omega_1/2$, $\omega_2/2$, and $(\omega_1+\omega_2)/2$ are seen to be relative equilibria because the function $\zeta_0(z) - \pi\rho\bar{z}$ is odd, has period L, and hence vanishes at these points.] Some lattices have two additional relative equilibrium configurations, which are minima of H, although configurations may not be minima of the total energy when variations in the lattice shape are considered. The morphology of energy minimizing configurations of two vortex lattices has been explored numerically.[5]

If the total circulation Σ is zero, the two-lattice case is uninteresting: All configurations translate uniformly, that is, $(z_1 - z_2)$ is a constant of the motion. However, the level sets of H can be used to trace out the relative motion of three vortex lattices. Writing the relative positions as $w_1 = z_1 - z_3$, $w_2 = z_2 - z_3$, we have $\Gamma_1 w_1 + \Gamma_2 w_2 = M = \text{const}$; thus there is only one independent coordinate for fixed M. Thus the level sets of H as a function of w_1 will give the trajectories of z_1 in the frame moving with z_3. Such curves for a triangular lattice and $M = 0$ are presented in Figs. 3 and 4 for two different choices of circulation: Here the critical points represent configurations that translate uniformly (w_1 and w_2 are constant). If the three vortex lattice positions are such that the differences z_{jk} are all half-periods, then the configuration will be at rest.

Because M is fixed, H as a function of w_1 is not in general periodic with period L. If the ratio of circulations Γ_1/Γ_2 is rational, then H will be periodic on some multiple of L; however, if the ratio is irrational H will be aperiodic. This can be seen in Figs. 3 and 4.

One can find uniformly translating configurations by solving the single equation $\zeta(w_1 - w_2) = \zeta(w_1) - \zeta(w_2)$, or in terms of the Weierstrass elliptic functions,

$$0 = (\wp' w_1 + \wp' w_2)/(\wp w_1 - \wp w_2).$$

If M is near zero, examination of the Laurent expansions

$$\wp(z) = z^{-2} + \cdots, \quad \wp'(z) = z^{-3} + \cdots$$

shows that there will be a pair of solutions near $w_1 = 0$ corresponding to the equilateral triangle configurations of three vortices which translate uniformly.[14]

ACKNOWLEDGMENT

I would like to thank L. J. Campbell for introducing me to this subject.

[1]L. J. Campbell and R. M. Ziff, "Vortex patterns and energies in a rotating superfluid," Phys. Rev. B **20**, 1886 (1979).
[2]C. E. Seyler, "Thermodynamics of two-dimensional plasmas or discrete line vortex fluids," Phys. Fluids **19**, 1336 (1976).
[3]L.F. Greengard, "The rapid evaluation of potential fields in particle systems," Ph.D. thesis, Yale University, 1987.
[4]V. K. Tkachenko, "On vortex lattices," Sov. Phys. JETP **22**, 1282 (1966).
[5]L. J. Campbell, "Vortex lattices in theory and practice," in *Proceedings of the SIAM Workshop on Mathematical Aspects of Vortex Dynamics*, Leesburg, VA, edited by R. Caflisch (Society for Industrial and Applied Mathematics, Philadelphia, 1989); M. M. Doria, J. B. Kadtke, L. J. Campbell, "Energy of infinite vortex lattices," to be published.
[6]M. L.Glasser, "The evaluation of lattice sums. III," J. Math. Phys. **15**, 188 (1974).
[7]N. E. Kochin, I. A. Kibel, and N. V. Roze, *Theoretical Hydromechanics* (Interscience, New York, 1964), p. 201.
[8]A. L. Fetter, "Quantum theory of superfluid vortices. I," Phys. Rev. **162**, 143 (1967).
[9]Reference 7, p. 214.
[10]V. P. Goncharov, "Dynamics of solitary dissipative vortices: vortex lattices and their stability," Sov. Phys. JETP **22**, 976 (1986).

[11]J. B. Kadtke, "Equilibria, lattices, and chaotic dynamics of point vortices," Ph.D. thesis, Brown University, 1987.
[12]A. L. Fetter, P. C. Hohenberg, and P. Pincus, "Stability of a lattice of superfluid vortices," Phys. Rev. **147**, 140 (1966).
[13]V. K. Tkachenko, "Stability of vortex lattices," Sov. Phys. JETP **23**, 1049 (1966).
[14]K. A. O'Neil, "Symmetric configurations of vortices," Phys. Lett. A **124**, 503 (1987).
[15]J. I. Palmore, "Relative equilibria of vortices in two dimensions," Proc. Nat. Acad. Sci. **79**, 716 (1982).

Nearest neighbor lattice statistics on semi-infinite two-dimensional rectangular lattices of various widths

J. M. Maeder and R. B. McQuistan
Department of Physics and Laboratory for Surface Studies, University of Wisconsin–Milwaukee, Milwaukee, Wisconsin 53211

(Received 26 January 1988; accepted for publication 25 January 1989)

Recent results of the shift operator matrix technique are utilized to calculate an exact expression for the lattice coverage of simple indistinguishable interacting molecules adsorbed on a semi-infinite two-dimensional $M \times \infty$ rectangular lattice. As an illustrated example, parameters are chosen which most likely model the adsorption of the noble gas Ne on a graphite surface. The lattice coverage is determined as a function of the activity, temperature, and interaction potentials for lattices of increasing width ($M = 1$–7).

I. INTRODUCTION

No chemical/physical system is without boundaries and so actual phase transitions occur in spaces of finite extent. It is clear from recent work that the results of the shift operator matrix (SOM) treatment of nearest-neighbor interacting particles on finite width lattice spaces can be used for certain systems to calculate successfully critical temperatures that agree quite well with experiment. Recently, George *et al.*,[1] using the results of previous work on a $2 \times N$ lattice,[2,3] have shown that even such narrow lattice spaces can be used to predict the critical temperatures of the two-dimensional phase transitions exhibited in the heat capacities of adsorbed noble gas submonolayer films. Thus questions arise as to the effect of increasing the width of such lattice spaces on the statistical/thermodynamic properties of a system, such as the calculated critical temperature for phase transitions of all kinds, including melting, heat capacities, etc., and on the sharpness of the "maxima" that occur near critical temperatures.

Previous investigations have shown that the SOM contains all the information necessary to treat the thermodynamics of a system of nearest-neighbor interacting, simple indistinguishable particles on $M \times N$ lattice spaces.[4] The useful result obtained is that the grand canonical partition function, in the thermodynamic limit ($N \to \infty$), can be determined completely from the largest eigenvalue of the SOM. In light of this, thermodynamic calculations on lattices wider than the $2 \times N$ lattice become feasible.

In the present paper, we shall calculate the coverage in terms of the largest eigenvalue of the SOM and examine the dependence of selected adsorption isotherms on the width of the lattice space, i.e., for increasing values of M. In particular, adsorption isotherms are calculated for various values of M for a system which has received considerable attention recently,[1] the noble gas Ne adsorbed on graphite basal planes. Here the adsorbed gas atoms are modeled as indistinguishable, simple particles that occupy a single lattice site, distributed on a $M \times N$ rectangular lattice space experiencing only nearest-neighbor interactions and interactions with the lattice. It is to be expected that as M is increased, the finite width of the lattice space should become less important in the sense that calculated statistical/thermodynamic quantities, such as lattice coverage and heat capacity signatures, are expected to approach their limiting behavior for two-dimensional lattices of infinite extent.

II. THE SHIFT OPERATOR MATRIX

To determine the SOM, we first take a $M \times N$ lattice and decompose it into 2^M lattice subspaces $a_i(N)$ ($i = 0$–M, with $m = 2^M - 1$) categorized by the state of occupation of the Nth column of the $M \times N$ lattice. Each site in the Nth column of the $M \times N$ lattice can exist in only one of two states of occupation, either the site is occupied by a particle or it is vacant. Thus we can think of each site in this column as a bit in the binary representation of a decimal number, being unity if the site is occupied and zero otherwise. Using this correspondence we define each $a_i(N)$ as a subspace with the Nth column in a configuration specified by the binary bit pattern of the decimal number i. By convention we will associate the least significant bit of i with the top site in the Nth column. The occupation of the sites in the remaining columns on each lattice subspace is not specified.

In order to simplify the matrices that we shall encounter later it will be convenient to distinguish between vertical nearest-neighbor pairs and horizontal nearest-neighbor pairs. Here $A_i[N,q,n_v,n_h]$ is the number of ways of arranging q particles on a $a_i(N)$ lattice subspace in such a way as to create n_v vertical nearest-neighbor pairs and n_h horizontal nearest-neighbor pairs. To illustrate the procedure for determining the SOM consider a $2 \times N$ lattice (see Fig. 1). By definition, $a_0(N)$ has both lattice sites of the Nth column unoccupied, $a_1(N)$ has only the top lattice site of the Nth column occupied, $a_2(N)$ has only the bottom lattice site of the Nth column occupied, and $a_3(N)$ has both sites of the Nth column occupied.

The system of coupled recursion relations which lead to the SOM is formed by considering how many particles, n_v and n_h pairs, are removed when one removes the Nth column of a particular lattice subspace. With these considerations, and the use of Fig. 1, we obtain the four coupled linear recursion equations:

 0022-2488/89/061380-05$02.50

$$A_0[N,q,n_v,n_h] = A_0[N-1,q,n_v,n_h] + A_1[N-1,q,n_v,n_h] + A_2[N-1,q,n_v,n_h] + A_3[N-1,q,n_v,n_h], \tag{1a}$$

$$A_1[N,q,n_v,n_h] = A_0[N-1,q-1,n_v,n_h] + A_1[N-1,q-1,n_v,n_h-1] + A_2[N-1,q-1,n_v,n_h] + A_3[N-1,q-1,n_v,n_h-1], \tag{1b}$$

$$A_2[N,q,n_v,n_h] = A_0[N-1,q-1,n_v,n_h] + A_1[N-1,q-1,n_v,n_h] + A_2[N-1,q-1,n_v,n_h-1] + A_3[N-1,q-1,n_v,n_h-1], \tag{1c}$$

$$A_3[N,q,n_v,n_h] = A_0[N-1,q-2,n_v-1,n_h] + A_1[N-1,q-2,n_v-1,n_h-1] + A_2[N-1,q-2,n_v-1,n_h-1] + A_3[N-1,q-2,n_v-1,n_h-2]. \tag{1d}$$

If we now use the set of shift operators $\mathbf{R} = (R,S,T,U)$ defined such that

$$R^r S^s T^t U^u A_i[N,q,n_v,n_h] = A_i[N-r,q-s,n_v-t,n_h-u],$$

Eqs. (1a)–(1d) can be written in matrix form

$$[\mathbf{P}(\mathbf{R}) - \mathbf{I}] \begin{pmatrix} A_0[N,q,n_v,n_h] \\ A_1[N,q,n_v,n_h] \\ A_2[N,q,n_v,n_h] \\ A_3[N,q,n_v,n_h] \end{pmatrix} = 0, \tag{2}$$

where

$$\mathbf{P}(\mathbf{R}) = \begin{pmatrix} R & R & R & R \\ RS & RSU & RS & RSU \\ RS & RS & RSU & RSU \\ RS^2T & RS^2TU & RS^2TU & RS^2TU^2 \end{pmatrix} \tag{3}$$

and $\mathbf{I}$ is the 4×4 identity matrix.

Examination of the manner in which $\mathbf{P}(\mathbf{R})$ is formed from the recursion relations Eqs. (1a)–(1d) show that each matrix element can be written as

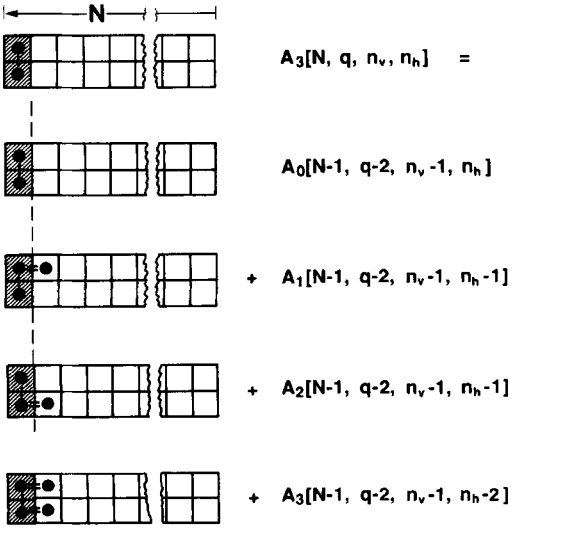


FIG. 1. The decomposition of the degeneracy A_3 of a $a_3(N)$ lattice in terms of the degeneracies A_0, A_1, A_2, A_3 of the $a_0(N-1)$, $a_1(N-1)$, $a_2(N-1)$, and $a_3(N-1)$ lattice formed by removing the Nth column of the original $a_3(N)$ lattice.

$$P_{ij}(\mathbf{R}) = RS^{s_i}T^{t_i}U^{u_{ij}}, \tag{4}$$

where s_i, t_i, and u_{ij} are the number of particles, vertical nearest-neighbor pairs, and horizontal nearest neighbor pairs that are removed when one starts out with a $a_i(N)$ lattice subspace and removes the Nth column to form a $a_i(N-1)$ lattice subspace. In terms of the row number i and column number j of the matrix $\mathbf{P}(\mathbf{R})$, s_i equals the number of ones in the binary representation of i, t_i equals the number of ones that are adjacent to each other in the binary representation of i, and u_{ij} equals the number of ones in the binary representations of i which match up bit by bit with the ones in the binary representation of j. Thus the functional form for each matrix element of $\mathbf{P}(\mathbf{R})$ can be determined completely from its row and column number. Since the way in which the elements $P_{ij}(\mathbf{R})$ are formed from the set of coupled recursion relations does not depend on the width of the lattice, Eq. (4) holds for a $M\times N$ lattice occupied by simple particles as prescribed above. It follows that the matrix equation for the $M\times N$ lattice analogous to Eq. (2) for a $2\times N$ lattice is given as

$$[\mathbf{P}(\mathbf{R}) - \mathbf{I}] \begin{pmatrix} A_0[N,q,n_v,n_h] \\ A_1[N,q,n_v,n_h] \\ \vdots \\ A_m[N,q,n_v,n_h] \end{pmatrix} = 0, \tag{5}$$

where the elements $P_{ij}(\mathbf{R})$ are given by Eq. (4) with i and j ranging from 0 to $m = 2^M - 1$, and $\mathbf{I}$ is now the $(m+1)\times(m+1)$ identity matrix.

III. DETERMINATION OF THE GRAND CANONICAL PARTITION FUNCTION

The grand canonical partition function is written

$$g(N,x,y,z) = \sum_{\{q,n_v,n_h\}} A[N,q,n_v,n_h] x^{n_h} y^{n_v} z^q, \tag{6}$$

where $A[N,q,n_v,n_h]$ is the total degeneracy factor for the complete $M\times N$ lattice. It is to be understood that we are working with some particular finite value of M in this and the following sections. The parameters q, n_v, and n_h range over all permissible values with

$$x = \exp[-\beta V_h], \tag{7a}$$

$$y = \exp[-\beta V_v], \tag{7b}$$

$$z = \exp[-\beta(V_0 - \mu)], \tag{7c}$$

in which μ is the chemical potential of the adsorbed parti-

cles, $\beta = 1/(k_B T)$ with k_B the Boltzmann constant and T the absolute temperature, V_v is the interaction energy between two vertical nearest-neighbor particle–particle pairs, V_h is the interaction energy between two horizontal nearest-neighbor particle–particle pairs, and V_0 is the interaction energy between a particle and the surface. To calculate $g(N,x,y,z)$ it will be convenient to define the "super grand canonical" generating function:

$$h(x,y,z,\eta) = \sum_{N=1}^{\infty} g(N,x,y,z)\eta^N \tag{8a}$$

$$= \sum_{\{N,q,n_v,n_h\}} A\,[N,q,n_v,n_h]\,x^{n_h}y^{n_v}z^q\eta^N. \tag{8b}$$

This function can be calculated explicitly, as will be shown subsequently. Once we obtain $h(x,y,z,\eta)$ we can determine $g(N,x,y,z)$, using Eq. (8a), by the simple relationship

$$g(N,x,y,z) = \frac{1}{N!}\frac{\partial^N h(x,y,z,\eta)}{\partial \eta^N}\bigg|_{\eta=0}. \tag{9}$$

For each $A_i\,[N,q,n_v,n_h]$ define generating functions:

$$h_i(x,y,z,\eta) = \sum_{\{N,q,n_v,n_h\}} A_i\,[N,q,n_v,n_h]\,x^{n_h}y^{n_v}z^q\eta^N. \tag{10}$$

If we now multiply both sides of Eq. (5) by $x^{n_h}y^{n_v}z^q\eta^N$ and sum over all permissible values of $\{N,q,n_v,n_h\}$, we obtain

$$[\mathbf{P}(R,S,T,U) - \mathbf{I}]\begin{pmatrix} h_0(x,y,z,\eta) \\ h_1(x,y,z,\eta) \\ \vdots \\ h_m(x,y,z,\eta) \end{pmatrix} \tag{11}$$

By rearranging summation indices, it is easy to show[5,6] that

$$[\mathbf{P}(x,y,z,\eta) - \mathbf{I}]\begin{pmatrix} h_0(x,y,z,\eta) \\ h_1(x,y,z,\eta) \\ \vdots \\ h_m(x,y,z,\eta) \end{pmatrix} = \begin{pmatrix} f_0(x,y,z,\eta) \\ f_1(x,y,z,\eta) \\ \vdots \\ f_m(x,y,z,\eta) \end{pmatrix}, \tag{12}$$

where $\mathbf{P}(x,y,z,\eta)$ is obtained from $\mathbf{P}(R,S,T,U)$ by the replacement of the shift operators R, S, T, and U by the variables η, z, y, and x, respectively. The functions $f_k(x,y,z,\eta)$ are polynomials that depend only on the initial conditions imposed on the $A_i\,[N,q,n_v,n_h]$. Applying Cramer's rule to Eq. (12), we have

$$h_l(x,y,z,\eta) = \frac{F_l(x,y,z,\eta)}{D(x,y,z,\eta)}, \tag{13}$$

where again $F_l(x,y,z,\eta)$ is a polynomial depending only on the initial conditions imposed on the $A_i\,[N,q,n_v,n_h]$, and $D(x,y,z,\eta)$ is the determinant of $[\mathbf{P}(x,y,z,\eta) - \mathbf{I}]$. By the way in which the $M \times N$ lattice is decomposed into the $a_i(N)$ lattice subspaces it is clear that the total degeneracy factor $A\,[N,q,n_v,n_h]$ for the $M \times N$ lattice is just the sum of the degeneracy factors for the individual lattice subspaces. With this is mind, Eqs. (10) and (13) show that $h(x,y,z,\eta)$ in Eq. (8) is now given by

$$h(x,y,z,\eta) = \frac{F(x,y,z,\eta)}{D(x,y,z,\eta)},$$
$$F(x,y,z,\eta) = \sum_{l=0}^{m} F_l(x,y,z,\eta). \tag{14}$$

Consider $F(x,y,z,\eta)$ and $D(x,y,z,\eta)$ in Eq. (14) as polynomials in η. If we perform a partial fraction decomposition[7] on $h(x,y,z,\eta)$ and use Eq. (9), we find that in the thermodynamic limit of large N, $g(N,x,y,z)$ is given asymptotically by[6,8]

$$g(N,x,y,z) = c(1/\eta_1)^N, \tag{15}$$

where η_1 is the smallest root of $D(x,y,z,\eta)$ which is considered as a polynomial in η and where c may be a function of η_1 but does not depend on N. It should be emphasized that we are only concerned with finite width lattices (M finite) at the moment. The thermodynamic limit envisioned concerns only the length of the lattice. When this model is used to calculate adsorption isotherms, the lattice is exposed to an essentially infinite reservoir of particles in the form of a gas phase above the lattice. For a macroscopically long lattice under these conditions it is entirely consistent to use the expression of Eq. (15). This calculational procedure is well known and covered extensively in the literature.[9–13] In this same limit, the logarithm of the grand canonical partition function becomes

$$\ln[g(N,x,y,z)] = N \ln(1/\eta_1). \tag{16}$$

At this point it is clear that we need only calculate the smallest root of $D(x,y,z,\eta)$, the determinant of the matrix $\mathbf{P}(x,y,z,\eta) - \mathbf{I}$.

Each element of the matrix $\mathbf{P}(x,y,z,\eta)$ contains a single factor of η, allowing us to write

$$\mathbf{P}(x,y,z,\eta) = \eta\mathbf{Q}(x,y,z), \tag{17}$$

where $\mathbf{Q}(x,y,z) = \mathbf{P}(x,y,z,\eta)/\eta$. Thus solving $\det[\mathbf{P}(x,y,z,\eta) - \mathbf{I}] = 0$, considered as a polynomial in η for the smallest root η_1, is equivalent to solving

$$\det[\mathbf{Q}(x,y,z) - \lambda_1\mathbf{I}] = 0 \tag{18}$$

for the largest eigenvalue $\lambda_1 = 1/\eta_1$ of the matrix $\mathbf{Q}(x,y,z)$. In terms of λ_1, Eq. (16) becomes

$$\ln[g(N,x,y,z)] = N \ln \lambda_1. \tag{19}$$

With this explicit expression for the logarithm of the grand canonical partition function, we can now calculate the expectation of the lattice coverage.

IV. EXPECTATION OF THE LATTICE COVERAGE

To determine the expectation of the lattice coverage $\langle\theta\rangle_{M\times N}$, we define

$$\langle\theta\rangle_{M\times N} = \langle q\rangle_{M\times N}/MN, \tag{20}$$

where

$$\langle q\rangle_{M\times N} = \frac{\Sigma_{\{N,q,n_v,n_h\}} qA\,[N,q,n_v,n_h]\,x^{n_h}y^{n_v}z^q\eta^N}{\Sigma_{\{N,q,n_v,n_h\}} A\,[N,q,n_v,n_h]\,x^{n_h}y^{n_v}z^q\eta^N}. \tag{21}$$

Thus

$$\langle\theta\rangle_{M\times N} = \frac{z}{MN}\frac{\partial}{\partial z}\ln[g(N,x,y,z)] \tag{22a}$$

$$= \frac{z}{M\lambda_1}\frac{\partial\lambda_1}{\partial z}. \tag{22b}$$

Since each element of the matrix $\mathbf{P}$ in Eq. (4) is strictly greater than zero, the Perron–Frobenius theorem[14] guarantees that the largest eigenvalue is nondegenerate. The derivative $\partial\lambda_1/\partial z$ can therefore be determined from the corresponding eigenvector. Consider the eigenvalue equation for λ_1

$$\mathbf{Q}\mathbf{v} = \lambda_1\mathbf{v}. \tag{23}$$

From Eq. (4) it follows that each element in any given row of $\mathbf{Q}(x,y,z)$ contains the same function of y and z, allowing us to write $\mathbf{Q}(x,y,z)$ as

$$\mathbf{Q}(x,y,z) = \mathbf{Q}_2(y,z)\mathbf{Q}_1(x), \tag{24}$$

where $(\mathbf{Q})_{ij} = y^{l_{ij}}z^{s_{ij}}\delta_{ij}$ and $(\mathbf{Q}_1)_{ij} = x^{u_{ij}}$ with δ_{ij} the Kronecker delta. If we now substitute Eq. (24) into Eq. (23) and use the renormalization

$$\mathbf{v}^T\mathbf{Q}_2^{-1}\mathbf{v} = 1 \tag{25}$$

we obtain

$$\mathbf{v}^T\mathbf{Q}_1\mathbf{v} = \lambda_1. \tag{26}$$

By differentiating Eqs. (25) and (26) with respect to z and inserting the identity matrix in the form $\mathbf{Q}_2\mathbf{Q}_2^{-1}$ or $\mathbf{Q}_2^{-1}\mathbf{Q}_2$ where appropriate we are led to the result

$$\frac{1}{\lambda_1}\frac{\partial\lambda_1}{\partial z} = -\mathbf{v}^T(\mathbf{Q}_2^{-1})'\mathbf{v}. \tag{27}$$

Combining this with Eq. (22b) gives us an expression for the coverage

$$\langle\theta\rangle_{M\times N} = z\,\mathbf{v}^T\mathbf{Q}_2^{-2}\mathbf{Q}_2'\mathbf{v}/M. \tag{28}$$

The form of Eq. (28) allows us to calculate $\langle\theta\rangle_{M\times N}$ from the eigenvector [with normalization defined by Eq. (25)] corresponding to the largest eigenvalue of the SOM $\mathbf{Q}(x,y,z)$ by standard numerical techniques,[15,16] the power method[17] being most suitable in this case.

We have chosen an interaction potential,[1,18,19] which most likely describes a real two-dimensional system and calculated numerically the coverage from Eq. (28) for values of M ranging from 1 to 7. The relevant interaction potentials for the rare gas Ne adsorbed on graphite basal planes measured in kelvins are $V_v/k_B = V_h/k_B = -34.6$ and $V_0/k_B = -378$. For fixed interaction energies the variables x and y remain constant along any particular adsorption isotherm leaving only z to vary. Under these conditions, z varies through the chemical potential μ, which in turn depends on the pressure p of the surrounding gas phase. In fact, for an ideal gas

$$\mu = k_B T\ln(p) + C(T), \tag{29}$$

where C is a function of T only. Therefore, in plotting $\langle\theta\rangle$ vs $\ln(z)$ at constant temperature as in Figs. 2–4, we are actually calculating adsorption isotherms of Ne adsorbed on graphite and observing how the coverage changes when the pressure of the gas phase is varied.

Figures 2 and 3 show the dependence of the adsorption isotherms on M for two opposite extremes of the kelvin scale. The transition from Fig. 2 to Fig. 3 through the temperature scale is found to occur gradually, where in all instances the adsorption isotherms approach their limiting value realized for an infinite two-dimensional lattice quite quickly (as indicated by the $M = 7$ curves). Figure 4 displays a set of adsorption isotherms for the rare gas Ne adsorbed on graphite modeled on a semi-infinite two-dimensional lattice seven sites wide. Figures 2 and 3 in turn indicate that this set of adsorption isotherms can be considered predictive of thermodynamic behavior for Ne adsorbed on graphite modeled on an infinite two-dimensional lattice.

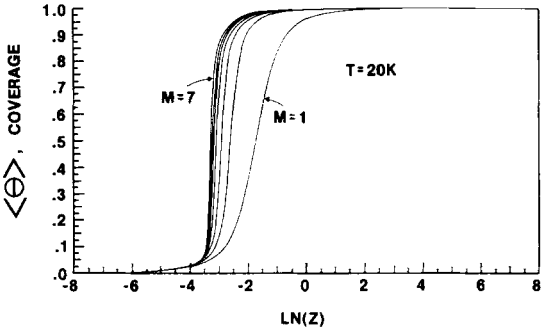


FIG. 3. The coverage as a function of $LN(z)$ under the conditions of Fig. 2 at the lower temperature of 20 K.

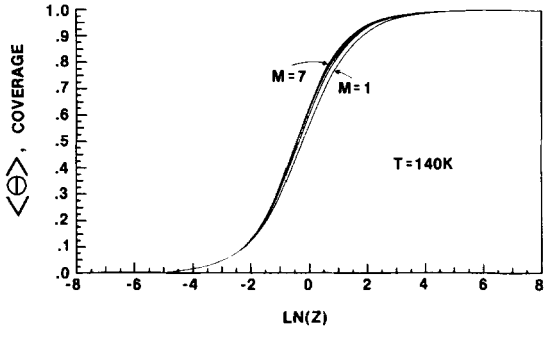


FIG. 2. The coverage for lattices ranging from one to seven sites wide as a function of $LN(z)$ with $V_v/k_B = V_h/k_B = -13.6$ at the temperature of 140 K.

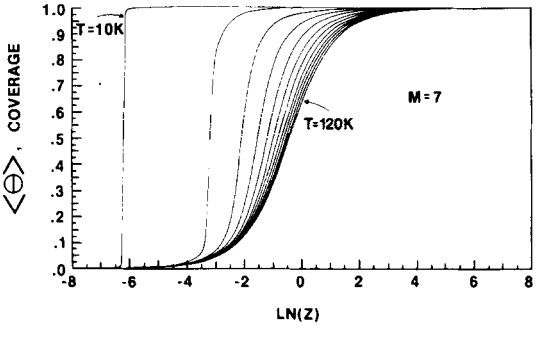


FIG. 4. The coverage as a function of $LN(z)$ over the temperature range of 10–120 K, in steps of 10 K, for a lattice seven sites wide.

V. CONCLUSION

A procedure has been presented for calculating the grand canonical partition function, starting from a set of coupled recursion relations derived from the lattice system of interest. Through the grand canonical partition function one then obtains an exact expression for the lattice coverage of an $M \times N$ rectangular lattice which depends on the width of the lattice. It is evident from the plots of the adsorption isotherms for Ne that the width of the lattice plays less of a role at high temperatures and an increasingly important role at very low temperatures.

ACKNOWLEDGMENT

One of the authors (R.B.M.) wishes to acknowledge the encouragement and support from the Air Force Office of Scientific Research under Grant No. AFOSR 88-0288.

[1]F. Battaglia, Y. S. Kim, and T. F. George, J. Phys. Chem. **91**, 414 (1987).
[2]R. B. McQuistan and J. L. Hock, J. Math. Phys. **25**, 261 (1984).
[3]J. L. Hock and R. B. McQuistan, J. Math. Phys. **26**, 2392 (1985).
[4]R. B. McQuistan and J. L. Hock, J. Math. Chem. **2**, 49 (1988).
[5]A. J. Phares, J. Math. Phys. **25**, 2169 (1984).
[6]J. Maeder, Master's thesis, University of Wisconsin–Milwaukee, 1987 (unpublished).
[7]A. Spitzbart, *Calculus with Analytic Geometry* (Scott, Foresman and Co., Glenview, IL, 1975).
[8]R. B. McQuistan and J. L. Hock, J. Math. Phys. **27**, 599 (1986).
[9]A. J. Phares, J. Math. Phys. **25**, 1756 (1984).
[10]J. L. Hock and R. B. McQuistan, J. Math. Phys. **26**, 2392 (1985).
[11]D. Walikainen and R. B. McQuistan, J. Math. Phys. **26**, 815 (1985).
[12]L. K. Runnels and L. L. Combs, J. Chem. Phys. **45**, 2482 (1966).
[13]T. L. Hill, *Statistical Mechanics* (McGraw-Hill, New York, 1956), Chap. 7.
[14]M. Marcus and H. Ming, *A Survey of Matrix Theory and Matrix Inequalities* (Allyn and Bacon, Boston, 1964).
[15]T. Yamamoto, Numer. Math. **8**, 324 (1966).
[16]I. S. Berezin and N. P. Zhidkov, *Computing Methods* (Pergamon, New York, 1965), Vol. 2.
[17]C. F. Gerald and P. O. Wheatley, *Applied Numerical Analysis* (Addison–Wesley, Reading, MA, 1984).
[18]S. Rauber, J. R. Klein, and M. W. Cole, Phys. Rev. B **27**, 1314 (1983).
[19]G. Vidali, M. W. Cole, and J. R. Klein, Phys. Rev. B **28**, 3064 (1983). References 18 and 19 are a good source of interatomic potentials for noble gases adsorbed on graphite.

Spectral properties of the Kronig–Penney Hamiltonian with a localized impurity

Silvestro Fassari
Department of Mathematics, and Center for Transport Theory and Mathematical Physics, Virginia Polytechnic Institute and State University, Blacksburg, Virginia 24061

(Received 1 March 1988; accepted for publication 15 February 1989)

It is shown that there exist bound states of the operator $H_{\pm\lambda} = -(d^2/dx^2) + \Sigma_{m\in Z}\delta(\cdot - (2m+1)\pi) \pm \lambda W$, W being an $L^1(-\infty, +\infty)$ non-negative function, in every sufficiently far gap of the spectrum of $H_0 = -d^2/dx^2 + \Sigma_{m\in Z}\delta(\cdot - (2m+1)\pi)$. Such an operator represents the Schrödinger Hamiltonian of a Kronig–Penney-type crystal with a localized impurity. The analyticity of the greatest (resp. lowest) eigenvalue of H_λ (resp. $H_{-\lambda}$) occurring in a spectral gap as a function of the coupling constant λ when W is assumed to have an exponential decay is also proven.

I. INTRODUCTION

In this paper we investigate some properties of the spectrum of the one-dimensional Schrödinger operator $H_{\pm\lambda} = H_0 \pm \lambda W$ with $\lambda > 0$,

$$H_0 = -\frac{d^2}{dx^2} + \sum_{m\in Z}\delta(\cdot - (2m+1)\pi)$$

and W being a non-negative L^1 function.

The Hamiltonian represents the Kronig–Penney model of a crystal with a localized impurity given by the short-range potential W.

There have been several papers[1–5] investigating the spectrum of the operator $H_{\pm\lambda}$ when H_0 is the Schrödinger Hamiltonian with a piecewise continuous periodic potential. In Ref. 6 the case of a nonperiodic potential V having a "short-range" order, so that $H_0 = d^2/dx^2 + V$ still has gaps in its spectrum, is studied. The main tool in our analysis will be the Birman–Schwinger kernel. Furthermore, we will exploit the Gel'fand expansion for the resolvent of H_0 (see Refs. 7 and 8) in order to have a convenient expression for the Birman–Schwinger kernel.

By doing so we show that there is a band in the spectrum of H_0 such that there exist eigenvalues of $H_{\pm\lambda}$ in each gap on the right of that band and if λ is sufficiently small we can find eigenvalues in each gap of $\sigma(H_0) = \sigma_{\rm ess}(H_{\pm\lambda})$.

Furthermore, we prove that under the stronger assumption of an exponential falloff of W, the greatest (resp. lowest) eigenvalue of $H_{+\lambda}$ (resp. $H_{-\lambda}$) occurring in a spectral gap is analytic as a function of the coupling constant λ.

The other important problem related to the asymptotics of the number of bound states in each gap will be studied in another paper.

II. BOUND STATES OF $H_0 \pm \lambda W$ IN THE GAPS OF $\sigma(H_0)$

In this section we shall be concerned with the existence of bound states of $H_0 \pm \lambda W$ inside the gaps of $\sigma(H_0)$.

First of all, let us recall that the spectrum of the unperturbed Hamiltonian H_0 is given by

$$\sigma(H_0) = \left(\bigcup_{k=0}^{\infty}[E_{2k+1}(0), (k+\tfrac{1}{2})^2]\right) \cup \left(\bigcup_{k=1}^{\infty}[E_{2k}(\pi), k^2]\right),$$

$E_n(\theta)$ being the nth root of the well-known Kronig–Penney equation

$$\cos 2\pi\sqrt{E} + (1/2\sqrt{E})\sin 2\pi\sqrt{E} = \cos\theta \qquad (2.1)$$

with $\theta\in[0,\pi]$ (see Ref. 9).

For each fixed θ, $\{E_n(\theta)\}_{n=1}^{\infty}$ are the eigenvalues of the reduced Hamiltonian $H_0(\theta) = (-d^2/dx^2)_\theta + \delta(\cdot - \pi)$ whose eigenfunctions are given by

$$\phi_n^{(\theta)}(x) = A_n^{(\theta)}\begin{cases}\cos\sqrt{E_n(\theta)}\,x \\ -((1-e^{i\theta}/1+e^{i\theta}))\cot\sqrt{E_n(\theta)}\,\pi\sin\sqrt{E_n(\theta)}\,x, & 0\leqslant x\leqslant\pi, \\ e^{i\theta}\left[\cos\sqrt{E_n(\theta)}\,(x-2\pi)\right. \\ \left.-((1-e^{i\theta}/1+e^{i\theta}))\cot\sqrt{E_n(\theta)}\,\pi\sin\sqrt{E_n(\theta)}\,(x-2\pi)\right], & \pi\leqslant x\leqslant 2\pi,\end{cases} \qquad (2.2)$$

$A_n^{(\theta)}$ being the normalization constant.

In the particular cases when $\theta = 0$, $\theta = \pi$ the eigenfunctions can be written as

$$\phi_{2k+1}^{(0)}(x) = \left(\pi + \frac{\sin 2\pi\sqrt{E_{2k+1}(0)}}{2\sqrt{E_{2k+1}(0)}}\right)^{-1}\begin{cases}\cos\sqrt{E_{2k+1}(0)}\,x, & 0\leqslant x\leqslant\pi, \\ \cos\sqrt{E_{2k+1}(0)}\,(x-2\pi), & \pi\leqslant x\leqslant 2\pi,\end{cases} \qquad (2.3)$$

 0022-2488/89/061385-08$02.50

$$\phi_{2k}^{(0)}(x) = (1/\sqrt{2\pi}) \sin kx \tag{2.4}$$

and

$$\phi_{2k+1}^{(\pi)}(x) = (1/\sqrt{2\pi}) \cos(k + \tfrac{1}{2})x, \tag{2.5}$$

$$\phi_{2k}^{(\pi)}(x) = \left(\pi - \frac{\sin 2\pi\sqrt{E_{2k}(\pi)}}{2\sqrt{E_{2k}(\pi)}}\right)^{-1} \times \begin{cases} \sin\sqrt{E_{2k}(\pi)}x, & 0 \leqslant x \leqslant \pi, \\ \sin\sqrt{E_{2k}(\pi)}(2\pi - x), & \pi \leqslant x \leqslant 2\pi. \end{cases} \tag{2.6}$$

By means of some boring algebra we can determine the normalization constant $A_n^{(\theta)}$, precisely

$$A_n^{(\theta)} = \left[\pi + \frac{\sin 2\pi\sqrt{E_n(\theta)}}{2\sqrt{E_n(\theta)}} + \left(\frac{1-\cos\theta}{1+\cos\theta}\right) \times \cot^2 \pi\sqrt{E_n(\theta)}\left(\frac{\sin 2\pi\sqrt{E_n(\theta)}}{2\sqrt{E_n(\theta)}}\right)\right]^{-1/2}. \tag{2.7}$$

This immediately leads to the following result whose proof is omitted since it only consists of tedious calculations.

Lemma 2.1: Let $\phi_n^{(\theta)}$ be the nth eigenfunction of the reduced Hamiltonian $H_0(\theta) = (-d^2/dx^2)_\theta + \delta(\cdot - \pi)$ acting on $L^2[0,2\pi]$ with $\theta \in [0,\pi]$. Then the following estimate holds:

$$\|\phi_n^{(\theta)}\|_\infty \leqslant \frac{1}{\sqrt{\pi}}\left[1 - \left(\frac{\sin 2\pi\sqrt{E_1(0)}}{2\pi\sqrt{E_1(0)}}\right)^2\right]^{-1/2} \tag{2.8}$$

for any $n \in N$ and any $\theta \in [0,\pi]$.

Remark: Since $\phi_n^{(2\pi-\theta)} = \overline{\phi_n^{(\theta)}}$, because of the antiunitarity of $H_0(\theta)$ and $H_0(\theta - 2\pi)$, the estimate (2.8) actually holds for any $\theta \in [0,2\pi]$. We shall use this property later.

After these preliminaries we consider the Birman–Schwinger operator in our particular case. Since W is a definite-sign function our Birman–Schwinger operator is self-adjoint. It is not difficult to show that $W^{1/2}(H_0 - E)^{-1}W^{1/2}$ ($W \geqslant 0$) is trace class for any $E \in \rho(H_0)$. One can first prove that this holds for $E < 0$ since

$$\|W^{1/2}(H_0 - E)^{-1}W^{1/2}\|_1 \leqslant \|W^{1/2}\left(-\frac{d^2}{dx^2} - E\right)^{-1}W^{1/2}\|_1 \tag{2.9}$$

for any $E < 0$.

Then, by using the first resolvent equation and the connectedness of the resolvent set of H_0, it follows that the property holds for any $E \in \rho(H_0)$.

This implies that we are allowed to use the KLMN theorem (see Refs. 10 and 11) in order to define the self-adjoint operator whose quadratic form is given by

$$(\psi, (H_0 \pm \lambda W)\psi) = (\psi, H_0\psi) \pm \lambda(\psi, W\psi)$$

for any $\psi \in Q(H_0)$.

Furthermore, we can apply the Fredholm theory to our Birman-Schwinger kernel. In particular, we can see that if $W \geqslant 0$, E is an eigenvalue of $H_0 \pm \lambda W$ if and only if ∓ 1 is an eigenvalue of $\lambda W^{1/2}(H_0 - E)^{-1}W^{1/2}$.

In order to obtain some information about the eigenvalues of $W^{1/2}(H_0 - E)^{-1}W^{1/2}$ when E belongs to some gap of $\sigma(H_0)$, we express the BS kernel by means of the Gel'fand transform, namely,

$$W^{1/2}(H_0 - E)^{-1}W^{1/2} = \sum_{n=1}^{\infty} \int_0^{2\pi} \frac{|W^{1/2}\phi_n^{(\theta)}\rangle\langle W^{1/2}\phi_n^{(\theta)}|}{E_n(\theta) - E} \frac{d\theta}{2\pi} \tag{2.10}$$

($\phi_n^{(\theta)}$ having been extended to the whole real axis by means of the θ condition), which makes sense for any $E \in \rho(H_0)$ because of Lemma 2.1 and the fact that $W \in L^1(R)$ as written in the Introduction.

If $E \in (-\infty, E_1(0))$, $W^{1/2}(H_0 - E)^{-1}W^{1/2}$ is a positive operator which implies that only $H_0 - \lambda W$ can have bound states lying in $(-\infty, E_1(0))$.

If E lies inside a gap the series on the rhs of (2.10) gives rise to a negative term in the expectation value of the BS kernel with respect to ψ due to the integrals related to the bands on the left of the gap containing E. For simplicity, let us only consider the case related to $H_0 - \lambda W$.

For our specific purpose we can neglect the negative term related to

$$W^{1/2}(H_0 - E)_-^{-1}W^{1/2} = -\sum_{n=1}^{2N} \int_0^{2\pi} \frac{|W^{1/2}\phi_n^{(\theta)}\rangle\langle W^{1/2}\phi_n^{(\theta)}|}{E - E_n(\theta)} \frac{d\theta}{2\pi} \tag{2.11}$$

since it can only lower the eigenvalues of the positive operator

$$W^{1/2}(H_0 - E)_+^{-1}W^{1/2} \tag{2.12}$$

similarly defined.

At this point we can start our analysis about the bound states of $H_0 - \lambda W$. First of all, we show the existence of bound states of $H_0 - \lambda W$ in every sufficiently far gap of $\sigma(H_0)$.

Theorem 2.2: Let

$$H_0 = -\frac{d^2}{dx^2} + \sum_{m=-\infty}^{+\infty} \delta(\cdot - (2m+1)\pi)$$

and W be a positive function belonging to $L^1(R)$. Then, for any fixed $\lambda > 0$, there exists a certain band of $\sigma(H_0)$ such that every gap on the right of that band contains eigenvalues of $H_\lambda = H_0 - \lambda W$.

Proof: First of all, we notice that the integral related to the $(2N+1)$st band in the expression of $W^{1/2}(H_0 - E)_+^{-1}W^{1/2}$ diverges when $E \to E_{2N+1}(0)_-$ on the one-dimensional subspace $\{W^{1/2}\phi_{2N+1}^{(0)}\}$.

Therefore the norm of $W^{1/2}(H_0 - E)_+^{-1}W^{1/2}$ increases without limit when $E \to E_{2N+1}(0)_-$.

Since this operator is compact and positive its greatest eigenvalue is equal to the norm of the operator. If $\lambda\|W^{1/2}(H_0 - N^2)_+^{-1}W^{1/2}\| < 1$, there must be an $\tilde{E} \in (N^2, E_{2N+1}(0))$ such that $\lambda W^{1/2}(H_0 - \tilde{E})^{-1}W^{1/2}\psi = \psi$, for some $\psi \in L^2(R)$. Of course, a completely similar analysis can be carried out when $E \in ((N+\frac{1}{2})^2, E_{2(N+1)}(\pi))$ for any fixed N. Therefore we need to show that $\|W^{1/2}(H_0 - N^2)_+^{-1}W^{1/2}\| \to 0$ when $N \to +\infty$.

For any ψ we have

$$(\psi, W^{1/2}(H_0 - N^2)_+^{-1} W^{1/2}\psi)$$
$$= \sum_{n=2N+1}^{\infty} \int_0^{2\pi} \frac{|(W^{1/2}\phi_n^{(\theta)}, \psi)|^2}{E_n(\theta) - N^2} \frac{d\theta}{2\pi}$$
$$\leqslant \frac{\|W\|_1 \|\psi\|_2^2}{2\pi^2} \left[1 - \left(\frac{\sin 2\pi\sqrt{E_1(0)}}{2\pi\sqrt{E_1(0)}}\right)^2\right]^{-1}$$
$$\times \left(\sum_{n=2N+1}^{\infty} \int_0^{2\pi} \frac{1}{E_n(\theta) - N^2} d\theta\right). \tag{2.13}$$

At this point we must prove that the series on the rhs of (2.13) goes to zero as N goes to infinity. First of all, let us consider

$$\sum_{n=2(N+1)}^{\infty} \int_0^{2\pi} \frac{1}{E_n(\theta) - N^2} d\theta. \tag{2.14}$$

The term with $n = 2N + 1$ will be considered later.

We can bound each integral in the series by replacing $E_n(\theta)$ by means of $E_{2k}(0)$ if $n = 2k + 1$ or $E_{2k-1}(\pi)$ if $n = 2k$.

Hence

$$(2.14) \leqslant 2\pi \left(\sum_{n=0}^{\infty} \frac{1}{(n + \frac{1}{2})(2N + n + \frac{1}{2})} + \sum_{n=0}^{\infty} \frac{1}{(n+1)(2N+n+1)}\right). \tag{2.15}$$

Since for any fixed N

$$\left\{\frac{1}{(n + \frac{1}{2})(2N + n + \frac{1}{2})}\right\}_{n=0}^{\infty}$$

and

$$\left\{\frac{1}{(n+1)(2N+n+1)}\right\}_{n=0}^{\infty}$$

are l_1 sequences dominated by

$$\left\{\frac{1}{(n + \frac{1}{2})^2}\right\}_{n=0}^{\infty} \in l_1,$$

by means of the dominated convergence theorem we get that the rhs of (2.15) goes to zero as $N \to \infty$.

In order to complete the proof of the theorem we only need to prove that

$$\int_0^{2\pi} \frac{1}{E_{2N+1}(\theta) - N^2} d\theta = 2\int_0^{\pi} \frac{1}{E_{2N+1}(\theta) - N^2} d\theta \to 0 \tag{2.16}$$

as $N \to \infty$.

For any $0 < \epsilon < \pi$ we have

$$\int_0^{\epsilon} \frac{1}{E_{2N+1}(\theta) - N^2} d\theta + \int_{\epsilon}^{\pi} \frac{1}{E_{2N+1}(\theta) - N^2} d\theta \leqslant \frac{\epsilon}{E_{2N+1}(0) - N^2} + \frac{\pi - \epsilon}{E_{2N+1}(\epsilon) - N^2} \tag{2.17}$$

since $E_{2N+1}(\theta)$ is an increasing function of θ (see Ref. 12).

By applying Taylor's theorem with remainder to (2.1) near $\sqrt{E} = \sqrt{E_{2N+1}(0)}$ we get

$$\cos 2\pi\sqrt{E_{2N+1}(0)} + \frac{1}{2\sqrt{E_{2N+1}(0)}} \sin 2\pi\sqrt{E_{2N+1}(0)} - \left[\left(2\pi + \frac{1}{2\tilde{E}_{2N+1}}\right)\sin 2\pi\sqrt{\tilde{E}_{2N+1}} - \frac{\pi}{\sqrt{\tilde{E}_{2N+1}}} \cos 2\pi\sqrt{\tilde{E}_{2N+1}}\right]$$
$$\times \left(\sqrt{E_{2N+1}(\theta)} - \sqrt{E_{2N+1}(0)}\right) = \cos\theta \tag{2.18}$$

[with $\tilde{E}_{2N+1} \in (E_{2N+1}(0), E_{2N+1}(\theta))$], which implies that for $\theta = \epsilon$

$$\sqrt{E_{2N+1}(\epsilon)} - \sqrt{E_{2N+1}(0)} \geqslant \frac{1 - \cos\epsilon}{2\pi + \pi/\sqrt{E_{2N+1}(0)} + 1/E_{2N+1}(0)} \geqslant \frac{1}{4\pi}(1 - \cos\epsilon) \tag{2.19}$$

for N large, since

$$\cos 2\pi\sqrt{E_{2N+1}(0)} + [2E_{2N+1}(0)]^{-1} \times \sin 2\pi\sqrt{E_{2N+1}(0)} = 1.$$

Thus we obtain

$$([E_{2N+1}(\epsilon) - E_{2N+1}(0)] + E_{2N+1}(0) - N^2)^{-1} \leqslant ([E_{2N+1}(0)/2\pi](1 - \cos\epsilon) + [E_{2N+1}(0) - N^2])^{-1}, \tag{2.20}$$

which implies

$$(\pi - \epsilon)/[E_{2N+1}(\epsilon) - N^2] \to 0 \tag{2.21}$$

as $N \to \infty$ for any $0 < \epsilon < \pi$ since $\lim_{N\to\infty} [E_{2N+1}(0) - N^2] = 1/\pi$ (see Ref. 12).

Therefore, for any $0 < \epsilon < \pi$, we have

$$\lim_{N\to\infty} \int_0^{\pi} \frac{1}{E_{2N+1}(\theta) - N^2} d\theta \leqslant \lim_{N\to\infty} \frac{\epsilon}{E_{2N+1}(0) - N^2} = \pi\epsilon \tag{2.22}$$

which gives (2.16). Q.E.D.

As a consequence of Theorem 2.2, we have the following corollary.

Corollary 2.3: If

$$\lambda < \frac{\pi[1 - (\sin 2\pi\sqrt{E_1(0)}/2\pi\sqrt{E_1(0)})^2]}{\|W\|_1[(\inf_{1<n<\infty}[a_{n+1} - b_n])^{-1} + 2\Sigma_{n=0}^{\infty}((n + \frac{1}{2})^2)^{-1}]} \tag{2.23}$$

[(b_n, a_{n+1}) being the nth gap of $\sigma(H_0)$] and with the same assumptions of Theorem 2.2, $H_0 - \lambda W$ has eigenvalues in every gap of $\sigma(H_0)$ (the inf is different from zero since there are no connected bands and $a_{n+1} - b_n$ converges to π^{-1} as follows from the analysis in Ref. 12).

Proof: From (2.13) and (2.15) we get

$$\lambda \| W^{1/2}(H_0 - N^2)_+^{-1} W^{1/2} \|$$

$$\leqslant \lambda \frac{\|W\|_1}{\pi} \left[1 - \left(\frac{\sin 2\pi \sqrt{E_{2N+1}(0)}}{2\pi \sqrt{E_{2N+1}(0)}} \right)^2 \right]^{-1}$$

$$\times \left[\frac{1}{E_{2N+1}(0) - N^2} + 2 \sum_{n=0}^{\infty} \frac{1}{(n+\frac{1}{2})^2} \right], \quad (2.24)$$

which is less than one for any N since λ satisfies (2.23). By means of similar estimates we obtain the same result for the gaps of the type $((N-\frac{1}{2})^2, E_{2N}(\pi))$. Q.E.D.

Of course, completely similar results hold in the case of $H_0 + \lambda W$. At this point we are going to investigate the behavior of the eigenvalues of $H_0 + \lambda W$ occurring in the gaps of $\sigma(H_0)$ for small values of the coupling constant λ.

III. COUPLING CONSTANT THRESHOLD BEHAVIOR FOR $H_0 - \lambda W$ AT $\lambda = 0$

In this section we are going to show that in the case of a non-negative potential W having an exponential falloff $\tilde{E}_{2N+1}(\lambda)$, the smallest eigenvalue of $H_0 - \lambda W$ inside the spectral gap $(E_{2N}(0), E_{2N+1}(0))$, is an analytic function of λ at $\lambda = 0$. First of all, we notice that

$$\lim_{\lambda \to 0} \tilde{E}_{2N+1}(\lambda) = E_{2N+1}(0),$$

since $H_0 - \lambda W \to H_0$ as $\lambda \to 0$ in the norm resolvent sense [in the case of the gap $((N-\frac{1}{2})^2, E_{2N}(\pi))$, $\tilde{E}_{2N}(\lambda)$ clearly converges to $E_{2N}(\pi)$].

In order to show the analyticity of $\tilde{E}_{2N+1}(\lambda)$ at $\lambda = 0$ we shall follow the strategy used in Ref. 8 (p. 337) since we know that $E \in (E_{2N}(0), E_{2N+1}(0))$ is an eigenvalue of $H_0 - \lambda W$ if and only if

$$\det(I - \lambda W^{1/2}(H_0 - E)^{-1} W^{1/2}) = 0. \quad (3.1)$$

First of all, the BS kernel can be expressed as follows:

$$W^{1/2}(H_0 - E)^{-1} W^{1/2}$$

$$= M_E^{(1)} + M_E^{(2)} + \frac{1}{2\pi} \left(\int_0^{2\pi} \frac{1}{E_{2N+1}(\theta) - E} \, d\theta \right)$$

$$\times |W^{1/2} \phi_{2N+1}^{(0)}\rangle \langle W^{1/2} \phi_{2N+1}^{(0)}| \quad (3.2)$$

with

$$M_E^{(1)} = \int_0^{2\pi} \frac{|W^{1/2}\phi_{2N+1}^{(\theta)}\rangle\langle W^{1/2}\phi_{2N+1}^{(\theta)}| - |W^{1/2}\phi_{2N+1}^{(0)}\rangle\langle W^{1/2}\phi_{2N+1}^{(0)}|}{E_{2N+1}(0) - E} \frac{d\theta}{2\pi} \quad (3.3)$$

and

$$M_E^{(2)} = \sum_{n \neq 2N+1} \int_0^{2\pi} \frac{|W^{1/2}\phi_n^{(\theta)}\rangle\langle W^{1/2}\phi_n^{(\theta)}|}{E_n(\theta) - E} \frac{d\theta}{2\pi}. \quad (3.4)$$

First of all, we notice that the operator-valued function $M_E^{(2)}$ is analytic at $E = E_{2N+1}(0)$. At this point we are going to show the boundedness of $M_{E_{2N+1}(0)}^{(1)}$ under a suitable assumption on W.

Proposition 3.1: If $W^{1/2} \in D(x^2)$ then $M_{E_{2N+1}(0)}^{(1)}$ is a bounded operator on $L^2(-\infty, +\infty)$.

Proof: First of all, the integral expression $M_{E_{2N+1}(0)}^{(1)}$ can be written in a more convenient form by using the antiunitarity of the operators $H_0(\theta)$ and $H_0(2\pi - \theta)$, i.e.,

$$M_{E_{2N+1}(0)}^{(1)} = \int_0^{\pi} \frac{|W^{1/2}\overline{\phi_{2N+1}^{(\theta)}}\rangle\langle W^{1/2}\overline{\phi_{2N+1}^{(\theta)}}| + |W^{1/2}\overline{\phi_{2N+1}^{(\theta)}}\rangle\langle W^{1/2}\overline{\phi_{2N+1}^{(\theta)}}| - 2|W^{1/2}\overline{\phi_{2N+1}^{(0)}}\rangle\langle W^{1/2}\overline{\phi_{2N+1}^{(0)}}|}{E_{2N+1}(\theta) - E_{2N+1}(0)} \frac{d\theta}{2\pi}, \quad (3.5)$$

which can also be written

$$2 \int_0^{2\pi} \frac{|W^{1/2}\Re\phi_{2N+1}^{(\theta)}\rangle\langle W^{1/2}\Re\phi_{2N+1}^{(\theta)}| + |W^{1/2}\Im\phi_{2N+1}^{(\theta)}\rangle\langle W^{1/2}\Im\phi_{2N+1}^{(\theta)}| - |W^{1/2}\phi_{2N+1}^{(0)}\rangle\langle W^{1/2}\phi_{2N+1}^{(0)}|}{E_{2N+1}(\theta) - E_{2N+1}(0)} \frac{d\theta}{2\pi}. \quad (3.6)$$

At this point we want to show that the norm of the operator defined by (3.6) is finite. From the Kronig–Penney relation we know that the function $E_{2N+1}(\theta) - E_{2N+1}(0)$ has only a zero at $\theta = 0$.

Furthermore, $E_{2N+1}(\theta) - E_{2N+1}(0)$ goes to zero like θ^2 when θ goes to zero since $E_{2N+1}(\theta)$ is an even function and

$$\frac{d^2 E_{2N+1}(\theta)}{d\theta^2}\Big|_{\theta=0} > 0 \quad (3.7)$$

as can be seen by computing the implicit function derivative by means of the Kronig–Penney relation.

Therefore $\theta = 0$ is a removable singularity for the function $\theta^{-2}[E_{2N+1}(\theta) - E_{2N+1}(0)]$ which implies also

$$\inf_{\theta \in [0,\pi]} \theta^{-2}[E_{2N+1}(\theta) - E_{2N+1}(0)] > 0 \quad (3.8)$$

since the function has no zeros in $[0,\pi]$.

Consequently the operator norm of (3.6) is bounded by

$$2\Big(\inf_{\theta\in[0,\pi]}\theta^{-2}[E_{2N+1}(\theta)-E_{2N+1}(0)]\Big)$$
$$\times\Big|\Big|\int_0^{\pi}\theta^{-2}[\,|W^{1/2}\Re\phi_{2N+1}^{(\theta)}\rangle\langle W^{1/2}\Re\phi_{2N+1}^{(\theta)}|+|W^{1/2}\Im\phi_{2N+1}^{(\theta)}\rangle$$
$$\times\langle W^{1/2}\Im\phi_{2N+1}^{(\theta)}|-|W^{1/2}\phi_{2N+1}^{(0)}\rangle\langle W^{1/2}\phi_{2N+1}^{(0)}|\,]\,\frac{d\theta}{2\pi}\Big|\Big|\,. \tag{3.9}$$

Since $\Re\phi_{2N+1}^{(\theta)}$ and $\Im\phi_{2N+1}^{(\theta)}$ are real analytic functions of θ we have the following McLaurin expansions:

$$\Im\phi_{2N+1}^{(\theta)}=\left[\frac{d}{d\theta}\Im\phi_{2N+1}^{(\theta)}\right]_{\theta=\tilde\theta}\theta \tag{3.10}$$

for some $\tilde\theta\in(0,\theta)$ and

$$\Re\phi_{2N+1}^{(\theta)}=\phi_{2N+1}^{(0)}+\left[\frac{d^2}{d\theta^2}\Re\phi_{2N+1}^{(\theta)}\right]_{\theta=\theta^*}\frac{\theta^2}{2} \tag{3.11}$$

for some $\theta^*\in(0,\theta)$.

Therefore (3.9) becomes

$$2\Big(\inf_{\theta\in[0,\pi]}\theta^{-2}[E_{2N+1}(\theta)-E_{2N+1}(0)]\Big)\Big|\Big|\int_0^{\pi}\Big[\Big|W^{1/2}\Big[\frac{d}{d\theta}\Im\phi_{2N+1}^{(\theta)}\Big]_{\theta=\tilde\theta}\Big\rangle\Big\langle W^{1/2}\Big[\frac{d}{d\theta}\Im\phi_{2N+1}^{(\theta)}\Big]_{\theta=\tilde\theta}\Big|$$
$$+\frac{1}{2}\Big(|W^{1/2}\phi_{2N+1}^{(0)}\rangle\Big\langle W^{1/2}\Big[\frac{d^2}{d\theta^2}\Re\phi_{2N+1}^{(\theta)}\Big]_{\theta=\theta^*}\Big|+\Big|W^{1/2}\Big[\frac{d^2}{d\theta^2}\Re\phi_{2N+1}^{(\theta)}\Big]_{\theta=\theta^*}\Big\rangle\langle W^{1/2}\phi_{2N+1}^{(\theta)}|\Big)$$
$$+\frac{\theta^2}{4}\Big|W^{1/2}\Big[\frac{d^2}{d\theta^2}\Re\phi_{2N+1}^{(\theta)}\Big]_{\theta=\theta^*}\Big\rangle\Big\langle W^{1/2}\Big[\frac{d^2}{d\theta^2}\Re\phi_{2N+1}^{(\theta)}\Big]_{\theta=\theta^*}\Big|\Big\rangle\frac{d\theta}{2\pi}\Big|\Big|\,. \tag{3.12}$$

By using the explicit expression of $\phi_n^{(\theta)}$ given in Sec. II, it is not difficult to show that there are two positive constants $A_1^{(N)}$, $B_1^{(N)}$ such that

$$\Big|\frac{d}{d\theta}\Im\phi_{2N+1}^{(\theta)}(x)\Big|\leqslant A_1^{(N)}|x|+B_1^{(N)}$$

for any real x and any $\theta\in[0,\pi)$. Similarly we have

$$\Big|\frac{d^2}{d\theta^2}\Re\phi_{2N+1}^{(\theta)}(x)|\leqslant A_2^{(N)}x^2+B_2^{(N)}|x|+C^{(N)}$$

for some positive constants $A_2^{(N)}$, $B_2^{(N)}$, $C^{(N)}$.

These estimates together with our assumption on $W^{1/2}$ imply that the operator-valued integral inside the norm in (3.12) has norm bounded by

$$2\int_0^{\pi}[\,\|W^{1/2}(A_1^{(N)}|x|+B_1^{(N)})\|_2^2+\|W^{1/2}\phi_{2N+1}^{(0)}\|_2\|W^{1/2}(A_2^{(N)}x^2+B_2^{(N)}|x|+C^{(N)})\|_2$$
$$+\frac{\theta^2}{4}\|W^{1/2}(A_2^{(N)}x^2+B_2^{(N)}|x|+C^{(N)}\|_2^2\Big]\frac{d\theta}{2\pi}<\infty, \tag{3.13}$$

which completes the proof of the proposition. Q.E.D.

Since $M_E^{(1)}$ and $M_E^{(2)}$ are both bounded at $E=E_{2N+1}(0)$ it follows that, if $E(\lambda)$ is a solution of (3.1) with the BS kernel given by (3.2), $[I-\lambda(M_{E(\lambda)}^{(1)}+M_{E(\lambda)}^{(2)})]$ is invertible for any λ sufficiently small. Therefore, following Ref. 8, we only need to study the equation

$$\det\Big[I-\frac{\lambda}{2\pi}\Big(\int_0^{2\pi}\frac{1}{E_{2N+1}(\theta)-E}\,d\theta\Big)|W^{1/2}\phi_{2N+1}^{(0)}\rangle$$
$$\times\langle W^{1/2}\phi_{2N+1}^{(0)}|[I-\lambda(M_E^{(1)}+M_E^{(2)})]^{-1}\Big]=0. \tag{3.14}$$

Since for any rank 1 operator B we have $\det(1+B)=1+\mathrm{Tr}(B)$, (3.14) becomes

$$1-\frac{\lambda}{2\pi}\Big(\int_0^{2\pi}\frac{1}{E_{2N+1}(\theta)-E}\,d\theta\Big)\langle W^{1/2}\phi_{2N+1}^{(0)},$$
$$[I-\lambda(M_E^{(1)}+M_E^{(2)})]^{-1}W^{1/2}\phi_{2N+1}^{(0)}\rangle=0. \tag{3.15}$$

At this point we state and prove the main theorem of this section.

Theorem 3.2: Let H_0 be the Kronig–Penney Hamiltonian. If $W \geqslant 0$ and $W^{1/2} \in D(e^{\alpha|x|})$ for some $\alpha > 0$, then (a) $\tilde{E}_n(\lambda)$, the smallest eigenvalue of $H_0 - \lambda W$ in the spectral gap $(E_{n-1}(0), E_n(0))$ for n odd, $(E_{n-1}(\pi), E_n(\pi))$ for n even, is analytic at $\lambda = 0$ and

$$\tilde{E}_n(\lambda) = \tilde{E}_n - \frac{\tilde{m}_n^*}{2}\left(\int_{-\infty}^{\infty} W(x)[\tilde{\phi}_n(x)]^2\,dx\right)^2 \lambda^2 + o(\lambda^2), \tag{3.16}$$

where

$$\tilde{E}_n = \begin{cases} E_n(0), & n = 2k+1, \\ E_n(\pi), & n = 2k, \end{cases}$$

$$m_n^* = \begin{cases} m_n^*(0), & n = 2k+1, \\ m_n^*(\pi), & n = 2k, \end{cases}$$

$$\tilde{\phi}_n(x) = \begin{cases} \phi_n^{(0)}(x), & n = 2k+1, \\ \phi_n^{(r)}(x), & n = 2k, \end{cases}$$

$m_n^*(\theta)$ being the effective mass; (b) $\lambda = 0$ is a coupling constant threshold.

Proof: By setting $E = E(\eta) = E_{2N+1}(0) - \eta^2$, $\eta > 0$ and multiplying both sides of Eq. (3.15) by η we get the equation

$$\eta - \frac{\lambda}{2\pi}\left(\int_0^{2\pi} \frac{\eta}{[E_{2N+1}(\theta) - E_{2N+1}(0)] + \eta^2}\,d\theta\right) \times (W^{1/2}\phi_{2N+1}^{(0)}, [I - \lambda(M_{E(\eta)}^{(1)} + M_{E(\eta)}^{(2)})]^{-1} W^{1/2}\phi_{2N+1}^{(0)}) = 0. \tag{3.17}$$

Thus we must show the existence of the function $\eta(\lambda)$ solution of Eq. (3.17) in a neighborhood of $\lambda = 0$ and its analyticity at $\lambda = 0$.

In order to achieve this result we shall use the implicit function theorem applied to Eq. (3.17). Therefore we have to prove that $F(\eta,\lambda)$ given by the left-hand side of Eq. (3.17) is jointly analytic in η and λ at $\langle \eta,\lambda \rangle = \langle 0,0 \rangle$ and that $F(\eta,\lambda)$ satisfies the conditions $F(0,0) = 0$, $(\partial F/\partial \eta)_0 \neq 0$.

First of all, we notice that $M_{E(\eta)}^{(2)}$ is an analytic function of η at 0 since

$$(\psi, M_{E(\eta)}^{(2)}\psi) = \sum_{n \neq 2N+1} \int_0^{2\pi} \frac{|(\psi, W^{1/2}\phi_n^{(\theta)})|^2}{[E_n(\theta) - E_{2N+1}(0)] + \eta^2} \frac{d\theta}{2\pi} \tag{3.18}$$

is an analytic function of η at $\eta = 0$ for any $\psi \in L^2(R)$ which implies the analyticity of the operator-valued function $M_{E(\eta)}^{(2)}$ (see Ref. 13 for the relation between analyticity in the weak operator topology and norm analyticity). Now we must show the analyticity of the functions

$$f(\eta) = \int_0^{2\pi} \frac{\eta}{[E_{2N+1}(\theta) - E_{2N+1}(0)] + \eta^2}\,d\theta$$

and

$$M_{E(\eta)}^{(1)} = \int_0^{2\pi} \frac{|W^{1/2}\phi_{2N+1}^{(\theta)}\rangle\langle W^{1/2}\phi_{2N+1}^{(\theta)}| - |W^{1/2}\phi_{2N+1}^{(0)}\rangle\langle W^{1/2}\phi_{2N+1}^{(0)}|}{[E_{2N+1}(\theta) - E_{2N+1}(0)] + \eta^2} \frac{d\theta}{2\pi}$$

at $\eta = 0$.

Let us begin by considering the first integral which can also be written as

$$\int_{-\pi}^{\pi} \frac{\eta}{g^2(\theta) + \eta^2} \frac{d\theta}{2\pi}$$

with

$$g^2(\theta) = E_{2N+1}(\theta) - E_{2N+1}(0) = \sum_{l=1}^{\infty} \beta_{2l}\theta^{2l}.$$

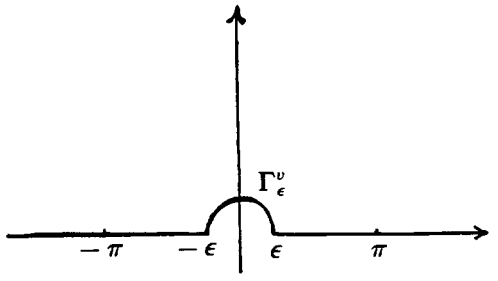
Furthermore

$$\int_{-\pi}^{\pi} \frac{\eta}{g^2(\theta) + \eta^2} \frac{d\theta}{2\pi} = \frac{i}{4\pi}\left[\int_{-\pi}^{\pi} \frac{1}{g(\theta) + i\eta}\,d\theta - \int_{-\pi}^{\pi} \frac{1}{g(\theta) - i\eta}\,d\theta\right]. \tag{3.19}$$

First of all, we note that $g(\theta)$ is real analytic and has an analytic extension to a complex neighborhood of $(-\pi,\pi)$.

If $\eta > 0$ the singularity of $1/[g(\theta) + i\eta]$ is located in the lower half-plane as can be seen by noticing that the leading term of the MacLaurin expansion of $g(\theta)$ is given by $\beta_2^{1/2}\theta$ with $\beta_2^{1/2} > 0$ and therefore the solution of $g(\theta) = -i\eta$ lies on the negative imaginary semiaxis.

Thus we can choose the path of integration shown below



and we have

$$\int_{-\pi}^{\pi} \frac{1}{g(\theta) + i\eta}\, d\theta = \int_{-\pi}^{-\epsilon} \frac{1}{g^2(\theta) + i\eta}\, d\theta + \int_{\Gamma_\epsilon^v} \frac{1}{g(\theta) + i\eta}\, d\theta + \int_{\epsilon}^{\pi} \frac{1}{g(\theta) + i\eta}\, d\theta . \tag{3.20}$$

of course we must choose the opposite path for the other integral

$$\int_{-\pi}^{\pi} \frac{1}{g(\theta) - i\eta}\, d\theta = \int_{-\pi}^{-\epsilon} \frac{1}{g(\theta) - i\eta}\, d\theta + \int_{\Gamma_\epsilon^d} \frac{1}{g(\theta) - i\eta}\, d\theta + \int_{\epsilon}^{-\pi} \frac{1}{g(\theta) + i\eta}\, d\theta . \tag{3.21}$$

Thus for any $\epsilon > 0$ we can find a suitable complex neighborhood of the origin in which both integrals on the left-hand sides of (3.20) and (3.21) are analytic functions of η since the integrands on the respective right-hand sides have no singularities along the path of integration. Therefore the function $f(\eta)$ is analytic at $\eta = 0$.

Furthermore, we obtain

$$\lim_{\eta \to 0} \int_{-\pi}^{\pi} \frac{\eta}{g^2(\theta) + \eta^2} \frac{d\theta}{2\pi} = \left(2 \frac{d^2 E_{2N+1}(\theta)}{d\theta^2} \right)_{\theta = 0}^{-1/2} . \tag{3.22}$$

By adopting a notation widely used in solid-state physics we can write the right-hand side of (3.22) as

$$(m^*_{2N+1}(0)/2)^{1/2},$$

$m^*_{2N+1}(0)$ being the so-called effective mass evaluated at the left boundary of the $(2N+1)$st band.

Now we must show the analyticity of the operator function $M^{(1)}_{E(\eta)}$ at the origin which is equivalent to showing that the function

$$(\psi, M^{(1)}_{E(\eta)} \psi) = \int_0^{2\pi} \frac{|(\psi, W^{1/2}\phi^{(\theta)}_{2N+1})|^2 - |(\psi, W^{1/2}\phi^{(0)}_{2N+1})|^2}{[E_{2N+1}(\theta) - E_{2N+1}(0)] + \eta^2} \frac{d\theta}{2\pi} \tag{3.23}$$

is analytic at $\eta = 0$ for any $\psi \in L^2(R)$. Let us show that the function

$$p(\theta) = |(\psi, W^{1/2}\phi^{(\theta)}_{2N+1})|^2 - |(\psi, W^{1/2}\phi^{(0)}_{2N+1})|^2$$

can be analytically extended to a complex neighborhood of the origin. As follows from our analysis of Sec. II, the generalized eigenfunctions of our Hamiltonian are given by the Bloch functions

$$\phi^{(\theta)}_n(x) = A^{(\theta)}_n e^{im\theta} \left[\cos\sqrt{E_n(\theta)}\,(x - 2m\pi) - ((1 - e^{i\theta})/(1 + e^{i\theta}))\cot\sqrt{E_n(\theta)}\,\pi \times \sin\sqrt{E_n(\theta)}\,(x - 2m\pi) \right] , \tag{3.24}$$

for $x \in [(2m-1)\pi, (2m+1)\pi]$ and for any m.

Since the eigenfunctions (3.24) are real-analytic functions of θ it follows that

$$P(\theta) = |W^{1/2}\phi^{(\theta)}_{2N+1}\rangle\langle W^{1/2}\phi^{(\theta)}_{2N+1}|$$

is a real-analytic rank-one operator-valued function due to the fact that $W^{1/2}\phi^{(\theta)}_{2N+1}$ is a real-analytic $L^2(R)$-valued function. From (3.24) we have for θ in a small neighborhood of 0

$$|\phi^{(\theta)}_{2N+1}(x)| \leqslant 2e^{|\Im\sqrt{E_{2N+1}(\theta)}|\pi} e^{|\Im\theta|\,|m|} \tag{3.25}$$

for any $x \in [(2m-1)\pi, (2m+1)\pi)$.

We notice that for any $|m| > 2$, $e^{\Im\theta\,|m|}$ is bounded by $e^{(3/2)|\Im\theta|\,|x|}$.

Therefore if $|\Im\theta| < \frac{2}{3}\alpha$ it follows from our assumption on W that $W^{1/2}\phi^{(\theta)}_{2N+1} \in L^2(R)$, which is equivalent to saying that in any neighborhood of 0 satisfying $|\Im\theta| < \frac{2}{3}\alpha$, $P(\theta)$ defined as above is an analytic rank 1 operator-valued function which implies that $p(\theta)$ has an analytic extension to a complex neighborhood of the origin. This fact allows us to use a procedure similar to the one used for the function $f(\eta)$ in order to show that the function $(\psi, M^{1)}_{E(\eta)}, \psi)$ can be analytically extended to a complex neighborhood of $\eta = 0$.

By going back to Eq. (3.17) we obtain that the left-hand side is jointly analytic in η and λ.

Furthermore we get

$$\begin{cases} F(\eta,\lambda)|_{\eta=\lambda=0} = 0, \\ \left.\dfrac{\partial F}{\partial \eta}\right|_{\lambda=0} = 1 . \end{cases} \tag{3.26}$$

Thus we are allowed to apply the implicit function theorem in order to obtain the existence of the function $\eta(\lambda)$ solution of (3.17) in a complex neighborhood of 0 and its analyticity at $\lambda = 0$.

By computing the first term in the Taylor expansion of $\eta(\lambda)$ around $\lambda = 0$ we get

$$\eta(\lambda) = (m^*_{2N+1}(0)/2)^{1/2} \| W^{1/2}\phi^{(0)}_{2N+1} \|_2^2 \lambda + o(\lambda) , \tag{3.27}$$

which implies

$$\eta^2(\lambda) = (m^*_{2N+1}(0)/2) \| W^{1/2}\phi^{(0)}_{2N+1} \|_2^4 \lambda^2 + o(\lambda^2) \tag{3.28}$$

and

$$\tilde{E}_{2N+1}(\lambda) = E_{2N+1}(0) - \frac{m^*_{2N+1}(0)}{2} \times \left(\int_{-\infty}^{+\infty} |W(x)| [\phi^{(0)}_{2N+1}(x)]^2 \, dx \right)^2 \lambda^2 + o(\lambda^2).$$

In the case of a gap of the type $((N-\frac{1}{2})^2, E_{2N}(\pi))$ we have the analogous formula given in the statement of the theorem. Q.E.D.

Remark: A similar result can be shown in the case of a piecewise continuous periodic potential since also in that case the Bloch eigenfunctions can be analytically continued in a neighborhood of $\theta = 0$, if W has an exponential decay and similar formulas can be found for the bound states occurring in the gaps of the spectrum.

ACKNOWLEDGMENTS

I would like to thank Professor M. Klaus for having taken a great interest in this work and having revised the drafts of the paper.

I also wish to thank the referee for his helpful comments and remarks.

Preparation of this manuscript was supported by Department of Energy Grant No. De-FG05-87ER25033 and National Science Foundation Grant No. DMS-8701050.

[1] V. A. Zheludev, "Eigenvalue of the perturbed Schrödinger operator potential," *Topics in Mathematical Physics*, Vol. 2 (Consultants Bureau, New York, 1968).

[2] V. A. Zheludev, "Perturbation of the spectrum of the one-dimensional self-adjoint Schrödinger operator with a periodic potential," *Topics in Mathematical Physics*, Vol. 4 (Consultants Bureau, New York, 1971).

[3] N. E. Firsova, "Levinson formula for the perturbed Hill operator," *Theoretical and Mathematical Physics* (Consultants Bureau, New York, 1985), Vol. 62, pp. 130–140.

[4] L. B. Zelenko, "Asymptotic distribution of eigenvalues in a lacuna of the continuous spectrum of the perturbed Hill operator," *Mathematical Notes of the Academy of Sciences of the USSR* (Consultants Bureau, New York, 1976), Vol. 20, pp. 750–755.

[5] M. Klaus, "Some applications of the Birman–Schwinger principle," Helv. Phys. Acta **55**, 49 (1982).

[6] P. A. Deift and R. Hempel, "On the existence of eigenvalues of the Schrödinger operator $H - \lambda W$ in a gap of $\sigma(H)$," Commun. Math. Phys. **103**, 461 (1986).

[7] I. M. Gel'fand, "Expansion in series of eigenfuncions of an equation with periodic coefficients," Dokl. Akad. Nauk. SSSR **73**, 1117 (1950).

[8] M. Reed and B. Simon, *Methods of Modern Mathematical Physics* (Academic, New York, 1978), Vol. IV.

[9] R. L. Kronig and W. G. Penney, "Quantum mechanics in crystal lattices," Proc. R. Soc. London **130**, 499 (1931).

[10] B. Simon, *Quantum Mechanics for Hamiltonians Defined as Quadratic Forms* (Princeton U. P., Princeton, 1971).

[11] M. Reed and S. Simon, *Methods of Modern Mathematical Physics* (Academic, New York 1975), Vol 11.

[12] S. Albeverio, F. Gesztesy, R. Hoegh-Krohn, and H. Holden, *Solvable Models in Quantum Mechanics, Texts and Monographs in Physics* (Springer, Berlin, 1988).

[13] T. Kato, *Perturbation Theory for Linear Operators*, 2nd ed. (Springer, Berlin, 1976).